N91-20709

Enhancing the Usability of CRT Displays in Test Flight Monitoring

Michael M. Granaas, and Victoria E. Sredinski
Human Factors Laboratory
The University of South Dakota
Vermillion, SD 57069

Enhancing the usability of Mission Control Center (MCC) CRT displays stands to improve the quality, productivity, and safety of flight-test research at the NASA Ames-Dryden Flight Research Facility. The research reported in this paper involves three experiments aimed at improving the usability of the CRT displays in the Ames-Dryden MCCs. The results of this research suggest that much can be done to assist the user, and improve the quality of flight research through the enhancement of current displays. The research reported has applications to a variety of flight data monitoring displays.

Introduction

In the years since World War II, the amount of data collected in flight-test research has increased from a handful of parameters to several hundred parameters (Granaas and Rhea, 1988). Also increasing is the amount of data available to the flight-test researcher in real-time (Moore, 1986). As technology improves, there is every reason to believe that the amount of data available to the researcher will continue to increase.

While this is generally good news for the flight-test researcher, it does not come without some cost. Specifically, increasing amounts of data have lead to cluttered screens and increased mental work load for the user. This in turn reduces the overall effectiveness of flight-test programs.

In order to minimize the negative impact of increasing amounts of data on the flight-test researcher, we have begun a program of research intent on determining appropriate and effective design criteria for CRT flight data displays. This paper reports the preliminary findings from three of these studies.

Experiment 1

Research has shown that reducing screen clutter and increasing information organization can lead to improved task performance (Ramsey and Atwood, 1979). This experiment was done to determine if some of the displays in the Dryden MCC could be reorganized to reduce screen clutter and improve user performance.

During the course of a test flight, MCC users will frequently monitor a set of flight parameters until they match a predetermined set of values that define a test point. Once this match occurs, the user takes some action until the matching ceases to occur. Often the task performed involves making a record of the time at which the aircraft reached the test point so that the data of interest can be retrieved at a later time. Since an aircraft is "on point" for only a few seconds, delays in recognizing that the aircraft is on point can be very costly in terms of the amount and quality of data collected.

This task was chosen because in one form or another, many users monitor one or more parameters. This task should also generalize to a number of tasks outside of the control room setting.

The current screen layout now requires that the user either memorize the three to four target values that define the test point, or scan multiple locations on the screen to determine what those values are. With only three values to monitor this is probably not too mentally taxing for the user. However, as the number parameters to be monitored increases, the mental work load of the user should increase to the point where user performance suffers.

This research utilized a modified display format to test the efficacy of modified displays in the MCC. Target and actual parameter values were placed in adjacent columns to reduce user scanning time. This placement also allowed for the elimination of some redundant parameter labels which helped reduce screen clutter.

Methods

Subjects. Sixty undergraduate students from a small midwestern university were recruited to act as subjects for this study. Subjects received class credit for their participation.

Apparatus. The displays were programmed in Microsoft-C on a CSS 286-A AT compatible computer running at ten megahertz. Displays were presented on a NEC Multisync II monitor in EGA mode. Reaction times were collected via a Microsoft Mouse and software that uses the system hardware timer to measure times with better than millisecond accuracy (Granaas, 1989).

Displays. Two display formats were developed for this study. The first replicated one of the actual alphanumeric displays currently used in the Dryden MCCs. The second was modified so that parameters being monitored appeared in adjacent columns of the display in the upper left quadrant, and redundant labels were eliminated. Each of these displays was tested with subjects monitoring three, five, or seven parameters.

Procedure. After receiving informed consent, subjects were seated at an experimental workstation with the display and mouse. Each subject participated in only one of the six possible display format by number of monitored parameters conditions. The subject was given instructions in both verbal and written form for the condition in which they participated. Subjects were given one practice trial followed by an opportunity to ask questions. After answering any questions, the experimenter left the subject to complete the experiment.

Each subject completed 20 trials as part of the experiment. The first five trials were later treated as practice trials due to the large number of missing data points in those trials.

In each trial subjects were instructed to press the left mouse button when the monitored parameters matched their target values, and again when one or more of the monitored parameters ceased to match its target value. Following each button press one of two tones was presented. One tone indicated a correct response and the other a false alarm. Thus, subjects had feedback to assist them in determining if they had responded correctly.

The computer recorded the reaction times and number of false alarms for both the matching and dematching tasks. So each trial had four data points associated with that trial: Reaction time for the matching task, reaction time for the dematching task, number of false alarms for the matching task and number of false alarms for the dematching task.

Results

For purposes of these preliminary analyses, the only data analyzed was the mean reaction time for each subject on the matching and dematching tasks. Each of these scores was analyzed independently using a 2(display format) by 3(parameters monitored) completely between groups analysis of variance (ANOVA).

Matching Task. The display format by number of parameters monitored interaction was significant ($F(2,54) = 11.53$, $p < .0001$). The main effects for display format and number of parameters were also significant ($F(1,54) = 22.02$, $p < .0001$ and $F(2,54) = 5.08$, $p < .01$).

Table 1

Mean Reaction Times for Matching Task for Old and New Display Formats by Number of Parameters Monitored

Number of Parameters	Format: Old	New
3	1.26 (0.433)	1.21 (0.282)
5	2.87 (1.032)	1.32 (0.437)
7	3.02 (1.394)	1.75 (0.516)

An examination of the means in Table 1 indicates that there is little difference between the two formats when only three parameters are monitored. As the number of parameters increases, reaction times for the unorganized display climbed sharply, while those for the organized display climbed only modestly.

Dematching Task. This analysis found no significant effects.

Discussion

This study shows that display organization is an important component in display design. Reorganizing of poorly organized displays can significantly improve performance. And, reorganization can also reduce the effects of further increases in the difficulty of the task being performed. However, this improvement in performance only occurs when that task is sufficiently difficult that the user cannot effectively compensate for the increased workload.

EXPERIMENT 2

This experiment was designed to look at the effects of different screen colors on task performance. Current guidelines suggest that there are no differences in performance as a function of screen color as long as the foreground/background colors selected are of high contrast (Mitchell, Stewart, Bocast, and Murphy, 1982; NASA, 1989). However, such research is open to the criticism that the tasks performed are relatively simple and of short duration (less than 20 minutes), or do not reflect real tasks. What happens in real task settings over a period of time seems to be unknown.

Methods

Subjects. Fifty-six subjects from the same population as Experiment 1 were recruited to participate in this study. Subjects received course credit for their participation.

Apparatus. The computer hardware used is the same as that used in Experiment 1. The programs from Experiment 1 were modified to meet the needs of this study. The modifications are discussed below.

Displays. The reformatted, seven parameter display was taken from Experiment 1 for use in this study. The programs were modified so that the program would run in one of four foreground/background display modes: White on black, amber on black, green on black, and black on white.

Procedure. The procedure for this study followed that of Experiment 1, except that five practice trials were given, and only 12 trials were used for collecting data. This reduced the number of trials from 20 to 17 total. This task required approximately 50 minutes to complete, and according to subjects from experiment 1, was fairly difficult. Subjects participated in only one of the display mode conditions.

Results

Mean reaction times for the matching and dematching tasks were used in a one-way analysis of variance (ANOVA) to test for differences between the screen color conditions. No significant effects were found for either the matching or dematching tasks.

Discussion

In a relatively difficult task under normal lighting conditions, screen foreground/background color did not affect task performance. Since all color combinations were of relatively high contrast, it can be argued that high contrast between colors is more important than which colors are used.

Since the number of foreground/background combinations used in this study were limited, some caution must be included as part of the recommendation that high contrast is all that is important. The color combinations selected for this study were those that are typically available on commercial CRT displays.

Blue has been suggested as a color to avoid as a character color due to the human eye's inability to focus precisely on wave lengths at either end of the visible spectrum. All of the colors tested can be focused on with a high degree of precision by the human eye. Thus, these results may not be duplicated with a high contrast blue characters on black background display.

EXPERIMENT 3

This experiment was designed to examine the influence of color highlighting on the matching and dematching tasks used in Experiments 1 and 2. To date, the results using color highlighting have been mixed. Many studies show that color highlighting does little or nothing to improve performance. However, a smaller number of studies suggest that under some conditions color highlighting can improve performance (Christ, 1975).

We do know that color highlighting that does not consistently provide useful information does not help, and may actually detract from, user performance (Christ, 1975; Fisher and Ten, 1989). This indicates that color use needs to be consistent if it is to be of any value at all (Schneiderman, 1987).

A limitation of color highlighting, and other highlighting research has been the task difficulty level. Frequently research in this area uses tasks that the subject can already perform with relative ease. Thus, any potential benefits of highlighting are masked due to ceiling effects (Christ, 1977).

In the research reported here, a sufficiently difficult task was used so that any positive effects of highlighting could be detected. In addition, some subjects received an extra task to increase their mental workload. This was done to insure that their mental workload was sufficiently taxing so that the benefits of highlighting could be detected if they exist.

Methods

Subjects. Twenty-four subjects from the same population used in Experiments 1 and 2 were recruited for the current study. Subjects received class credit for their participation.

Apparatus. The computer hardware is the same as used in Experiments 1 and 2. The program used in Experiment 2 was modified to meet the needs of this study. These modifications are described below.

Displays. The base display for this research was the display used in Experiment 2 with white characters on a black background. The program for this display was modified to include different forms of color highlighting to assist the user in the matching and dematching tasks. Four highlighting conditions were used.

The first was a no highlighting condition. In this condition, subjects had to scan the actual and target values for all seven parameters after each screen refresh to determine if all seven matched. Thus the subject was required to make seven comparisons before being able to respond to the matched condition.

The second highlighting condition was labeled individual highlighting. In this condition, each parameter's label, actual, and target values were highlighted in yellow when that parameter matched it's target value. In this condition, the subject was required to make seven yes/no decisions before being able to respond positively to the matched condition.

The third highlighting condition was labeled group highlighting. In this condition, the seven parameter's labels, actual and target values were highlighted in green as a group only when all seven parameters matched their target values. This reduced the matching task to a simple signal detection task. The subject needed to respond only when the highlighting occurred.

The fourth highlighting condition was labeled combined highlighting. This condition combined individual and group highlighting.

Procedure. In a fashion similar to Experiments 1 and 2, subjects received informed consent and general instructions. The highlighting conditions were presented to subjects in one of four presentation orders to balance practice effects. Subjects were presented with five practice trials prior to the first highlighting condition to familiarize them the task and their first highlighting condition. Prior to each subsequent highlighting conditions, subjects received two practice trials to familiarize them with that highlighting condition. Subjects completed a block of ten trials in each highlighting condition.

In addition to the matching and dematching tasks, half of the subjects also performed a safety task. For the safety task subjects were expected to monitor two additional parameters on the upper right portion of the display. When both of these values exceeded predefined limits the subject was to respond by pressing the right mouse button. This condition occurred twice during each block of ten trials and reaction times were recorded.

Results

For each of the highlighting conditions, mean reaction times for the matching task, the dematching task, and, where appropriate, the safety task were calculated for each subject. The matching and dematching reaction times were analyzed using independent 2(number of tasks) by 4(highlighting conditions) split plot ANOVAs where the number of tasks was a between groups factor, and the highlighting conditions were repeated across subjects. The reaction time data for the safety task was analyzed with a repeated measures ANOVA for highlighting conditions.

Matching Task. The data for the matching task showed a significant main effect for highlighting condition ($F(3,66) = 49.92$, $p < .0001$). The difficulty main effect and the difficulty by highlighting condition interaction were not significant.

Table 2

Mean Reaction Times for Matching Task by Highlighting Condition

Type of Highlighting	Mean RT (SD)
none	2.10 (.675)
individual	1.26 (.284)
group	0.93 (.192)
combined	1.05 (.305)

Dematching Task. There were no significant effects for the dematching task.

Safety Task. Highlighting condition failed to have a significant effect on response time for the safety task.

Discussion

These results indicate that color highlighting can provide a display user with information that improves performance in some cases. Taken with the findings for the dematching and safety tasks, this work suggests that the way in which highlighting is used is an important consideration. Highlighting needs to substantially reduce the cognitive workload of the user in order to provide a performance enhancement. An analysis of the matching and dematching tasks suggest an explanation for why highlighting works in one case but not the other. The matching task required confirmation on each of seven items that a match has occurred. The dematching task requires only that one of the seven items has ceased to match. We would expect then that color highlighting would be much more effective in assisting with the more complex task due to the increased mental workload involved in that task.

A further analysis of the task would suggest that group highlighting should have provided a performance advantage over individual highlighting. That these were not different was somewhat surprising. Two explanations for this lack of a significant difference between group and individual highlighting are readily apparent. First, the reduction in cognitive workload may not be great enough to produce a significant difference between these conditions. Second, this study may have lacked the power to reliably detect such a difference if it did exist. These, and other possibilities need to be explored.

That there was no difference in the safety task due to highlighting is not so easily explained. The design of the study may have been flawed. The safety task did not take place very often during the experiment. It also never took place at the same time as the matching task. Thus, subjects may have divided their attention successfully between the two tasks. Again, further research is required.

General Discussion

Taken as a group, these studies indicate two things. First, that laboratory research can be used in the process of display design. While some of the findings of this research are relatively intuitive, others are not. Experiment 2 contradicts those who advocate a particular foreground/background combination for displays under normal lighting conditions. Using the most common foreground/background color combinations available, no performance differences for a relatively involved task were detected.

The second important aspect of this work is that it indicates that there is still a great deal of need for additional basic research in the area of CRT display designs. Experiment 1 demonstrated the usefulness of organizing displays. It did not, however, address the issue of an optimal organization for this or any other display application. Experiment 3 demonstrated that color highlighting can assist the user in task performance under some conditions. This experiment did not, however, explore the full range of when such highlighting is or is not useful.

Acknowledgements

This work was supported in part by NASA Grant NAG2-492 awarded to the first author.

References

Christ, R. E. (1977). Four years of color research for visual displays. Proceedings of the Human Factors Society 21st Annual Meeting. 319-321.

Christ, R. E. (1975). Analysis of Color and Its Effectiveness. Las Cruces, NM., New Mexico State University, NMSU-ONR-FR-75-1.

Fisher, D. L. and Ten, K. C. (1989). Visual displays: The highlighting paradox. Human Factors, 31(1), 17-30.

Granaas, M. M. (1989). TIMEX2: A modified C-language timer for PC-AT class machines. Behavior Research Methods, Instruments, & Computers, 21(6), 619 - 622.

Granaas, M. M. & Rhea, D. C. (1989). Techniques for Optimizing Human-Machine Information Transfer Related to Real-Time Interactive Display Systems. NASA TM 100450.

Mitchell, C. M., Stewart, L. J., Bocast, A. K., & Murphy, E. D. (1982, December). Human Factors Aspects of Control Room Design: Guidelines and Annotated Bibliography. Fairfax, VA. George Mason University, NASA TM 84942.

Moore, A. L. (1986). The Role of a Real-Time Flight Support Facility in Flight Research Programs. NASA TM 86805.

NASA (1988). Space Station Freedom Program Human-Computer Interface Guidelines. Houston, TX., Johnson Space Center, NASA USE 1000 V. 2.1.

Ramsey, H. R. & Atwood, M. E. (1979, September). Human Factors in Computer Systems: A review of the Literature. Englewood, CO. Sciences Applications, Inc., (NTIS No. AD-A075-679).

Schniderman, B. (1987). Designing the User Interface. Massachusetts: Addison-Wesly Publishing Company.