NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

HUMAN FACTORS EVALUATION OF COLOR USE IN THE TARGET DATA PROCESSOR RELEASE 10 (TDP R10)

by

Claudia J. Schmidly

March 1989

Thesis Advisor: J. H. Lind

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable) 55	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO.

11 TITLE (Include Security Classification)
HUMAN FACTORS EVALUATION OF COLOR USE IN THE TARGET DATA PROCESSOR RELEASE 10 (TDP R10)

12 PERSONAL AUTHOR(S) SCHMIDLY, Claudia J.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ____ TO ____	1989, March	64

16 SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Color; Color CRT; Operator Machine Interface; IUSS; Test and Evaluation; Human Computer Interface; Military Displays; (KR)

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This thesis provides color use guidelines for static military CRT display formats. A total of 13 guidelines are discussed, relating to color as a coding dimension, the quantity of colors to include, selection of colors to use, ambient luminence, display legibility and readability, human color deficiencies, and operator fatigue. Guidelines are then applied to the operator-machine interface of the U.S. Navy's Target Data Processor Release 10 (TDP R10), a tactical computer workstation for sue in the Integrated Undersea Surveillance System. Specific color related design recommendations are included for the TDP R10 alphanumeric and geographic display screens with the goal of enhancing user performance. Since the TDP R10 is being developed using an iterative design process (design, test, redesign, etc.), test and evaluation considerations also are discussed at length. Various types of user self-report techniques are discussed, along with user (Cont.)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL J. H. Lind	22b TELEPHONE (Include Area Code) 408-646-2453; 22c OFFICE SYMBOL 55Li

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office: 1986—606-24.

Block 19. Abstract (Continued)

performance testing, sample sizes, and data analysis procedures.

Keywords: Cont'd previous page

Approved for public release; distribution is unlimited.

Human Factors Evaluation of Color Use in the Target Data Processor Release 10 (TDP R10)

by

Claudia J. Schmidly
Lieutenant, United States Navy
B.S., University of New Mexico, 1982

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1989

Author: Claudia J. Schmidly

Approved by: Judith H. Lind, Thesis Advisor

LCDR Thomas Mitchell, Second Reader

Peter Purdue, Chairman,
Department of Operations Research

Kneale T. Marshall
Dean of Information and Policy Sciences

ABSTRACT

This thesis provides color use guidelines for static military CRT display formats. A total of 13 guidelines are discussed, relating to color as a coding dimension, the quantity of colors to include, selection of colors to use, ambient luminance, display legibility and readability, human color deficiencies, and operator fatigue. Guidelines are then applied to the operator-machine interface of the U.S. Navy's Target Data Processor Release 10 (TDP R10), a tactical computer workstation for use in the Integrated Undersea Surveillance System. Specific color related design recommendations are included for the TDP R10 alphanumeric and geographic display screens with the goal of enhancing user performance. Since the TDP R10 is being developed using an iterative design process (design, test, redesign, etc.), test and evaluation considerations also are discussed at length. Various types of user self-report techniques are discussed, along with user performance testing, sample sizes, and data analysis procedures.

TABLE OF CONTENTS

ACKNOWLEDGMENT

This thesis would not have been possible without the support of many individuals within the Space and Navy Warfare Systems Command, Undersea Surveillance Program and the Naval Ocean Systems Center, San Diego, CA. Specifically I would like to thank CAPT Ray Witter, LCDR Sherry Deems, and Frank Gambino of SPARWARSYSCOM. At NOSC Tyrone McMurrey and the staff of the OSSD provided invaluable assistance in understanding and operating TDP R10.

A special thanks goes to CDR Leslie Skowronek, SPAWARSYSCOM whose faith and assistance made this all possible.

A thank you to my "co-advisors" Judith Lind and LCDR Tom Mitchell for their assistance and patience. And thanks to Professors G. Lindsay and J. Esary for their statistical expertise.

And finally to Olivia who never allowed me to loose sight of what my real priorities should be.

I. INTRODUCTION

A. EFFECTIVENESS OF COLOR FOR CRT DISPLAYS

Though color has proven effective in many fields, its use in cathode ray tube (CRT) displays has been limited until now. Early color CRTs were extremely expensive and had very poor resolution. Engineering developments have decreased the price and improved the resolution. However, those factors alone should not dictate the decision to use a multicolor (versus monochrome) display for any given purpose. Used correctly, color can improve operator performance. Used incorrectly, it can result in performance decrements.

According to Shneiderman [Ref. 1:pp. 336-337], there are several advantages to using color for computer software driven displays. Color can:

1. Be soothing or striking to the eye
2. Add accents to an uninteresting display
3. Facilitate suble discrimination in complex displays
4. Emphasize the logical organization of information
5. Draw attention to warnings
6. Evoke more emotional reactions of joy, excitement, fear, or anger.

Inappropriately used color can result in the opposite effects. Over use of color codes can increase error rates and reaction times. Inconsistent use of color can confuse the operator.

In order to design an effective color display the software designer needs a set of guidelines. However, color CRT display guidelines, like those for other aspects of human factors engineering, must be tailored to the specific operational task.

B. COLOR CRT DISPLAYS FOR MILITARY SYSTEMS

Color can also be effective for use in military systems. However, most research and development has focused on color CRT displays for aircraft. Aircraft sensors and onboard computers provide vast quantities of rapidly changing data which the operator must correctly interpret. The cost of an error can be a multimillion dollar aircraft and its pilot.

Sanders and McCormick [Ref. 2:p. 79] classify displays which provide rapidly changing data as being dynamic displays. A static display is one in which the information does not change or changes only at a slow rate.

The U.S. Navy's Target Data Processor (TDP) is an example of a static display system. This system provides data fusion and message processing for the Navy's Integrated Undersea Surveillance System (IUSS) community. Although the TDP does present an operator with continually updated tactical information, rate of information change is slow enough to classify this as a static system.

Early versions of the TDP used green monochrome displays. One of the development goals for TDP Release 10 (R10) is to incorporate multicolor displays into the design. Because this is one of the first multicolor displays developed for IUSS, little is known about how to use color effectively in formats for such displays.

At the request of the Space and Naval Warfare Systems Command (SPAWARSYSCOM) Undersea Surveillance Program, an evaluation of the possible use of color for TDP R10 display formats has been undertaken for this study. Guidelines for designing good color CRT formats are required for that evaluation. Several sets of guidelines have been developed for the design of aircraft displays [Ref. 3, Ref. 4, Ref. 5]

While some of the aircraft display design rules apply to both static and dynamic displays, others do not. At present there does not exist in one document a complete set of guidelines for designing static color CRT display formats such as those required for the TDP R10. Such a set of guidelines is critical for any meaningful evaluation, and also for system improvements prior to production and deployment.

C. TEST AND EVALUATION OF DESIGN ALTERNATIVES

Design guidelines provide a starting point in the development process. At some point in development the design will be tested. This may occur early, in the laboratory, as developmental testing or later, in the field, as operational testing. In general, the later in the process design flaws are detected, the more difficult and expensive they are to correct. One way to avoid this problem is the use of an iterative design process where a sequence of design, test, design, etc., is continued until the final product is ready for the user.

Whatever method is chosen, many alternative test and evaluation techniques are available for consideration. The type of technique to use depends on the system to be tested, time, money, operational tempo, etc. Color CRT displays fall into the category of man machine systems. Techniques for evaluating them must take human factors into consideration.

Two techniques frequently used to evaluate man machine systems are objective performance testing and subjective operator evaluations. Both techniques have advantages and disadvantages. However, they can be combined in order to provide a complete evaluation.

The end result of any test procedure is data. The process of analyzing those data in order to make decisions is evaluation. Many data analysis methods are available to study test results and to combine them in meaningful ways.

D. THESIS GOALS

This study has three goals.

1. Develop a set of color-use guidelines for static military CRT display formats.

2. Apply this set of guidelines to the TDP R10 system in the form of design recommendations.

3. Provide some general test and evaluation guidelines for consideration by SPAWARSYSCOM when developing the test plan for TDP R10 system.

To achieve these goals, the following steps need to be completed:

1. Conduct an extensive literature search of research in the field of color use for displays.

2. Based on the literature survey, identify those studies which apply to static military displays and, when possible, which have used modern CRT displays in recent experiments.

3. Develop the guidelines for static military CRT display formats, based on the identified applicable studies.

4. Compare the proposed TDP R10 prototype to the guidelines and note where the prototype does and does not follow them.

5. Recommend improvements to the prototype as appropriate and justify the need for changes.

6. Recommend techniques to test TDP R10 design alternatives and to analyze test results.

E. SCOPE OF THE THESIS

This study to focuses specifically on use of color for static military CRT display formats. Technical details of system development have been kept to a minimum. No attempt has been made to address engineering questions such as design of electronic display systems, generation of specific chromaticities on CRTs, etc.

II. GUIDELINES FOR STATIC COLOR CRT DISPLAY FORMATS

A. LITERATURE REVIEW

An intensive literature review was conducted to locate accepted guidelines for the design of static color CRT display formats. The review utilized several resources including:

1. Technical library database at the Naval Ocean Systems Center (NOSC), San Diego, CA
2. Technical reports and thesis database at the Naval Postgraduate School, Monterey, CA
3. Defense Technical Information Center (DTIC)
4. Ergonomics Abstracts

Information was sought concerning color, color coding, color vision, color displays, and color CRT displays. Hundreds of citations are available on these topics, but many are not applicable to static color CRT displays. This is usually due to one of two reasons:

1. The purpose of the study was to address dynamic display requirements; therefore the scope was limited.
2. The research did not utilize a modern CRT display.

In these cases, where it was uncertain whether the research results could be generalized to static color CRT displays, resulting guidelines were not included in this study.

The remaining material covered a wide range of design factors and proposed guidelines that have been grouped under seven topical areas:

1. Color as a coding dimension
2. Quantity of colors to use
3. Selection of the colors to use
4. Ambient luminance
5. Displ v legibility and readability

6. Human color vision deficiencies

7. Operator fatigue.

The grouping is based more on convenience than on any firm division of information. Each topic tends to be related to others (e.g., legibility is related to the specific color used).

Each topical area is discussed separately along with its resulting guidelines. For convenience, all guidelines are summarized at the end of this chapter.

B. COLOR AS A CODING DIMENSION

Coding of information is the conversion of some real stimulus into an abstract form that may be more easily dealt with by the user. A map is an abstract representation of land, roads, etc., which a driver can carry in the car.

The following discussion of coding is adapted from Sanders and McCormick [Ref. 2:pp. 50-53, 98-101]. Different types or dimensions of coding are available: color, shape, size, alphanumerics, position, etc. Coding can be limited to only a single dimension or multiple dimensions can be combined. Multidimensional coding can be either orthogonal or redundant. In orthogonal coding, each dimension represents unique information. For example, in a shape and color code, shape can represent platform type (e.g., submarine, surface ship, airplane) and color can represent status (e.g., friendly, enemy, unknown). Dimensions in a redundant code carry the same meaning (e.g., both color and shape can represent platform type).

The choice of coding type depends on the task. Some dimensions are thought to be better than others for certain tasks. A single dimension code is the simplest, but is limited by the number of absolute judgments the average person can make. This number depends on the dimension in question (color, size, etc.) and on how much practice the user has had. The maximum that can reliably be used without practice is thought to be 7 ± 2 [Ref. 6]. Orthogonal coding can increase the amount of information conveyed. In redundant coding, one dimension reinforces the others, making it useful if a correct identification is especially important.

Sanders and McCormick [Ref. 2:pp. 53-54] list the following characteristics of a good coding system:

1. Detectability; the stimulus can be perceived by the human sensory system in the ambient environment.

2. Discriminability; one coding symbol is obviously different from another.

3. Meaningfulness; the coding technique is conceptually compatible to the user's expectations.

4. Standardization; consistency is maintained between displays and systems.

5. Multidimensionality; the number and discriminability of coding stimuli are increased through the use of more than one coding dimension.

Color as a coding dimension can prove to be both an advantage and a disadvantage. On the positive side, color is useful in search and identification tasks, especially if symbol density is high [Ref. 7:p. 8-40]. Figure 1 shows how color coding can improve target identification accuracy when both symbol density and exposure time are varied.

Color can be used either as a redundant coding technique or an orthogonal technique. As noted above, a redundant code can improve symbol detectibility while an orthogonal code can increase the amount of information conveyed [Ref. 7:p. 8-38].

Many researchers have stressed the conventional meanings associated with certain colors: red always means danger, yellow caution, etc. As pointed out by Smith and Mosier [Ref. 8:p 184], "Other associations can be learned by a user if color coding is applied consistently." An on-screen color legend can help the operator remember the meanings assigned to the color code.

Color coding also can result in poorer performance. The disadvantages associated with color use can result when the factors mentioned above (and in other parts of this study) are disregarded.

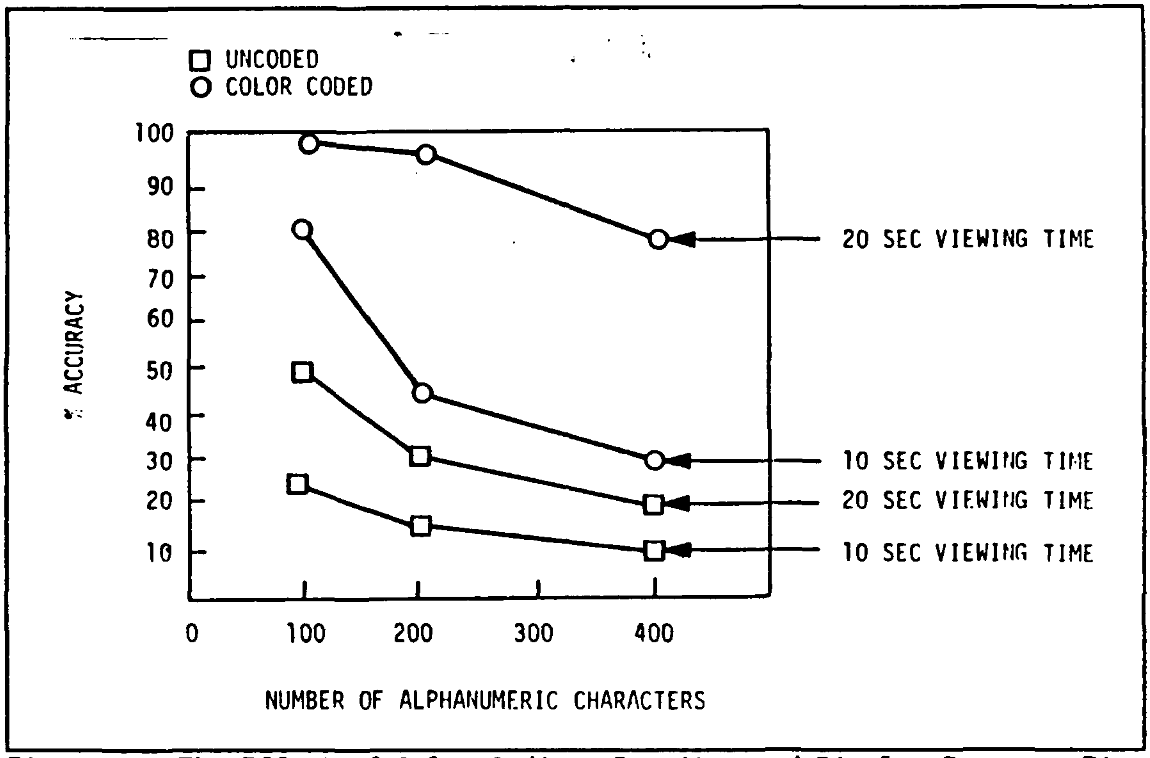

Figure 1. The Effect of Color Coding, Density, and Display Exposure Time on the Accuracy of Locating Targets. [Ref. 3:p. 60]

Color also can be a disadvantage when it is used for tasks for which it is not suited, such as coding quantitative information [Ref. 9:p. 1080]. If care is taken in system design, color coding can improve performance.

GUIDELINES

1. Use color when symbol density is high and in order to group information.

2. Use color consistently.

C. QUANTITY OF COLORS

There is a vast difference between the quantity of colors or hues a color normal individual can distinguish and the number that can be identified as part of a color code. The number of just noticeable color differences that can be distinguished has been estimated as high as

350,000 [Ref. 10]. The number that can be identified with training is approximately 50 [Ref. 11]. Even 50 is considered to be a much larger quantity than is operationally feasible for a color code.

It has been shown that as the quantity of colors in a set increases, operator error and reaction or detection time increase. However, there has been no agreement among researchers as to the maximum quantity recommended for use at one time. Shontz and others [Ref. 12] recommended a maximum of 23 depending on the task. The more widely known and recommended quantity is three to four [Ref. 3:p. 36]. However, this estimate is not based on any data "...from real-world displays," but is "...based on the expectation that ambient lighting may at times be high, that display reliability may be limited, and that fast reaction time of the operator may often be critical," as, for example, in the cockpit environment [Ref. 3:p. 36].

More recent studies have been conducted which raise the range to seven to ten colors. Jacobsen and Neri [Ref. 13] studied the effect of learning on error rates and on time to recognition for color sets of up to seven. Their results are presented in Figure 2. From that figure it can be seen that, although reaction time does increase with set size, the increase is very small (approximately 0.1 second difference in reaction time between sets of size 1 and of size 7). Analysis showed no statistically significant increase in error rates with increase of set size.

Luria and others [Ref. 14] studied the effect of set size on a color matching task involving sets of up to ten colors. It was found that, although reaction time did increase with set size, the increase was not so large as to preclude the use of a set of size 10. However, error rates did increase abruptly with set sizes above seven. Researchers noted that this might be related to the specific colors used and that a carefully chosen set of ten might still result in adequate color matching performance.

Based on these results, a larger number of colors than the traditional four to five may be used under some circumstances. This gives some increased flexibility in display design. Some tasks may require ten colors while others need only two. The choice depends on the specific system being designed. A slight increase in error rate and reaction time may be insignificant if other factors indicate the need for an increase in color code set size.

GUIDELINE

Limit the quantity of colors in the color set to no more than is necessary for task accomplishment, up to a maximum of ten.

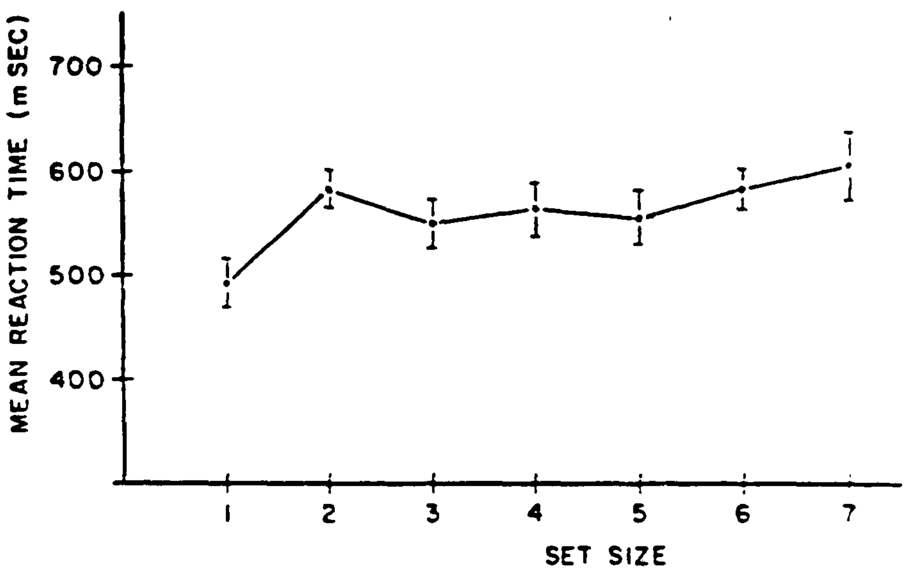

Figure 2. Mean Reaction Times as a Function of the Size of a Set of Colors. I = ± 1 Standard Error. [Ref. 13:p. 10]

D. SELECTION OF COLORS

Before discussion of the use of specific colors for coding, some definitions and concepts concerning color are useful. The following discussion is adapted from Meister [Ref. 7:pp. 180-204], Merrifield and Silverstein [Ref. 5:pp. 10-11], and Rossotti [Ref. 15:pp. 144-145].

Color is not a physical property of an object. What we perceive as color is light of varying composition and intensity. Color can be characterized using three attributes: hue, saturation, and lightness or brightness. Hue depends on the dominant wavelength of the light (i.e., green, red, blue, etc.). Saturation is a measure of how much white light is mixed with the dominant wavelength. For example, the colors red and pink are both the same hue (i.e., red); however, pink has more white light mixed with it, making it desaturated. Lightness or brightness is a subjective measure of luminance or luminous intensity and refers to how much light is transmitted. Figure 3 shows how these three attributes are related. Note the vertical axis showing that black and white have zero saturation and vary through shades of gray only in lightness.

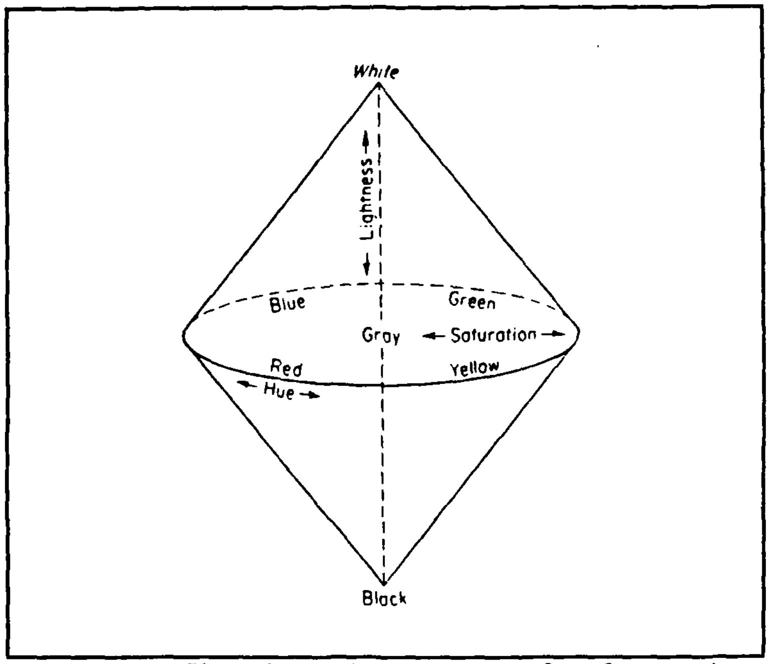

Figure 3. The Three Attributes of Color: Hue, Saturation, and Lightness. [Ref. 2:p. 392]

Several systems are available for describing and specifying colors. The one most often recommended for use with CRT displays is the Commission Internationale de l'Eclairage (CIE) chromaticity system [Ref. 16, Ref. 5]. This system describes colors by their coordinates on what is referred to as a chromaticity diagram. This form of description allows for color replication on different CRTs. This is particularly useful for the translation of research results into design specifications.

Most researchers agree that the ability to discriminate one color from another depends on the color contrast and luminous contrast between the two colors. Several metrics have been developed for measuring perceived color difference. Carter and Carter [Ref. 16] suggest that the CIELUV metric called (delta)E* be used for CRTs. This metric considers hue, saturation, and luminance when measuring the difference between two colors. Research by Carter and Carter [Ref. 17] showed that performance on a target location task deteriorated when the (delta)E* between two colors was less than approximately 40 units. For a more complete definition of this metric and its associated equations see Merrifield and Silverstein [Ref. 5] and Judd and Wyszecki [Ref. 18].

More recent research by Neri and others [Ref. 19] supports the work of Carter and Carter. They evaluated ten sets of seven colors and found that, on a color matching task, performance with seven of the sets was better than performance while using the remaining three. Although the seven sets did result in better performance than the remaining three, they were not significantly different from one another. The performance difference between the two groups of sets could not be attributed to the use of particular colors, but was related to the minimum (delta)E* values between colors in each set. The seven sets with the highest minimum (delta)E* resulted in better performance. For convenience, the seven preferred sets and information concerning them are provided in Appendix A.

The choice of background colors can also affect performance. Neri and others [Ref. 20] tested blue, green, yellow, red, and black (dark gray) backgrounds for mean reaction time to identify targets displayed in seven colors. Figure 4 presents the results of two experiments. In the first experiment, red, yellow, green, and blue were used as background colors. The mean reaction time for the blue background was faster than for all other backgrounds. This was due to the fact that the backgrounds were not matched for brightness. Hence, the blue background appeared brighter so that both color and luminance contrast were higher between it and all target colors.

In the second experiment, all the colored backgrounds were matched for color brightness, red was omitted, and a black background was added for these tests. Under these conditions, reaction time for the blue background was slowest, with black only slightly faster. Based on this, it would seem that black is not a good choice of background color. However, the experimenters note that on "...each of the colored backgrounds there was at least one opponent-colored target which was detected very quickly, whereas with the black background the RTs [reaction times] to all the target colors were of moderate magnitude and much less variable." [Ref. 20:p. 17] Further, the mean reaction times among all backgrounds varied less than 20 msec.

GUIDELINES

1. Choose sets of display colors so that the CIELUV (delta)E* value is maximized, such as those provided in Appendix A.

2. For approximately equal discriminability of all colors, choose an achromatic (gray to black) background.

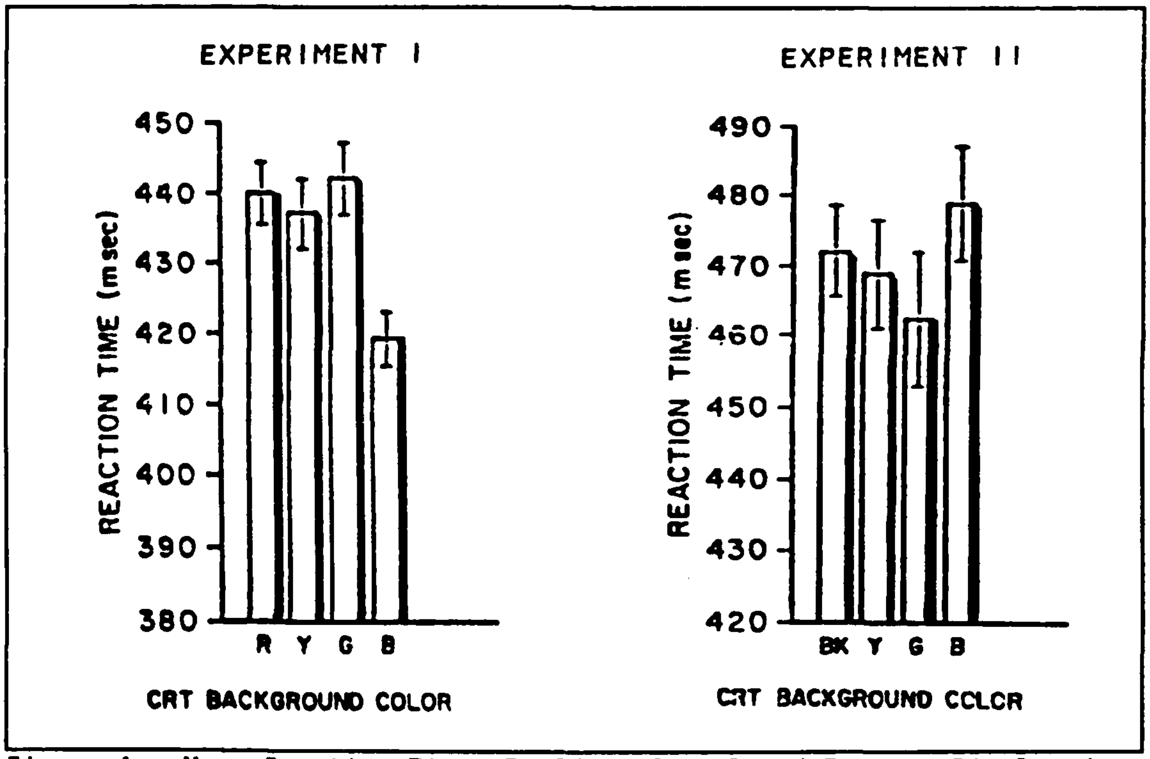

Figure 4. Mean Reaction Times To Identify Colored Targets Displayed on Five CRT Background Colors in Two Experiments. Error Bars Represent ± 1 Standard Error. R = Red, Y = Yellow, G = Green, B = Blue, BK = Black. [Ref. 20: pp. 6,14]

E. AMBIENT LUMINANCE

The design of CRT screens is such that light from sources within the display environment can be reflected on the screen. Light sources include overhead lights, windows, and light reflected off objects such as the operator. Reflections are either a diffuse luminance over the screen or are a mirror-like image known as specular reflection. Specular reflections can be distracting and annoying to the user. Both types affect legibility of the display. [Ref. 2:pp. 420-421]

Background luminance can affect color discrimination. Background luminance consists of CRT raster luminance combined with ambient luminance reflected from the screen. As ambient illumination increases, both color contrast and luminance contrast are decreased and color discriminability is reduced. [Ref. 21:p. 1]

Jacobsen [Ref. 22] compared the effects on performance of two raster luminance levels: black or low luminance and middle gray or intermediate luminance. He found that color set learning was faster and error rates were lower with the middle gray than with the black background. In this study, the CIELUV (delta)E* metric did not serve as a good predictor of performance. Jacobsen attributed this to the qualitative differences in color appearance that can be caused by the background. In this study the luminance level of the gray background was set so that its luminance was higher than the luminance of half of the color set and lower than that of the other half. Although the gray background actually resulted in less color contrast than the black, observers could determine color difference based on whether the sample was lighter or darker than the background instead of on how much lighter or darker [Ref. 22:p. 12].

In a later study, Jacobsen [Ref. 21] looked at the effects of both raster and ambient luminance on multicolor displays. He found that maximum discrimination among colors is achieved when background luminance is set at an intermediate level. "This means that under dark ambient conditions, the raster luminance should be set to an intermediate level but reduced as the ambient illumination increases." [Ref. 21:p. ii]

Colored ambient light has been found to alter perceived color in other display media, but does not appear to affect CRT displays [Ref. 7:p. 8-27]. Neri and others [Ref. 20] studied target-background color combinations under different colored ambient illuminations. They found that the color of ambient light did not affect performance on a target identification task. However, they cautioned that this may be due to the low levels of illumination used. They suggest that subdued white light be used for ambient light if color perception is important [Ref. 20:p. 17].

Various other methods may be used for reducing CRT screen reflections. The light source or the CRT screen can be repositioned. A coating or filter can be applied to the CRT screen. Many antireflective techniques are available, but they can themselves cause legibility problems. For instance, screen etching can blurr the edges of characters and reduce legibility. [Ref. 2:p. 422-423]

GUIDELINES

1. Display formats should be designed and evaluated under the same ambient light as will be present in the operational environment.

2. The raster luminance level should be set depending on ambient light conditions: at a middle gray or intermediate level if ambient lighting is dark, and at a black or low level is ambient light is bright.

3. If, at design time, ambient light conditions are unknown, allow for operator adjustment of raster luminance.

F. DISPLAY LEGIBILITY AND READABILITY

Legibility refers to how well one letter, number, or other symbol can be distinguished from another. Legibility depends on symbol size, stroke width, width-to-height ratio, display resolution, etc., as well as on color and luminance contrast. Readability refers to how well the user can interpret the alphanumerics and other symbols and recognize the information they convey, when grouped into words, sentences, or other collections. Readability depends on spacing between characters and lines and other formatting characteristics related to the grouping of symbols. [Ref. 2:pp. 85-96]

A distinction is made between the size required to detect a symbol and the size required to perceive the symbol's color. Color perception and identification require a larger size. Symbols and alphanumerics displayed on a CRT should be 21-45 minutes of arc in height, increasing as the quantity of colors used increases. The stroke width should be at least 2 minutes of arc and the width-to-height ratio 5:7 or 2:3. Graphic lines should be at least 4 minutes of arc wide. [Ref. 3:pp. 9-10]. These angular measurements may be converted to inches or millimeters using the following formula provided by Sanders and McCormick [Ref. 2:p. 81]:

$$H = (VA \times D) \div 3438$$

where H = symbol height in inches or millimeters

VA = visual angle in minutes

D = viewing distance in inches or millimeters.

The resolution of a CRT depends on how it forms characters. Most use a rectangular shaped matrix of dots to draw each symbol or character. The larger the matrix, the more legible the character. For color display, the matrix should be at least 5 dots wide by 7 dots high. [Ref. 3:p. 13]

Color can contribute to format readability by helping to group information on a single display or across multiple displays [Ref. 1:pp. 339-340]. However, two factors must be kept in mind when using color to format information. First, the colors used must have consistent meaning or confusion will occur. Second, it is usually best to format the display in monochrome first then add color as a redundant code. This is particularly important if a hard copy of the display will be used for training or documentation of the system [Ref. 8:p. 184].

GUIDELINES

1. Design alphanumerics, symbols, and graphic lines large enough to allow for color perception.

2. Choose a CRT display with as large a dot matrix as possible (at least 5 x 7) for the best resolution.

3. Design for monochrome display first, then add color.

G. HUMAN COLOR VISION DEFICIENCIES

One of the most important aspects of color display design is whether the user is physically capable of discriminating between colors. Color vision capability ranges from total color blindness to what is considered to be normal color vision. Table I summarizes the categories of color vision and the discriminations that can be made by each.

TABLE I
CATEGORIES OF COLOR VISION, DISCRIMINATIONS THAT CAN BE MADE BY EACH, AND THEIR INCIDENCE IN THE POPULATION. [Ref. 5:p. 54]

Designation by Number of Discriminations Possible and by Type	Discriminations Light Dark	Yellow Blue	Red Green	Incidence in Population (%) Male	Female
Trichromatism (3)					
Normal	X	X	X	--	--
Protanomaly (red weak)	X	X	weak	1.0	0.02
Deuteranomaly (green weak)	X	X	weak	4.9	0.38
Dichromatism (2)					
Protanopia (red blind)	X	X		1.0	0.02
Deuteranopia (green blind)	X	X		1.1	0.01
Tritanopia (yellow green blind)	X		X	0.002	0.001
Monochromatism (1)					
Congenital Total Color Blindness (cone blindness)	X			0.003	0.002
				8.005	0.433

Among Americans, approximately 8% of males and less than 1% of females have some form of color vision deficiency [Ref. 23:p. 129]. All active duty Navy personnel are tested for color vision using the Farnsworth Lantern test, the preferred method, or the Pseudoisochromatic Plate test [Ref. 24]. However, color vision tests are not infallible. According to the Navy Flight Surgeon's Manual [Ref. 25:p. 343], "The Farnsworth Lantern will pass 95 out of 100 people; in other words, it will pass the 90 percent of people who are normal and the best 5 of the 10 with color vision defects." Even if the test passed only those with normal color vision, there are differences in ability to make fine distinctions [Ref. 18:p. 69]. Both of these facts give support to the idea of allowing some operator selection of display colors.

User preference was studied by d'Ydewalle and others [Ref. 26]. Users were asked to select three preferred color combinations out of 256 possible. The five most commonly chosen were used in a detection task. The results showed that color per se had no effect on performance, but that ability to use a preferred color combination did improve performance.

GUIDELINE

When fine discriminations are necessary, allow user selection of the color palette.

H. OPERATOR FATIGUE

With the increasing use of CRTs there has been a growing number of complaints of operator fatigue. These complaints concern operator vision, headaches, muscular pain, nausea, etc. Many researchers have attempted to study this problem, but with limited success. Most attempts have focused on finding the factors inherent in CRTs that may cause operator fatigue. Many early studies did indeed show a causal relationship, but more recent work is not so conclusive.

One of the problems with earlier studies, as pointed out by Starr [Ref. 27], is the lack of or inappropriate use of control groups in a study. Virtually all workers in all types of jobs will report fatigue at the end of a work day. To alleviate this experimental problem, Starr conducted two studies in 1982 and 1984 which used questionnaires administered to both CRT users and a control group doing the same job with paper documents. In both studies subjects were asked about physical discomforts. In the second study they were also asked to rate the level of discomfort, if it existed. In the first study, CRT operators reported slightly higher numbers of discomforts of all types and significantly higher neck discomfort. However, when subjects were equated by age, the results were shown to be age related versus CRT related.

The results of the second study were somewhat different. The CRT users reported more incidence of blurred vision and discomfort in the buttocks, and the paper users reported more headaches and nausea. Both groups reported an almost equal number of users who felt their vision had deteriorated in the previous year. No correlation was found between age and type or level of discomfort. Buttocks discomfort can be explained by the fact that the CRT users spent more time seated than did the paper users. No explanation could be found for the blurred vision. Interestingly, more CRT users than paper users preferred their display medium over the highly legible questionnaire form used for the study. The study indicates that, although CRT operators do suffer physical discomforts, these discomforts are different in type rather than number from those experienced in other sedentary jobs.

Other researchers have noted the high incidence of visual and shoulder/neck discomforts among CRT operators [Ref. 28:p. 1637]. Zwahlen and others [Ref. 29:p. 1640] studied shoulder and neck discomfort in CRT operators and found that subjective ratings of discomfort increased after each work period, but were less after work periods which included short pauses.

Are there features inherent in CRT displays which can cause visual discomfort? The most common hypothesis to explain visual fatigue is that the frequent reaccomodation and convergence necessary in a visually demanding task causes fatigue of the eye muscles. Mourant and others [Ref. 30] did find that CRT use caused visual fatigue when the task involved uninterrupted viewing. Hedman and Briem [Ref. 31] found that visual fatigue increased with time on task, but fatigue was not limited to CRT use.

Another potential problem is chromatic aberration. This phenomenon refers to the inability of the eye to focus on more than one wavelength at a time [Ref. 9:p. 1081]. That is, for the eye at rest, violet light will focus in front of the retina, while red wavelengths focus behind. It is thought that this would require constant reaccommodation by the eye.

Weitzman points out that the accommodation lens is always in motion which may explain why a connection between reaccommodation and visual fatigue has not been found. He is supported by the findings of Matthews and Mertins [Ref. 32:pp. 1275] who did not find a relationship between color display and subjective discomfort.

CRT use can cause visual and shoulder/neck fatigue, but the incidence is not limited to CRTs and is more likely caused by other factors. These include age, uninterrupted sitting, frequency of breaks, etc.

GUIDELINE

The work routine associated with the display should include breaks which allow the operator to move around.

I. SUMMARY OF GUIDELINES

1. Use color when symbol density is high and in order to group information.

2. Use color consistently.

3. Limit the quantity of colors in the color set to no more than is necessary for task accomplishment, up to a maximum of ten.

4. Choose sets of display colors so that the CIELUV (delta)E* value is maximized, such as those provided in Appendix A.

5. For approximately equal discriminability of all colors, choose an achromatic (gray to black) background.

6. Display formats should be designed and evaluated under the same ambient light as will be present in the operational environment.

7. The raster luminance level should be set depending on ambient light conditions: at a middle gray or intermediate level if ambient lighting is dark, and at a black or low level if ambient light is bright.

8. If, at design time, ambient light conditions are unknown, allow for operator adjustment of raster luminance.

9. Design alphanumerics, symbols, and graphic lines large enough to allow for color perception.

10. Choose a CRT display with as large a dot matrix as possible (at least 5 x 7) for the best resolution.

11. Design for monochrome display first, then add color.

12. When fine discriminations are necessary, allow user selection of the color palette.

13. The work routine associated with the display should include breaks which allow the operator to move around.

III. TARGET DATA PROCESSOR RELEASE 10 (TDP R10)

A. BACKGROUND

The Integrated Undersea Surveillance System (IUSS) is an element of the U.S. Navy's Anti-Submarine Warfare program. It provides early warning and cueing of enemy submarine forces as well as maintaining current intelligence on their locations and movements [Ref. 33]. Information collected by various components of IUSS is reported to evaluation centers. To facilitate the processing of this information the Target Data Processor (TDP) was developed. This subsystem provides the following functions:

1. Tactical coordination, command, and control
2. Generation of tracking and fixing data
3. Generation and release of operational directives to IUSS components
4. Generation and release of tactical information to higher authority
5. Support intelligence and historical data gathering and analysis.

Since the original development of the TDP, the system has been updated and released in nine different versions. The current TDP Release 9 (TDP R9) consists of an AN/UYK-7 digital computer with two processing units and 192K of core memory. The primary user interface consists of a dual-screen computer workstation with an alphanumeric keyboard and a trackball. The left hand screen displays geographic data including target positions, target tracks, sensor locations, etc. The right hand screen displays alphanumeric data such as messages, target summaries, data input fields, etc. Both screens are monochrome green. [Ref. 34]

The evaluation center watch section includes an Ocean Systems Watch Officer (OWO) and several Ocean Systems Technician Analysts (OTAs) who hold Navy Enlisted Classification OT-0612, TDP Displays Analyst. Each member of the watch section is assigned a workstation and is responsible for specific data handling duties. Workstations are initialized with specific modes of operation depending on the watch position of the operator. For instance, the OWO mode allows for release of formal message traffic.

The current TDP release is considered to be limited in terms of main memory and processing time. This is of particular concern since its data handling requirements have been increasing with time. Also, the older hardware does not allow for expansion of the system's functional capabilities. Currently the system generates formatted tactical messages according to RAINFORM reporting requirements. To conform with current Navy requirements, the system must be converted to JINTACCS format. The system's sponsor, SPAWARSYSCOM Undersea Surveillance Program, has tasked the Naval Ocean Systems Center (NOSC), San Diego, with development of TDP R10.

The development goals for TDP R10 are:

1. To rehost current TDP R9 algorithms on commercial, off-the-shelf computers and peripheral equipment

2. To incorporate JINTACCS message generation capability

3. To provide an environment for prototyping and evaluation of new functions and subsystems

4. To enhance the OMI through use of state-of-the-art windowing, menu driven operations, and multicolor displays [Ref. 34].

The TDP R10 shares a standard operating environment with another IUSS subsystem, the Universal Communications Processor Release 6 (UCP R6). This subsystem is a smaller, single screen workstation that is limited to message generation and release functions. Both systems are considered to be early prototypes of a future workstation known as the Advanced Surveillance Workstation which will incorporate multiple subsystem capabilities and will replace both the TDP and the UCP.

B. SYSTEM DESCRIPTION

As part of TDP R10 development, multiple prototype versions will be developed and tested. The following system description is adapted from preliminary documentation of TDP R10 Version 3.0 provided by NOSC San Diego, CA [Ref. 35].

The hardware for the TDP R10 Version 3.0 consists of a Sun Workstation with two CRTs. One CRT is multicolor and displays geographic and alphanumeric data. The other CRT is monochrome and displays only alphanumeric data. User interaction is accomplished with an alphanumeric keyboard and a mouse. The functional capabilities of this prototype are very limited.

The TDP R10 uses the desktop management processing concept. This concept uses windowing and icons to provide user interface with the system's functions. To access a function, the on-screen cursor is positioned to highlight the icon representing the function; a keyboard or mouse button is then pressed to select the function. Icons can be either graphical or textual. For TDP R10 all icons are textual.

Windows partition a CRT screen into functional areas. A window may remain on screen at all times, it may appear due to operator selection of a function, or it may appear due to system processes. Windows may operate independently of one another, allowing multiple functions to occur at the same time. Window operations may also be dependent, when the action taken with one window affects another. Simultaneous display of multiple windows can be accomplished with tiling, where no window can overlay any portion of another, or with stacking, where windows are stacked one on top of another.

A design goal for TDP R10 is to provide a simple, consistent user interface with the system. To that end each function window is constructed of a standardized toolset of objects. An object can be considered a window within the function window. Each window uses a combination of objects which are themselves made up of lower level tools. Figure 5 shows the object toolset and lower level tools developed for the TDP R10. Figure 6 shows examples of three lower level tools.

As shown in Figure 5, objects are grouped under three families, each with a different purpose. The purpose of program control objects is to allow the operator to direct program flow, select options, and specify data. Data entry objects allow for the rapid entry and editing of data.

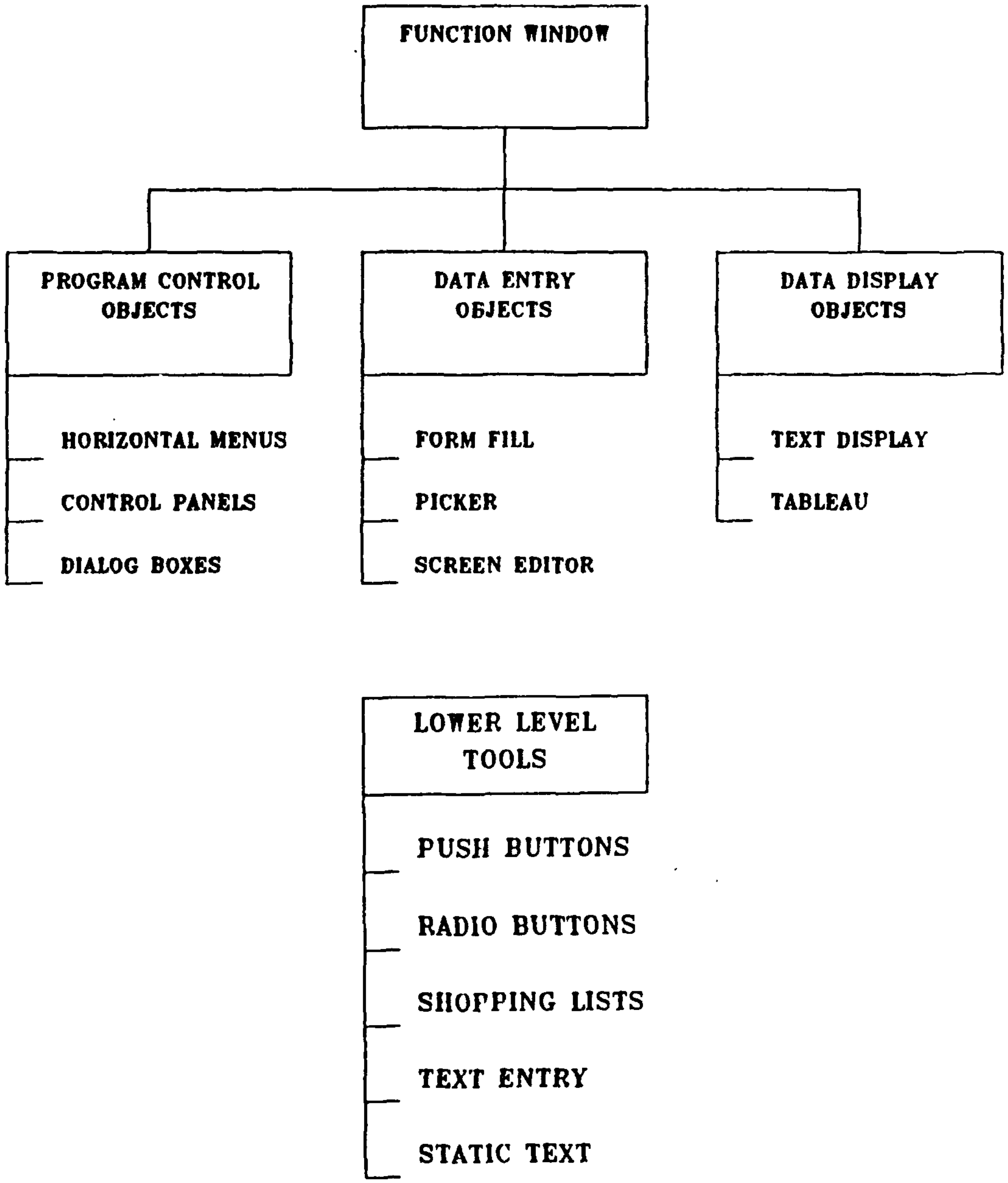

Figure 5. The Function Window Object Toolset and Lower Level Tools for TDP R10. [Ref. 35]

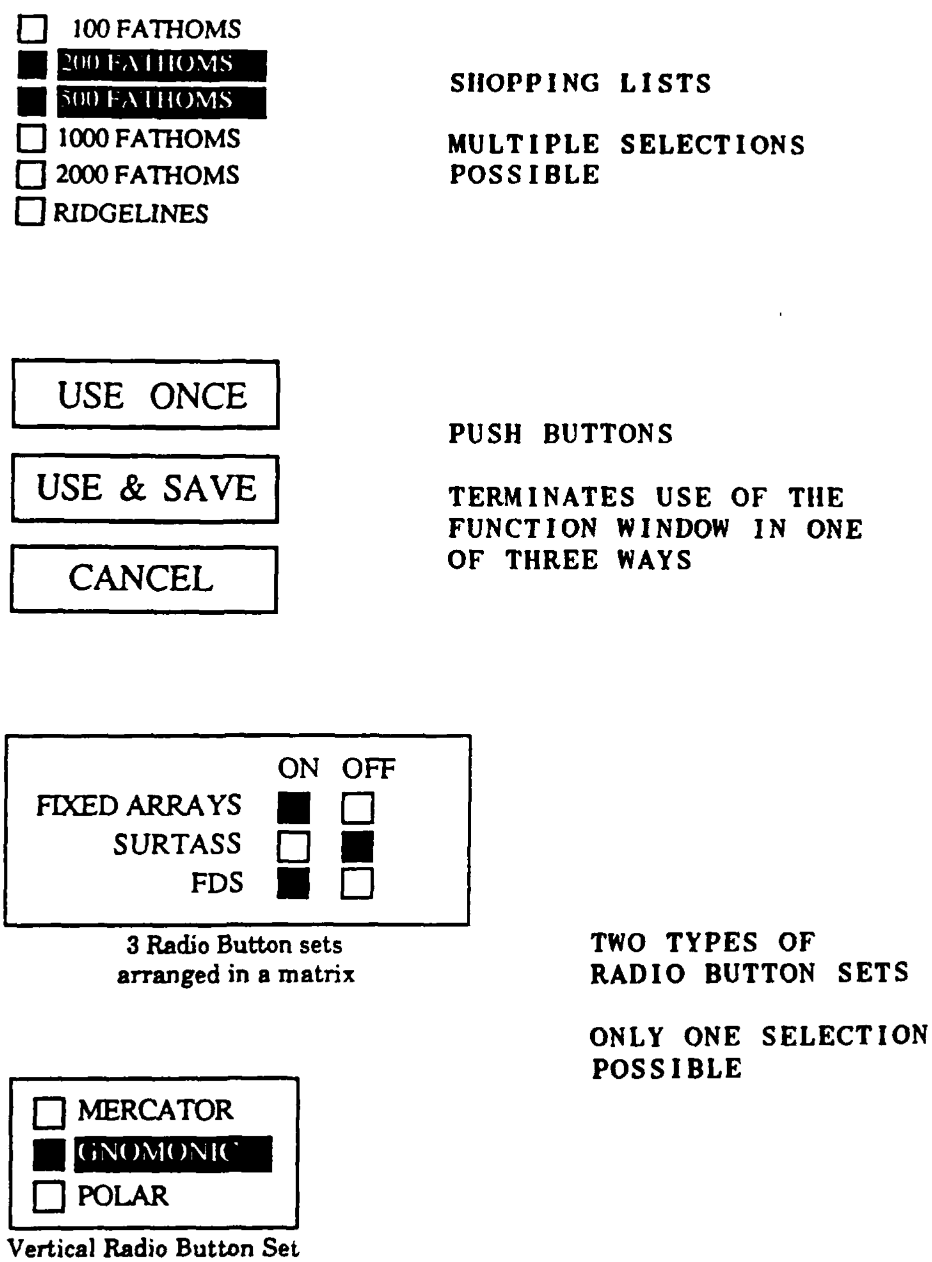

Figure 6. Examples of Three Lower Level Tools for TDP R10. [Ref. 35]

Data display objects provide only review and transfer of data. Figure 7 is an example of a TDP R10 function window using program control and data display objects.

1. Alphanumeric Screen

The monochrome alphanumeric screen is designed with a light gray background; characters and lines are shown in black. The screen area is divided into three areas: the applications status area, the applications menu area, and the applications display area. Figure 8 shows the screen layout. All three areas remain on-screen at all times.

The applications status area located at the top of the screen provides system information to the operator and is updated automatically by the system. No user interaction is allowed with this area.

The applications menu area along the right hand side of the screen allows the operator to access functions, and is divided into five functional areas: housekeeping functions, mission functions, operations support, system control, and hidden functions. The first four areas are used to initiate new functions. Hidden functions are those which have been temporarily suspended and hidden from view. Reactivation occurs by selecting the function name in the hidden function area. The menu area is accessible to the operator at all times and can be activated by either the mouse or the keyboard.

The applications display area is used to display all active function windows. It operates in two modes. In the normal mode, one active function window fills the entire area. In the split screen mode, one active function window and one suspended function window share the working area. The operator can switch back and forth between the two windows as needed.

2. Geographic Screen

The TDP R10 geographic (geo) screen operates in a manner similar to the alphanumeric screen, but includes some additional functional tools. This screen is divided into four areas: the geo menu bar, the geo status line, the geo title bar, and the geo map. Although this screen has a multicolor CRT, only the geo map uses the multicolor capabilities. The

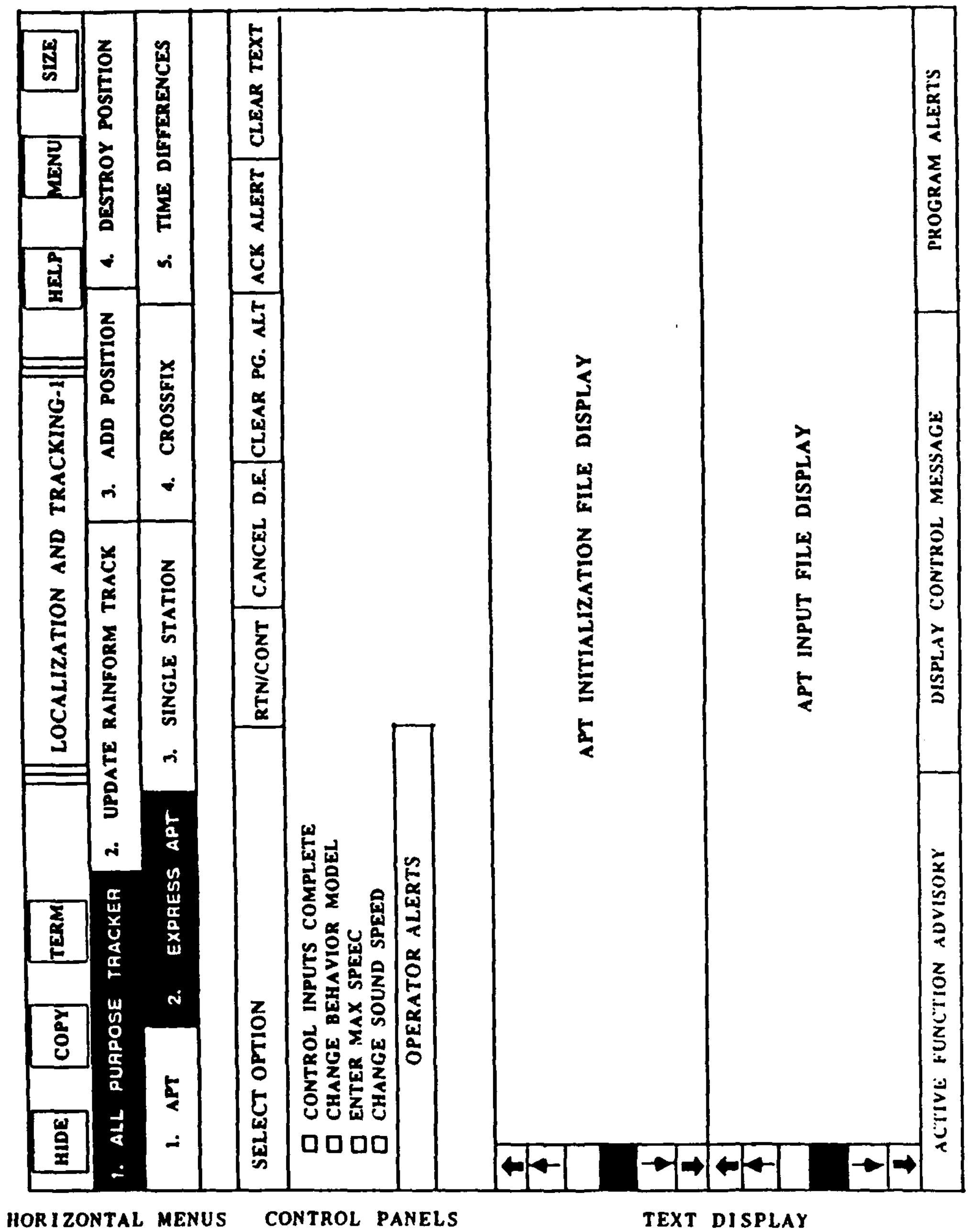

Figure 7. An Example of a TDP R10 Function Window Using Program Control and Data Display Objects. [Ref. 35]

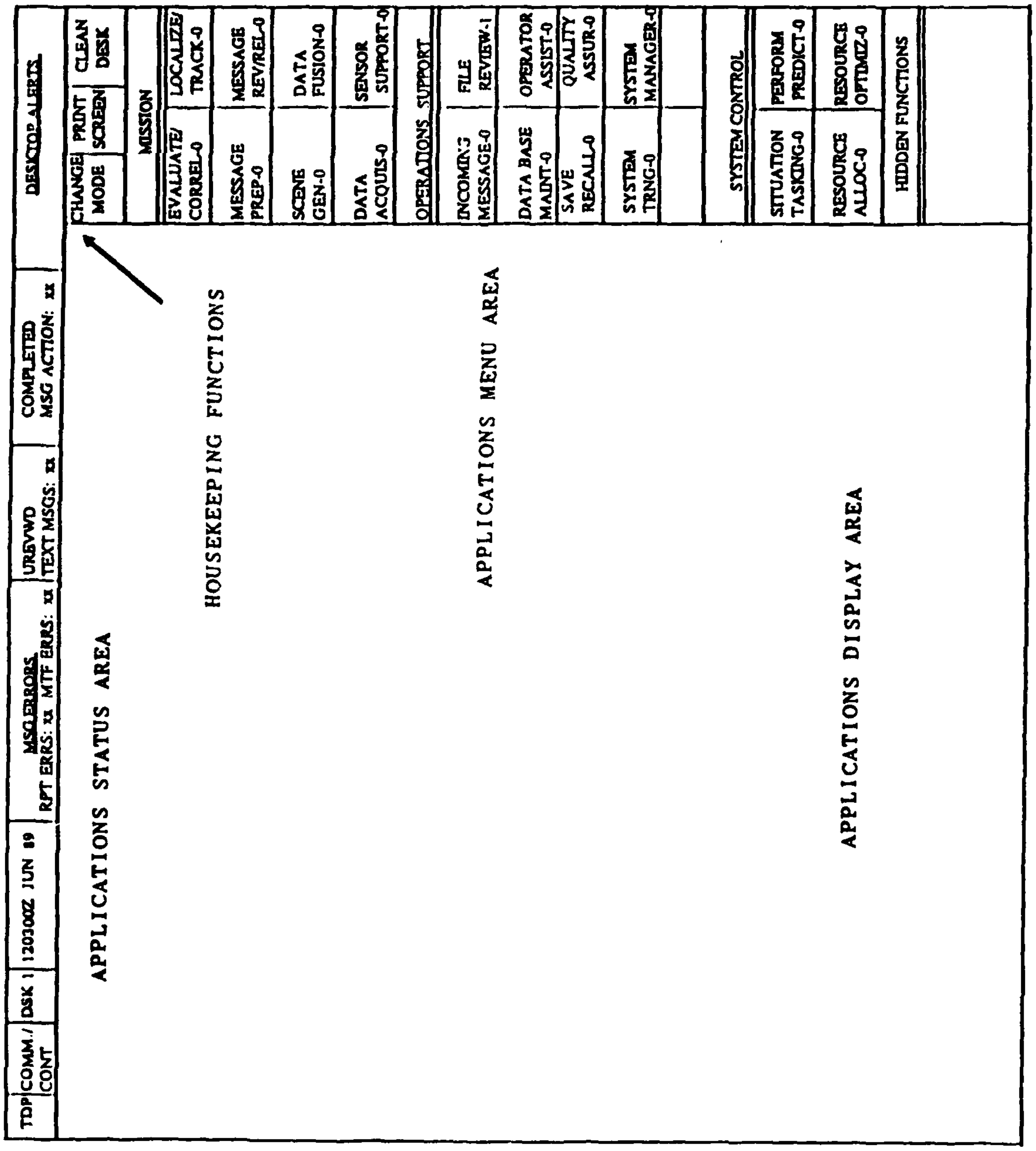

Figure 8. The TDP R10 Alphanumeric Screen Format. [Ref. 35]

remaining areas display information in the same manner as the alphanumeric screen. Figure 9 shows the geo screen format.

The geo menu lists categories of functions which are infrequently accessed by the operator. Selection of one of these categories causes a dialog box to appear on the geo map. In Figure 9, the dialog box which appears after selecting the geo display function is shown.

The geo status line is similar to the alphanumeric screen status area except that a operator input area is provided for some functions. Status information is updated continuously by the system.

The title bar area (so named because of its design in earlier versions) provides access to graphical display tools referred to as gadgets. Of note are the zoom in and zoom out gadgets which can magnify and reduce any area of the geo map.

The geo map covers most of the screen and is of the solid landfill type. The remaining background area on the map represents ocean. The operator has many options available for modifying the display based on current needs, including:

1. Map projection type
2. Modes of target data displayed
3. Types of arrays displayed
4. Coastline displayed
5. Bottom contours displayed
6. Map gridlines displayed
7. Color of background, coastline, contours, and gridlines.

Displayed targets are either red or green depending on the status (i.e., red for threat, green for friendly). The operator controls which targets are displayed from the alphanumeric screen.

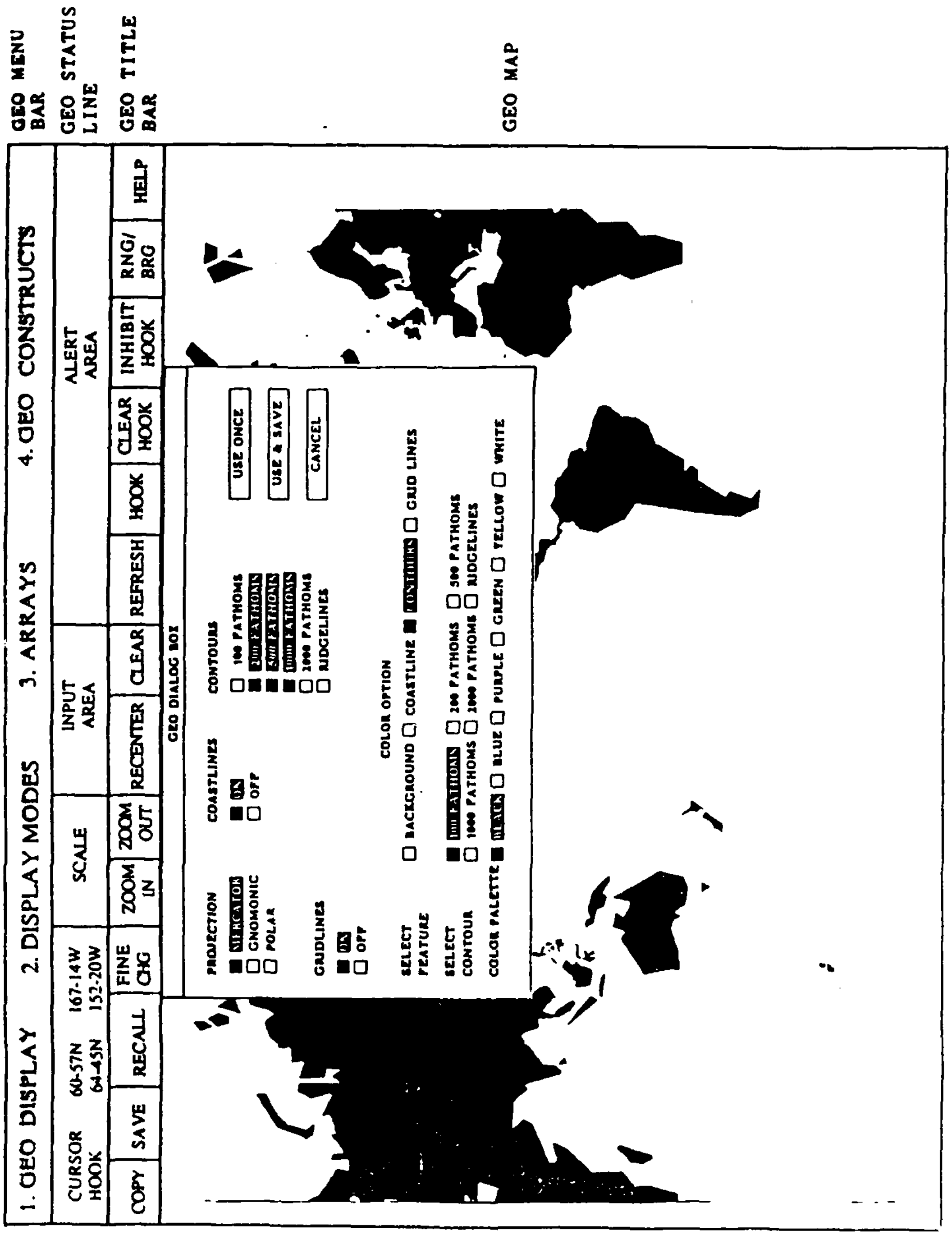

Figure 9. The TDP R10 Geo Screen Format. [Ref. 35]

IV. COLOR RECOMMENDATIONS FOR TDP R10

A. APPLICATION OF DESIGN GUIDELINES

Design guidelines such as those developed for static color CRT display formats tend to be very general in nature. However, when designing a specific system these general guidelines may need to be tailored to the requirements of that system. The system's tasks, users, operating conditions, and other design features must be considered when tailoring general design guidelines into specific design rules. [Ref. 8:pp. 8-9]

Some guidelines can be applied directly. For instance, "Display formats should be designed and evaluated under the same ambient light as will be present in the operational environment" is specific enough to be a design rule. However, to "use color consistently" is too vague a statement to form a design rule without further clarification.

B. COLOR DESIGN RULES FOR TDP R10

Currently, only the map display area of TDP R10 geo screen utilizes the multicolor capabilities of its display. For a more effective display, color use in this area could be improved. In addition, the option exists to extend color use to the windows of both the geo screen and the alphanumeric screen. Considerable effort has gone into designing the windows in a clear, consistent, and easy to read format. However, color could be added to improve the appearance and to assist the operator.

The general guideline that a monochrome display format should be designed before color is added has been satisfied for both the map displav area and the windows. Very little of the map area has been changed from the design used for the monochrome TDP R10. Window areas for TDP R10 geo and alphanumeric screens so far have been formatted in monochrome only. The map display area and the windows will be discussed separately.

1. Map Display Area

The current design of the map display area allows the operator to change the color of all items except targets, which are always either red or green. This results in three problems.

First, some color combinations such as green objects displayed on a white background are illegible. Legibility would be improved by limiting the background display colors to no more than two options: a middle gray and a black. This would satisfy three of the general guidelines: (1) that the display background be achromatic, (2) that the raster luminance levels be set at either middle gray or black, and (3) that the operator be allowed to adjust the raster luminance to ambient light conditions. However, this change will not totally relieve designers of considering ambient light during design.

The remaining color palette could be chosen from the seven sets listed in Appendix A. If a set of seven colors is considered insufficient for the items which must be displayed on the ocean background (e.g., targets, arrays, etc.), a further refinement might be to limit the coastline choices to some less discriminable colors not already included in the set chosen.

The second problem with the current design is that an operator who is working with several targets classified as threats in a small geographic area will have difficulty distinguishing one target from another since they will all be displayed in red. This limitation was imposed in order to meet U.S. Navy color coding standards (e.g., red for threats, green for friendly areas or objects, etc.). However, a standard practice at IUSS facilities where manual target plotting is done on paper charts is to allow the OTA to choose any available colored pencil to plot a specific target. A small legend on the chart lists which targets are drawn with which color. This practice could be extended to the TDP R10.

The third problem results from the fact that the cursor symbol is displayed in black. On a black or very dark background the cursor disappears. Since moving the cursor is the primary method to select functions and operate the graphical gadgets, not being able to locate it can cause considerable problems. Cursor color should be automatically linked to display background color. On the black background the cursor should be white (or light gray) and on the middle gray background it should be black.

There may be some concern that these recommendations allow too much user selection. Supervisors may worry that operators will spend too much time experimenting with the system's options. This may well be true when the system is initially used, but it is also equally likely that operators will settle on a set of options they prefer to work with and make changes only when needed. The discussion on human color vision deficiencies has already pointed out reasons why user selection should be encouraged, but an additional reason exists. The operational situation changes from moment to moment. At one time the operator may be working within a small area of the ocean where there is a need to distinguish each bottom contour by color code. Later, a larger area of ocean may be viewed and all contours can be in the same color or not displayed at all.

Recommendations included here for the map display do not include specifications as to how large the alphanumerics, symbols, and lines should be. The zoom function on the TDP R10 negates the need to follow the character size guideline. The operator can always magnify the area being viewed if the items are not legible.

The recommendations listed above can be implemented at the operator level be altering the geo display dialog box. A proposed example is provided in Figure 10.

2. TDP R10 Windows

The use of windowing for the TDP R10 partitions the display screen into distinct functional areas. Color could also be applied to provide further distinction between the areas and to improve readability. However, this is a case where overuse of color can cause more problems than it solves. The entire display should not be color coded, only portions of the display. The remaining areas should remain as currently designed, using black letters on a light gray background.

One way to use color for the window areas would be to use very pale colors as the display background for portions of the window while keeping all lettering in black. For example, the horizontal menu for a window could be pale blue while below it the control panels could be colored pale yellow (see Figure 7). A second way to highlight windows would be to use narrow bands of color around each window to set the windows off from each other.

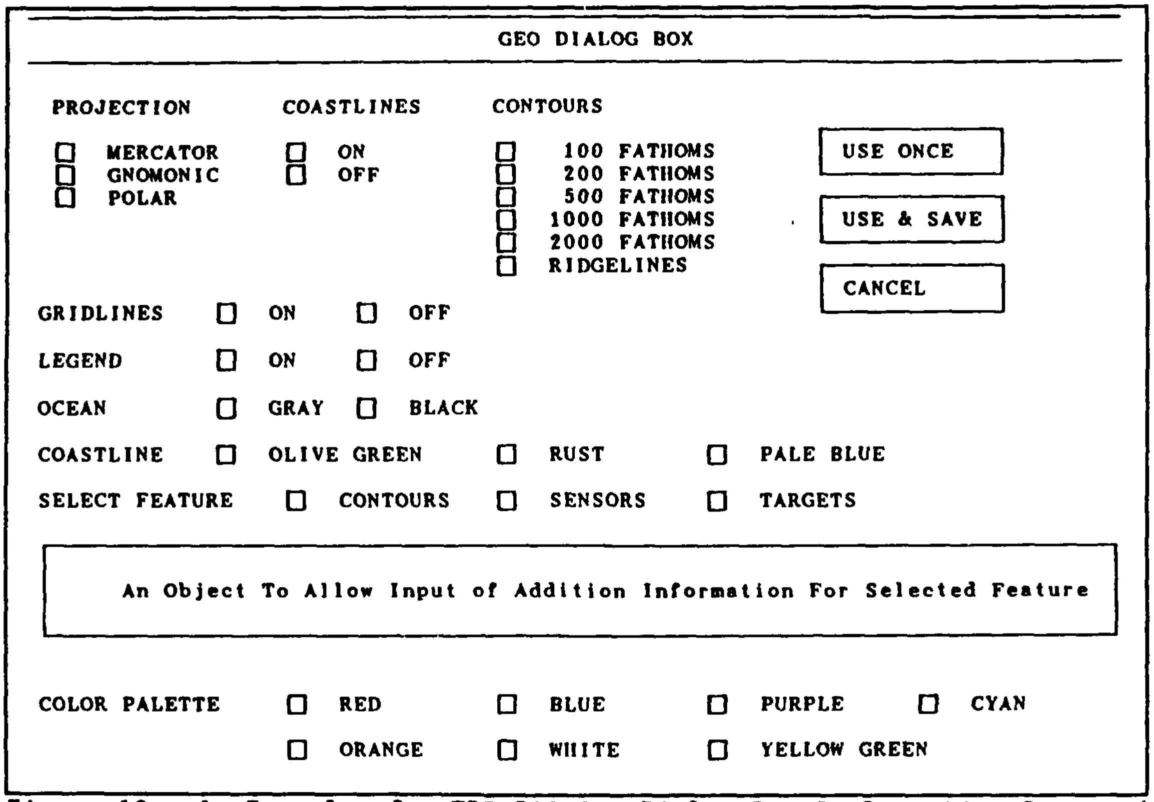

Figure 10. An Example of a TDP R10 Geo Dialog Box Implementing Proposed Recommendations.

The choice of which specific colors to use is not critical except that system alerts should always be displayed in red. Further, it is not essential that user selection of colors be provided. What is important is that consistent color coding must be used. Each window should be color coded based on the toolset defined for TDP R10. For example, program control objects, data entry objects, and data display objects each could be assigned a specific color. The color code could also be extended to lower level objects such as horizontal menus and dialog boxes. How low a level to code should be determined through test and evaluation.

Regardless of the level chosen, the color codes must be applied consistently on both the alphanumeric and geo displays.

In selecting the colors to use, adjacent windows should be displayed in colors with maximum contrast. For instance, blue and red would provide better contrast than blue and cyan. The color sets in Appendix A can be used to select a palette for color coding windows, as well as for the map display.

In addition to using color to improve readability, color can be used to link information between the alphanumeric screen and the geo screen. Data files viewed on the alphanumeric screen contain multiple entries relating to the targets which can be displayed on the geo map. A small colored dot placed next to all entries which refer to a given target would help the operator relate all information on that target. The current color setting of the target on the geo display will determine the dot color.

The recommendation that window lettering remain black gives some latitude to the character size guideline. However, if that guideline were met it would ensure character legibility.

3. General Design Rules

Three of the general guidelines for the use of color in static displays are specific enough for direct application to the TDP R10 geo display and alphanumeric display.

1. Display formats should be designed and evaluated under the same ambient light as will be present in the operational environment.

2. Choose a CRT with as large a dot matrix as possible (at least 5x7) for the best resolution.

3. The work routine associated with the display should include breaks which allow the operator to move around.

V. TDP R10 TEST AND EVALUATION CONSIDERATIONS

A. BACKGROUND

The development and application of guidelines is a first step in the design process, but not the last.

> The result of guidelines application will be a design for user interface software that may incorporate many good recommendations. However, even the most careful design will require testing with actual users in order to confirm the value of good features and discover what bad features may have been overlooked. Thus prototype testing must follow initial design, followed in turn by possible redesign and operational testing. [Ref. 8:p. 10]

Thus design is considered to be an iterative process. Gould and Lewis consider iterative design one of three principles for designing user interface systems. Early focus on users and tasks and empirical measurement are the other two. Their reasoning is that relatively little is known about human thought processes. Without user inputs and testing of the system with its expected users, many design problems will go undetected until the final product is operational. At that point, changes will be both expensive and difficult. [Ref. 36:pp. 300-311]

Recognizing the iterative process of designing an effective OMI, the developers of the TDP have utilized rapid prototyping as a primary development tool. A sequence of prototypes (demonstrators) which are not fully functional systems are being used to test and evaluate design alternatives. This allows designers and users a chance to see, operate, and evaluate the proposed system prior to final development.

In July and August 1987, TDP prototypes were provided to fleet users. They were given training on how to use the prototype and allowed to operate it. Users were then asked to complete a questionnaire evaluating the system and providing comments or suggestions. This feedback was used to solidify requirements, and to simplify and standardize the OMI.

Although costly, the use of rapid prototyping during TDP development has resulted in many advantages. Several design flaws have already been detected and corrected. Further, the fleet users have provided many original ideas for improvements. A side benefit of fleet involvement has been laying a foundation for favorable user acceptance of the final system. Most users were more interested in how soon the system would be available than in its design flaws.

The TDP R10 prototypes developed to date are not fully functional systems. Since the final TDP R10 design has not been determined, it is not presently possible to recommend a specific set of evaluation procedures. However, general guidelines are provided here for consideration when the test plan is developed.

The use of self-report techniques (e.g., questionnaires, interviews, and surveys) can provide valuable and unique information not determined using other evaluation techniques. However, special problems exist with these techniques which can bias the results. The data analyst must be aware of these problems when planning and conducting a survey. These problems are related to the fact that self-report data are subjective, no matter how objective the respondent tries to be.

This problem may be minimized by combining self-report techniques with performance testing. Users who take part in a performance test based on quantifiable measures of effectiveness could also participate in a survey about the system either during or immediately following the test [Ref. 36:p. 306-308]. Like self-report techniques, performance testing is also limited in what it can determine. It cannot measure cognitive processes such as attitudes, opinions, or perceptions [Ref. 37:p. 333].

The results from both self-report techniques and performance testing can then be analyzed alone or in combination. The advantage of combining results is that information can be determined which could not be determined by analyzing the data separately.

B. SELF-REPORT TECHNIQUES

The following discussion is adapted from Meister [Ref. 37: pp. 353-397]. A questionnaire is a written list of questions which require some form of written response. An interview is similar, but is conducted verbally and tends to be less structured. A survey is the completion of many interviews or questionnaires by a representative sample of the population. All three techniques are intended to gather information about attitudes, intentions, perceptions, or knowledge.

The decision to use an interview instead of a questionnaire depends on several factors. Generally, interviews require more time and money, require a trained interviewer, and may yield biased results due to loss of participant anonymity and/or influence of the interviewer. Questionnaires are easier to administer, once they have been prepared, and usually provide data that are easier to analyze.

The steps in completing a survey are generally the same for both questionnaire and interview formats. The steps include:

1. Decide what information is needed.
2. Determine the sample population and size.
3. Decide which data analysis techniques to use.
4. Search for existing questions on the subject.
5. Draft or revise new questions.
6. Format the entire list of questions.
7. Pretest the questions.
8. Revise them as needed.
9. Prepare administrative instructions.
10. Conduct the survey.
11. Analyze the results.
12. Report the results.

The first step in the most important. The information needed from the survey determines what questions to ask and the characteristics of the sample population. By considering this at the start of survey design, there is a better chance that the results will include all information that is needed without asking unnecessary questions which have no use or meaning.

After deciding what information is needed from the survey, the researcher must set about formulating the questionnaire or interview questions. The way in which questions are posed has a significant effect on the validity of the responses and on the types of data analysis techniques which can be used.

Validity is a measure of how well the question results in the intended answer. Four factors affect validity, i.e., are related to response error: memory, motivation, communication, and knowledge [Ref. 38:pp. 17-19]. Respondents may not give true answers because they have forgotten, because they fear to respond, because the question is confusing, or because they simply do not know the answer. Careful wording of questions and the use of specific types of questions can alleviate most of these errors.

In general, all questions should be grammatically and factually correct and as clear as possible. The respondent should not have to make assumptions about what is intended. The person (i.e., first, second, or third) in which the question is asked should be understood, along with the point of view the respondent should take. Each question should ask about only one topic. Compound questions which ask for a single opinion about multiple topics may result in confusion and invalid results.

Questions should normally not be loaded or leading (i.e., should not indicate which response to choose). Loading can occur if a reason for choosing a response is given or the preference of an influential group or person is stated. Leading occurs when the question is stated in such a way that a certain tone is established.

The order in which questions are presented can have the same affect as loading or leading. This may be intentional if the questionnaire designer wants to establish a frame of mind. Funneling, a technique where first general then specific questions are asked about a topic, can produce more valid results by clarifying the question's meaning.

Many types of question formats are available and all have advantages and disadvantages. The choice of which to use depends on the information desired and the way the data are to be analyzed. The same types of questions can be used for the entire questionnaire or interview format, or different types can be combined.

1. Open-Ended Items

With open-ended questions, each participant is asked to discuss, describe, or comment on an item. This is the easiest type of question to ask, but the hardest to analyze since unique answers are possible from all respondents. Further, in a written survey, there is no chance to probe the respondents on issues brought up in their answers. Open ended questions are best used to pretest questions in order to determine the range of possible responses, prior to actual questionnaire preparation.

2. Multiple Choice Items

When a multiple choice format is used, the participant is asked a question and given a list of two or more response alternatives. True/false questions are a form of multiple choice. These questions are easy to complete, analyze, and administer. However, all possible responses must be known ahead of time or the results will be invalid. The participant should not have to make a forced choice among responses that may not include the preferred answer. This can be avoided by including a noncommittal response, such as "none of the above" or "other".

3. Rating Scale Items

Given a rating scale, either verbal, numeric, or graphic, the participant is asked to rate an item along that scale. Whichever type of scale is chosen, it should represent the continuum of possible responses with equally spaced intervals. Verbal modifiers should have small variability in meaning and have parallel wording. A more complete discussion of rating scales and examples from the literature are provided by Meister [Ref. 37:pp. 320-329, 381-385].

Rating scales provide both a direction and degree of response, are easy to analyze, and take little time to complete. Though more reliable than multiple choice items, they are more susceptible to errors than some other question types.

4. Ranking Items

It is often useful to allow the respondents to rank a list of items according to some dimension stated in the question. The ranking represents a relative ordering without allowing the degree of difference between items to be specified. Surveys that utilize ranking are easy to administer, score, and code, but tend to be less precise than those that use rating techniques.

5. Checklists

When a checklist format is used in a survey, participants are given a list a statements and asked to check all those that are appropriate. If numeric values (as are obtained with rating scales) are not necessary, this type of question can be useful and is easier to format.

6. Arrangement of Items

It may be important to know in what sequence the operator thinks the tasks should occur. In this case, each participant is presented with a list of events or steps and is asked to arrange them in order of occurrence. Given the difficulty in scoring and analyzing the results, this question type is usually limited to task analysis.

C. PERFORMANCE TESTING

One of the most common methods of evaluating performance is through the use of quantifiable measures of effectiveness (MOEs). These measures are objective in that they do not require subjective judgments to be made. Objective MOEs can be used to describe system performance or to compare one system to another system or to an external standard. For example, the effectiveness of alternative OMI designs can be evaluated by comparing user performance results for the various designs.

In the field of human behavior, only a relatively few generic measures are available for this purpose. The time it takes to complete a task and the number of times an event occurs can be recorded. The counting of events can be combined with a time interval to give event frequency. [Ref. 37:pp. 332-334]

The choice of the measure to use depends on the objective of the test. Based on that objective, a detailed MOE must be stated and a procedure specified to measure it. This can be a difficult process when testing hardware (or software) alone; the addition of the human in the system further complicates the situation. The following discussion points out some of the problems which may occur.

The performance being tested must involve some physical or overt occurrence which can be observed and measured. For example, a target detection task may involve detection of a signal, followed by analysis, classification, and the reporting of the signal's occurrence. Only the time between the signal's occurrence and the operator's report can be measured. Measuring the time required for the observer simply to detect the signal is not possible, since detection, analysis, and classification are cognitive activities that cannot be observed. [Ref. 37:pp. 333-335]

The context in which a measured event occurs must be clearly understood. For example, when counting errors it is not useful simply to know that an error has occurred. The reason it occurred is what must be determined. This means that a considerable amount of information about the occurrence must be known. The type of error, how critical it was, when it occurred, who made it, etc., must be known in order to correct design flaws which may have caused the error. [Ref. 37:pp. 336-339]

D. SAMPLE POPULATION AND SIZE

When conducting surveys or performance tests, a group of study participants must be identified. Determination of the population to use is based on the purpose of the procedure. For both survey and performance testing of a system, the participant population should be representative of the ultimate user population. In the case of the TDP, the user population consists of those OTAs who have been or will be designated TDP Displays Analysts. Since the TDP is closely related in design to the UCP R6, the user population could be extended to include UCP operators.

It is normally infeasible to test the entire population. Therefore a subset or sample of the population is tested. The sample size can affect the validity of the results: the smaller the sample, the less likely the results obtained from testing will reflect the true values. For example, if a survey of n users asks how many prefer a gray display background to a black background, the sample proportion who answer yes (denoted Y/n, the number of yes answers divided by n) will be an estimate of the true proportion P.

To see how much P may vary from Y/n, the binomial distribution can be used to construct a confidence interval for P with a stated confidence coefficient (usually chosen to be 90% or 95%). Table II gives the confidence intervals for selected values of Y/n and sample size. For instance, if 100 respondents are asked their preferences and 50 say they prefer a gray background (i.e., Y/n = 0.50), it can be said with 90% confidence that the true population proportion lies between 0.37 and 0.63.

TABLE II
EFFECT OF SAMPLE SIZE ON THE 90% AND 95% CONFIDENCE INTERVALS FOR THE TRUE POPULATION PROPORTION P USING THE BINOMIAL DISTRIBUTION AND THREE VALUES OF THE SAMPLE PROPORTION Y/n.

Y/n	SAMPLE SIZE	90% CONFIDENCE INTERVAL	95% CONFIDENCE INTERVAL
0.25	10	0.02 - 0.69	0.04 - 0.61
	100	0.14 - 0.38	0.17 - 0.35
	1000	0.22 - 0.28	0.23 - 0.28
0.50	10	0.13 - 0.87	0.18 - 0.82
	100	0.37 - 0.63	0.40 - 0.60
	1000	0.46 - 0.54	0.47 - 0.53
0.75	10	0.31 - 0.98	0.39 - 0.96
	100	0.62 - 0.86	0.65 - 0.83
	1000	0.72 - 0.78	0.72 - 0.77

E. DATA ANALYSIS

A data analysis plan should be formulated early in the survey and performance testing process. The result of data analysis is the information needed to make a decision. If consideration is given to data analysis early on, the results are more likely to provide the data needed in a format that can easily be used with the most appropriate analysis techniques. This does not mean that data from surveys or performance testing cannot be analyzed without a prior plan or using techniques not planned for. However, a detailed analysis plan assures that the results will be useful.

The data analysis techniques to use depend on the information that is needed from the analysis. It may be enough to describe the numeric results or to compare them to a predetermined standard. For these purposes descriptive statistics, (e.g., the mean, median, variance, range, etc.) may be sufficient. For example, in the case of the TDP R10, determination that at least 95% of messages were released error free will strongly indicated that the system will be adequate for that task.

If other factors are thought to affect the value of a particular variable, analysis techniques can be used to determine if relationships exist between these factors or variables. Factors that may often influence results include differences in test conditions, differences among the test subjects, and differences that are revealed by survey questions. For example, if the TDP R10 were tested with two alternative color display designs, a lower error rate for message processing might be associated with one of the designs.

The following is a brief discussion of some statistical analysis techniques available to study such possible relationships among variables.

1. Regression

One way to analyze the relationship between two or more variables is to attempt to define a mathematical equation which relates one variable to another. In regression a dependent variable is estimated from one or more independent variables using an equation called a regression equation. If the true relationship between the variables is expressed by the regression equation, the value of the independent variable(s) can be used to predict the value of the dependent variable. A correlation coefficient is often calculated during a regression procedure. The coefficient measures how well the regression equation represents the true relationship between the variables.

An example of using regression for test and evaluation would be to analyze mean values of an MOE as a function of independent variables. Jacobsen and Neri [Ref. 13] studied recognition time for color sets of up to seven colors (results presented in Figure 2). They determined that the relationship between recognition time and set size was not significantly different from a line with a slope of zero [Ref. 13:p. 9]. In this example, the dependent variable or MOE was reaction time and the independent variable was a test condition, set size.

2. Analysis of Variance

Analysis of variance (ANOVA) is a special form of regression. This technique is used to study whether a specific condition or factor has an effect on the mean values for some variable. The observations can be classified by one or two factors at the same time.

ANOVA was used by d'Ydewalle and others [Ref. 26] to determine the factors that influence performance in a signal detection task. They showed that signal detection is influenced by signal strength and by whether the operators use their preferred color combination [Ref. 26:pp. 298-299]. The MOE for this study was the number of target detections while the two influencing factors were a test condition, signal strength, and a response on a survey question, preferred color combination.

3. Contingency Tables

A contingency table is formed by classifying observations (the results of performance tests or survey questions) according to two factors (e.g., color preference and response time). Each factor may have two or more categories (color preference may be red, blue, green, etc.). An r x c contingency table has r categories (or rows) for one factor and c categories (or columns) representing those of the other. The intersection of each row and column forms a cell containing the quantity of observations which fall into both that row category and that column category. Once the contingency table has been formed, a chi-square test is used to determine whether there is a relationship between the two factors.

Contingency tables can be used in many ways for analyzing survey and performance test data. For example, the two factors could be two different survey questions. Alternately, the row factor could be the population samples which were tested under different conditions while the column factor could be responses on a survey question. This would show whether there was any relationship between survey responses and test conditions. A contingency table analysis could be used if a survey was taken before before and after the operators took part in a performance test. The analysis would help determine whether the experience gained using the system influences operator responses.

APPENDIX A: RECOMMENDED COLOR SETS

The minimum (delta)E* per set, the luminances (Cd/m^2) and chromaticity coordinates (C.I.E. 1931) of seven color sets recommended for use in designing color CRT displays. Data adapted from Neri and others (1985, pp. 4, A4-A5).

SET	MIN SET (delta)E*	COLOR	Cd/m^2	x	y
1	49.1	Dark Blue	17.0	.15	.07
		Purple	19.7	.27	.14
		Red	56.3	.61	.34
		Aqua	85.6	.25	.36
		Pink	106.3	.35	.33
		Yellow	189.4	.42	.46
		White	239.7	.29	.30
2	47.4	Blue	28.9	.15	.07
		Red	63.3	.54	.32
		Purple	68.4	.27	.15
		Cyan	81.0	.21	.26
		Orange	101.8	.50	.41
		Yellow Green	104.7	.30	.54
		White	229.8	.28	.31
3	89.5	Dark Green	7.4	.31	.57
		Medium Blue	9.3	.17	.12
		Red	15.7	.49	.27
		Tan	41.8	.38	.36
		Orange	81.8	.54	.39
		White	183.6	.29	.31
		Yellow	190.5	.42	.47

SET	MIN SET (delta)E*	COLOR	Cd/m^2	x	y
4	35.6	Green	42.4	.24	.35
		Blue	59.2	.17	.12
		Red	62.5	.53	.31
		Amber	62.7	.53	.38
		Gray	68.9	.28	.28
		Yellow	75.0	.46	.43
		Magenta	85.1	.25	.18
5	34.3	Blue	27.1	.17	.09
		Red	55.1	.54	.33
		Orange	59.9	.51	.40
		Yellow Green	62.1	.30	.55
		Purple	68.6	.27	.15
		Gray	68.8	.28	.27
		Cyan	73.2	.19	.18
6	33.9	Red	63.9	.52	.31
		Blue	80.9	.17	.13
		Amber	92.0	.53	.38
		Magenta	132.9	.25	.16
		Yellow	140.8	.46	.44
		White	231.4	.29	.31
		Green	235.7	.25	.38
7	24.3	Medium Purple	10.3	.61	.32
		Dark Yellow Green	14.5	.31	.53
		Red	16.7	.61	.32
		Gray Red	31.0	.39	.31
		Pale Purple Blue	31.8	.23	.21
		Pale Orange Yellow	78.0	.33	.34
		Orange Yellow	82.8	.45	.44

APPENDIX B: ACRONYMS

CIE	Commission Internationale de l'Eclairage (International Commission on Illumination)
CRT	cathode ray tube (also known as VDU or VDT)
DTIC	Defense Technical Information Center
IUSS	Integrated Undersea Surveillance System
JINTACCS	Joint Interoperability of Tactical Command and Control Systems (message format which replaces RAINFORM)
K	kilobytes
MOE	measure of effectiveness
NOSC	Naval Ocean Systems Center
OMI	operator machine interface (also known as MMI or HCI)
OTA	Ocean Systems Technician Analyst
OWO	Ocean Systems Watch Officer
RAINFORM	RAINBOW message format
SPAWARSYSCOM	Space and Naval Warfare Systems Command
TDP	Target Data Processor
UCP	Universal Communications Processor

LIST OF REFERENCES

1. Shneiderman, B., Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison-Wesley Publishing Co., 1987.

2. Sanders, M.S., and McCormick, E.J., Human Factors in Engineering and Design, McGraw-Hill Book Co., 1987.

3. Office of Naval Research Report CR213-136-2F, Color Display Design Guide, by M.J. Krebs and others, October 1978.

4. Naval Postgraduate School Report 55-80-013, Application of Color Coding in Tactical Display S-3A, by D.E. Neil, April 1980.

5. Naval Air Development Center Report 86011-60, The Development and Evaluation of Color Systems for Airborne Applications: Fundamental Visual, Perceptual, and Display Systems Considerations, by R.M. Merrifield and L.D. Silverstein, February 1986.

6. Miller, G., "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information," Psychological Review, v. 63, pp. 81-97, 1956.

7. Navy Personnel Research and Development Center Report 84-51, Human Engineering Data Base for Design and Selection of Cathode Ray Tube and Other Display Systems, by D. Meister, July 1984.

8. Electronic Systems Division, Air Force Systems Command Report 86-278, Guidelines for Designing User Interface Software, by S.L. Smith and J.N. Mosier, August 1986.

9. Weitzman, D.O., "Color Coding Re-Viewed," Proceedings of the Human Factors Society 29th Annual Meeting, 1985.

10. Fogus, R.H., Perception, McGraw-Hill Book Co., 1966.

11. Hanes, R.M., and Rhoades, M.V., "Color Identification as a Function of Extended Practice," Journal of the Optical Society of America, v. 49, pp. 1060-1064, 1959.

12. Shontz, W.D., Trumm, G.A., and Williams, L.G., "Color Coding for Information Location," Human Factors, v. 13, pp. 237-246, 1971.

13. Naval Submarine Medical Research Laboratory Report 1042, *The Effect of Set Size on Color Recognition*, by A.R. Jacobsen and D.F. Neri, 25 January 1985.

14. Naval Submarine Medical Research Laboratory Report 1041, *The Effect of Set Size on Color Matching*, by S.M. Luria, D.F. Neri, and A.R. Jacobsen, 14 January 1985.

15. Rossotti, H., *Colour*, Princeton University Press, 1983.

16. Carter, R.C., and Carter, E.C., "High-contrast Sets of Colors," *Applied Optics*, v. 21, pp. 2936-2939, 15 August 1982.

17. Carter, E.C., and Carter, R.C., "Color and Conspicuousness," *Journal of the Optical Society of America*, v. 71(6), 723-729, 1981.

18. Judd, D.B., and Wyszecki, G., *Color in Business, Science and Industry*, John Wiley and Sons, 1975.

19. Naval Submarine Medical Research Laboratory Report 1068, *An Evaluation of Color Sets for CRT Displays*, by D.F. Neri, A.R. Jacobsen, and S.M. Luria, 31 December 1985.

20. Naval Submarine Medical Research Laboratory Report 1027, *Visibility of Various Target-Background Color Combinations Under Different Chromatic Ambient Illuminations*, by D.F. Neri, S.M. Luria, and D.A. Kobus, 24 August 1984.

21. Naval Submarine Medical Research Laboratory Report 1070, *The Effect of Background Luminance and Ambient Illumination on Color Discrimination in a CRT Display*, by A.R. Jacobsen, 16 January 1986.

22. Naval Submarine Medical Research Laboratory Report 1051, *The Effect of Background Luminance on Color-Coding*, by A.R. Jacobsen, 15 July 1985.

23. Rothwell, N.V., *Understanding Genetics*, Oxford University Press, 1979.

24. Department of the Navy, *Manual of the Medical Department*, Chapter 15, change 95, 25 November 1980.

25. Department of the Navy, *Flight Surgeons Manual*, prepared by BioTechnology, Inc., Arlington, VA, 1968.

26. d'Ydewalle, G., Van Rensbergen, J., and Huys, J., "The Efficiency of Letter Perception in Function of Color Combinations: A Study of Video-Screen Colors," *Man-Computer Interaction Research MACINTER-I*, North-Holland, 1986.

27. Starr, S.J., "Effects of Video Display Terminals in a Business Office," Human Factors, v. 26(3), pp. 347-356, 1984.

28. Hagberg, M., and Sundelin, G., "Discomfort and Load on the Upper Trapezius Muscle When Operating a Wordprocessor," Ergonomics, v. 29(12), pp. 1637-1645, 1986.

29. Zwahlen, H.T., Hartmann, A.L., and Rangarajulu, S.L., "Video Display Operator's Eye Scanning, Comfort/Discomfort and Performance in a Hard Copy - Screen and a Split Screen Data Presentation," Proceedings of the Human Factors Society 28th Annual Meeting, 1984.

30. Mourant, R.R., Lakshmanan, R., and Chantadisai, R., "Visual Fatigue and Cathode Ray Tube Display Terminals," Human Factors, v. 23(5), pp. 529-540, 1981.

31. Hedman, L.R., and Briem, V., "Short-Term Changes in Eyestrain of VDU Users as a Function of Age," Human Factors, v. 26(3), pp. 357-370, 1984.

32. Matthews, M.L., and Mertins, K., "The Influence of Color on Visual Search and Subjective Discomfort Using CRT Displays," Proceedings of the Human Factors Society 31st Annual Meeting, pp. 1271-1275, 1987.

33. Personal communication between CDR L. Skowronek, PMW 180, SPAWARSYSCOM and the author, November 1988.

34. Personal communication between LCDR S. Deems, PMW 180-3, SPAWARSYSCOM and the author, May 1988.

35. TDP R10 Prototype Version 3.0 Function Definitions, Part 1, preliminary, prepared by Computer Sciences Corp., San Diego, CA, 1988.

36. Gould, J.D., and Lewis, C., "Designing for Usability: Key Principles and What Designers Think," Communications of the ACM, v. 28, pp. 300-311, 1985.

37. Meister, D., Behavioral Analysis and Measurement Methods, John Wiley and Sons, 1985.

38. Sudman, S., and Bradburn, N.M., Asking Questions, A Practical Guide to Questionnaire Design, Jossey-Bass Publishers, Inc., 1982.

INITIAL DISTRIBUTION LIST

1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Commander Space and Naval Warfare Systems Command Attn: J.R. Boyett, PMW 180-3 Navy Department Washington, D.C. 20363-5100	2
4.	Commander Naval Ocean Systems Center Code 723 San Diego, CA 92152-5000	1
5.	Commander Naval Ocean Systems Center Code 7203 San Diego, CA 92152-5000	1
6.	Commander Naval Ocean Systems Center Attn: Suzanne V. Bemis, Code 441 San Diego, CA 92152-5000	1
7.	Commander Naval Submarine Medical Research Laboratory Attn: LT David F. Neri Naval Submarine Base NL Groton, CT 06349-5900	1
8.	LT Claudia J. Schmidly 8390 Fern Leaf Court Springfield, VA 22153	2
9.	Naval Postgraduate School Attn: Judith Lind, Code 55Li Monterey, CA 93943	2

10. Naval Postgraduate School 2
Attn: LCDR Tom Mitchell, Code 55Ml
Monterey, CA 93943

11. Naval Weapons Center 1
Attn: Kim Frankenberger, Code 3152
China Lake, CA 93555

12. Naval Weapons Center 1
Attn: Patrick Croyle, Code 3152
China Lake, CA 93555

13. Naval Air Development Center 1
Attn: Michael Barnes, Code 6021
Warminster, PA 18974

14. Naval Personnel Research And Development Center 1
Attn: Jeff Grossman, Code 71
San Diego, CA 92152-6800

15. Naval Facility Whidbey Island 1
Attn: LCDR Sharon A. Deems
Oak Harbor, WA 98278-9400