EGG-M-00982

PERFORMANCE-BASED EVALUATION OF GRAPHIC DISPLAYS FOR NUCLEAR POWER PLANT CONTROL ROOMS

Rohn J. Petersen, William W. Banks and David I. Gertman

EG&G Idaho, Inc.

INTRODUCTION

This paper reports several methodologies for evaluating the perceptual and perceptual/decision making aspects of displays used in the control rooms of nuclear power plants. This NRC funded study focuses upon the Safety Parameter Display System (SPDS) and relates the utility of the display to objective performance and preference measures obtained in experimental conditions. The first condition is a traditional laboratory setting where classical experimental methodologies can be employed. The second condition is an interactive control room simulation where the operator's performance is assessed while he/she operates the simulator. The third condition is a rating scale designed to assess operator preferences and opinions regarding a variety of display formats. The goal of this study is the development of a cost-efficient display evaluation methodology which correlates highly with the operators ability to control a plant.

The initial evaluation effort was directed toward evaluating three SPD prototypical configurations: Stars, Bar Graph, and Meters (see Figure 1). Displays similar to each of these configurations are currently being used in nuclear power plant control rooms. The SPDS serves as a status display, keeping the operator informed of the status of several critierical safety parameters.

Three experiments from the laboratory portion of the evaluation are described below. The other portions of the evaluation including a correlative analysis of the methodologies will be presented elsewhere.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results cf such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the U.S. Nuclear Regulatory Commission. Work supported by the U.S. Nuclear Regulatory Commission, under DOE Contract No. DE-AC07-76ID01570.

LABORATORY EXPERIMENTS

For the laboratory portion of this investigation psychophysical and information processing methodologies are very useful approaches to developing evaluation methodologies. In this context we are attempting to assess those characteristics of visual displays which most profoundly influence information extraction. In essence, we are adopting an information processing view of information extraction and focusing upon the early stages of the process. Although the perceptual elements of information extraction are of primary interest, the overall goal of relating SPDS information to safe and efficient nuclear power plant operation is also recognized. An assumption that pervades the experiments described below is that the information content of the displays can be controlled so that the configuration of the display can be manipulated independently.

This portion of the evaluation consists of four separate experimental tasks: Detection--addressing the alerting function; Spatial Localization--addressing the localization portion of problem recognition; Parameter Recognition--addressing the identification portion of problem recognition; and Event Recognition--addressing the problem diagnosis function. Detection and spatial localization were analyzed using a signal detection paradigm, while parameter recognition and event recognition are subjected to chronometric analysis.

EXPERIMENT 1: DETECTION

The theory of signal detection (TSD) offers a methodology for obtaining independent and quantitative estimates of both sensitivity and response bias (Tanner and Swets, 1954; and Green and Swets, 1966). With regard to our intial efforts in evaluating displays, we are not necessarily interested in the response and decision

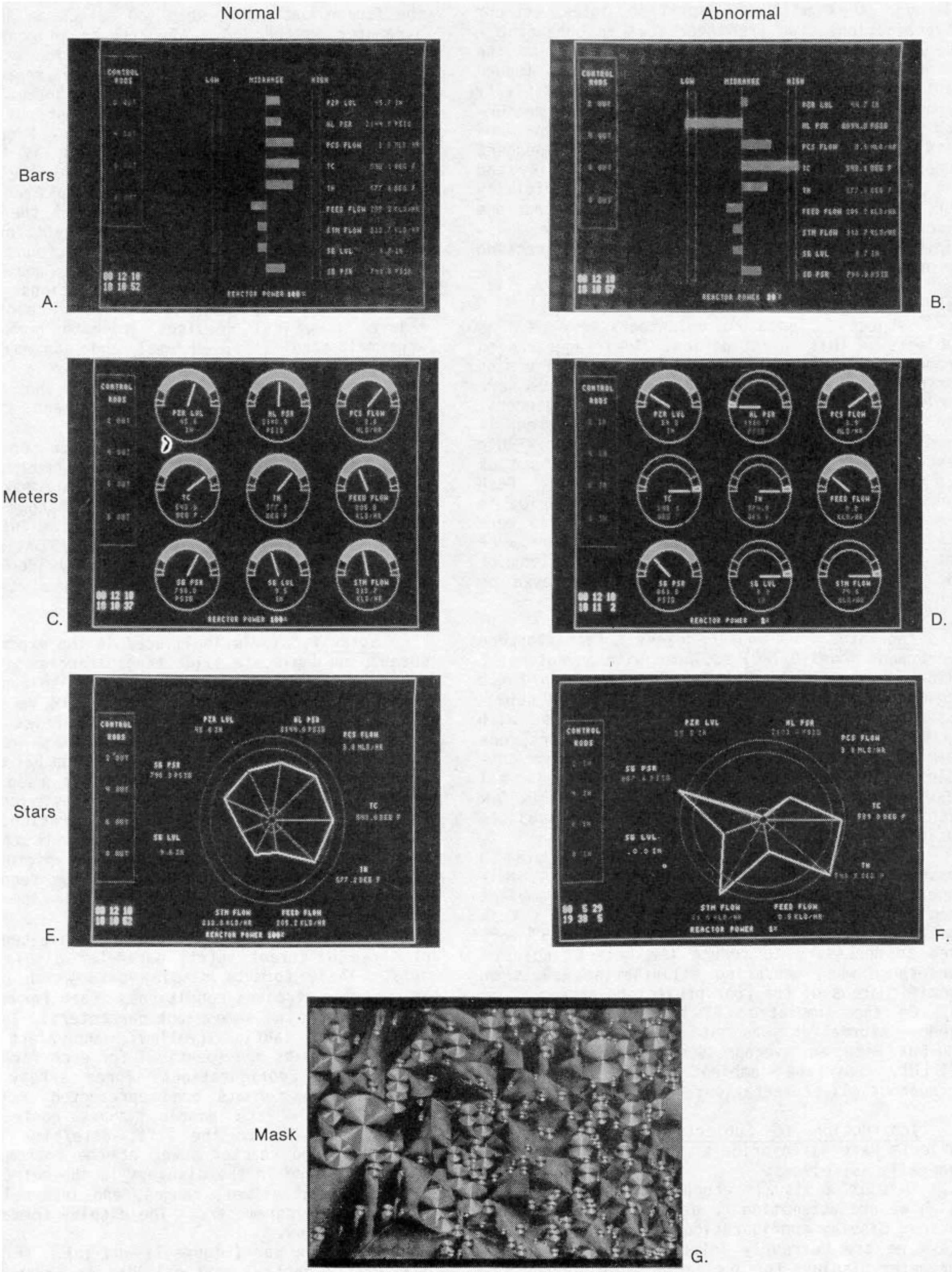

Figure 1. Examples of the stimulus material used in the detection experiment.

aspects of the operators mental processes, although they will be important later in our investigations. We are interested in obtaining a clear view of the operator's sensitivity to the information contained in the display. The important point is that an unbiased view of sensitivity cannot be obtained without controlling or measuring the observer's response bias (decision process). Signal Detection Theory allows independent quantitative values for both the sensitivity and the response bias. Determining the sensitivity for each of the three display configurations allows us to meaningfully compare the three displays, as to their relative ease of extracting information.

Method

Subjects. Ten adult volunteers were used as subjects in this investigation. Their ages ranged from 26 years to 44 years and all reported vision correctable to 20/20. Five of the subjects were nuclear plant operators and five were engineers. The first group of subjects are currently qualified reactor operators from the Loss-of-Fluid Test (LOFT) reactor plant. They have a mean of 9.4 years of reactor operating experience. Each operator received his initial reactor training in the U.S. Navy. The second group of subjects were EG&G Idaho, Inc. engineers. These engineers were not trained in the details of the LOFT plant or the significance of the parameters displayed on the SPDS formats.

Apparatus. A dual channel tachistoscope (Gerbrands Model G1180) equipped with an automatic slide changer (Model G1180) and adaptation field logic interface (Model G1159) was used for stimulus presentation. This device was equipped with 4 channel timer (Model 300-4T), two shutters, one beam splitter and associated shutter drive console. (Gerbrands Corp. Arlington, Mass.) All testing was conducted in a room 4.57 m by 6.10 m with 1.52 m partitions placed around the subjects position.

Illumination levels were recorded using a Gossen Cadmium-sulfide cell light meter. A hemispherical diffuser was used to measure ambient room illumination levels from the subject's test position. Spot attachments of 15° and 7.5° were used as necessary to reduce the meters angle of acceptance when measuring illumination levels on specific areas of the rear projection screen.

On the simulated CRT display the red and green information was at an illumination of 700 LUX with an average screen illumination of 525 LUX. Average ambient room illumination throughout all presentations was 1.75 LUX.

Instruction to Subjects. Instructions to subjects were given prior to any testing and were generally as follows.

This is a visual recognition experiment, in which we are attempting to determine the value of various display configurations. The type of displays we are currently interested in are Safety Parameter Displays for nuclear power plants.

During the test, you will be asked to observe the screen and report when you detect an abnormal parameter on the SPD. You will be in control of the presentation of the displays. Your task is to identify the state the display represents, i.e., all normal parameters or some abnormal parameters. There will be three different configurations for SPD used in this experiment. Figures 1a and 1b show a typical bar graph display in both normal and abnormal states. Note that the abnormal states are represented by red bars and by a red numerical reading which indicates the actual value of the parameters. These two forms of recognizing abnormal displays will be found on all display configurations. Figures 1c and 1d are normal and abnormal meter configurations. Meter needle positions, background field color and colored numerical readings indicate normal and abnormal conditions. Normal and abnormal star configurations are shown in Figures 1e and 1f. Star shape, spoke size and color and colored numerical readings indicate normal and abnormal conditions.

The displays will be shown to you for only a brief period of time. If you cannot determine the state of the display, abnormal/normal, make your best guess. The display will then be shown to you for a slightly longer period of time. This will continue until you are consistently making the correct response. That is, correctly identifying normal displays as "normal" and abnormal displays as "abnormal."

Stimuli. The stimuli used in the experiments were 35 mm duplicate slide transparencies of photographs of reactor transient data displays on a cathode ray tube. The photographs were taken with a Contax Model RTS Camera using a Zeiss Planar f2.8, 66 mm, macro lens. The CRT image was displayed through a Dunn Instrument Camera 631 system. Ektachrome 200 color slide film was used. The stimuli are described in three parts: content, parameter format, and display configuration.

Content: Stimuli content refers to the actual reactor transient data which is pictured on the test slides. The data comes from recordings of plant instrument readings during experiments on the LOFT reactor.

Parameter Format: Test slides have been made of three different safety parameter display formats. These formats display data which provide an overview of plant conditions. Each format displays exactly the same plant parameters. The normal (green), caution (yellow), and alert (red) parameter limits are identical for each format.

Display Configuration: Three safety parameter display formats are represented among the test slides. Each display shows control rod status in a box to the left, date/time in the lower left and reactor power at the bottom. The only difference in the displays is the method used to show noraml values, ranges, and interrelationships between parameters. The display formats are described below.

Deviation Bar (Figure 1a and 1b). This display uses a central vertical line to indicate the normal value. Parameter deviations from this value show as bars to the left or right of normal. High and low range values are shown as vertical lines. Parameter descriptions and digital values are on the right of the display.

Meter Display (Figure 1c and 1d). This display represents parameter values as needle positions on nine meters drawn on a CRT. The green, yellow, and red ranges are shown on the meters with only the color corresponding to the current parameter value lit. Digital values (color coded) and parameter descriptions are inside each meter.

Circular Plot (Star) (Figure 1e and 1f). This display represents parameter values as positions on the spokes of a circle. A small inner circle represents range minimums with an outer circle representing maximums. Current value spoke positions are tied together to form a nine-sided polygon. Digital values and parameter descriptions are shown around the outside of the maximum range ring.

Intertrial Mask: An intertrial mask display was presented after each trial for the duration of the intertrial interval. The mask consisted of an enlarged photograph which consisted of a pseudo-random color pattern (Figure 1g).

The displays subtended a horizontal visual angle of 13.4° and a vertical angle of 11.4°.

Procedure: Operator Training. Operator subjects were given more extensive training then engineer subjects. This training was to prepare them for future display testing that will be more complex than detection testing. Each operator was briefed on the three SPDS formats and on the normalization schemes for displayed parameters. An engineering simulator was used to drive the displays so that each operator subject observed the same simulated plant scenarios on each display format in real time. Following simulation training for operator subjects, each was required to correctly sketch each display format and explain the parameter normalizations.

The second group of subjects are EG&G Idaho, Inc. engineers. These engineers were not trained in the details of the plant or the significance of the parameters displayed. Training of the engineer subjects was limited to familiarization with each display. The subjects were shown each display in normal and abnormal states to ensure that they knew how these states were represented.

Procedure: Test. Three types of safety parameter display configurations were used as separate conditions in this experiment. Each subject was presented with a minimum of three blocks of trials for each configuration. Each block contained nine normal displays, and eighteen abnormal displays. The order and sequence of the trials were randomized. Each subject was familiarized with the displays as per the instructions and then given a series of thirty warmup trials before actual testing was initiated. In addition, they were given detailed instructions before the warmup session began. Following instructions and the warmup, testing began with a test display presented for 5.0 ms. The exposure duration was then increased by 10 ms per block until the subject made no errors during three successive blocks. The subjects responses were recorded at each intensity level. Between every presentation the masking slide was displayed to eliminate the possibility of establishing latent images. Each block of trials consisted of a single display type (e.g., meters). The order of presentation of the test blocks was balanced across subjects and type of display configuration. The subjects were given a fifteen minute rest between display configuration changes.

Design. A "within subjects" nested design was used in this experiment. Four independent variables were manipulated: three fixed variables, display configuration, exposure duration, and type of subject (operators vs non-operators), with the random variable, subjects, being nested within type of subject. Three dependent variables were examined: perceptual sensitivity (d'), response criterion (β), and response accuracy (percent correct).

Results

The data from this experiment are shown in Figure 2. An analysis of variance of these data was conducted in three separate parts. The first part analyzed the perceptual sensitivity (d') of the subjects as a function of display type and exposure duration as shown in Figure 2a and 2d. The analysis revealed that both display type ($F(2,16) = 10.88$, $p < 0.01$) and exposure duration ($F(7,56) = 21.17$, $p < 0.01$) are significant main effects, i.e., a change in sensitivity was observed as a function of display type and of exposure duration. In addition, a significant interaction was shown for the display type and exposure duration ($F(14,112) = 2.15$, $p < 0.01$).

Since one of the objectives of this experiment was to evaluate the three display formats, orthogonal planned comparisons of the data were conducted. The first comparison was meters versus bars and star. This comparison revealed a significant difference ($t(16) = 4.02$, $p < 0.01$), i.e., bar and star formats were better for detection than meters. The second comparison, bars versus star, showed no significant difference for detections.

The second part of the analysis considered the accuracy of the subject's responses in terms of percent correct. These data are plotted in Figures 2b and 2e. As with sensitivity, two significant main effects were found, display type ($F(2,16) = 12.94$ $p < 0.01$) and exposure duration ($F(7,56 = 24.13$, $p < 0.01$). Again, a significant interaction was found, display type by exposure duration ($F(4,112) = 5.15$, $p < 0.01$). Orthoganal planned comparisons of the data on the type of display format showed the same pattern of results as the first part of the analysis, i.e., the difference between meters versus stars and bars was significant ($t(16) = 4.90$, $p < 0.01$). The second comparison, bars versus star, showed no significant difference.

The third part of the analysis examined the data in terms of the subject's response criterion (β). These data are plotted in Figures 2c and 2f. The only significant main effect was exposure

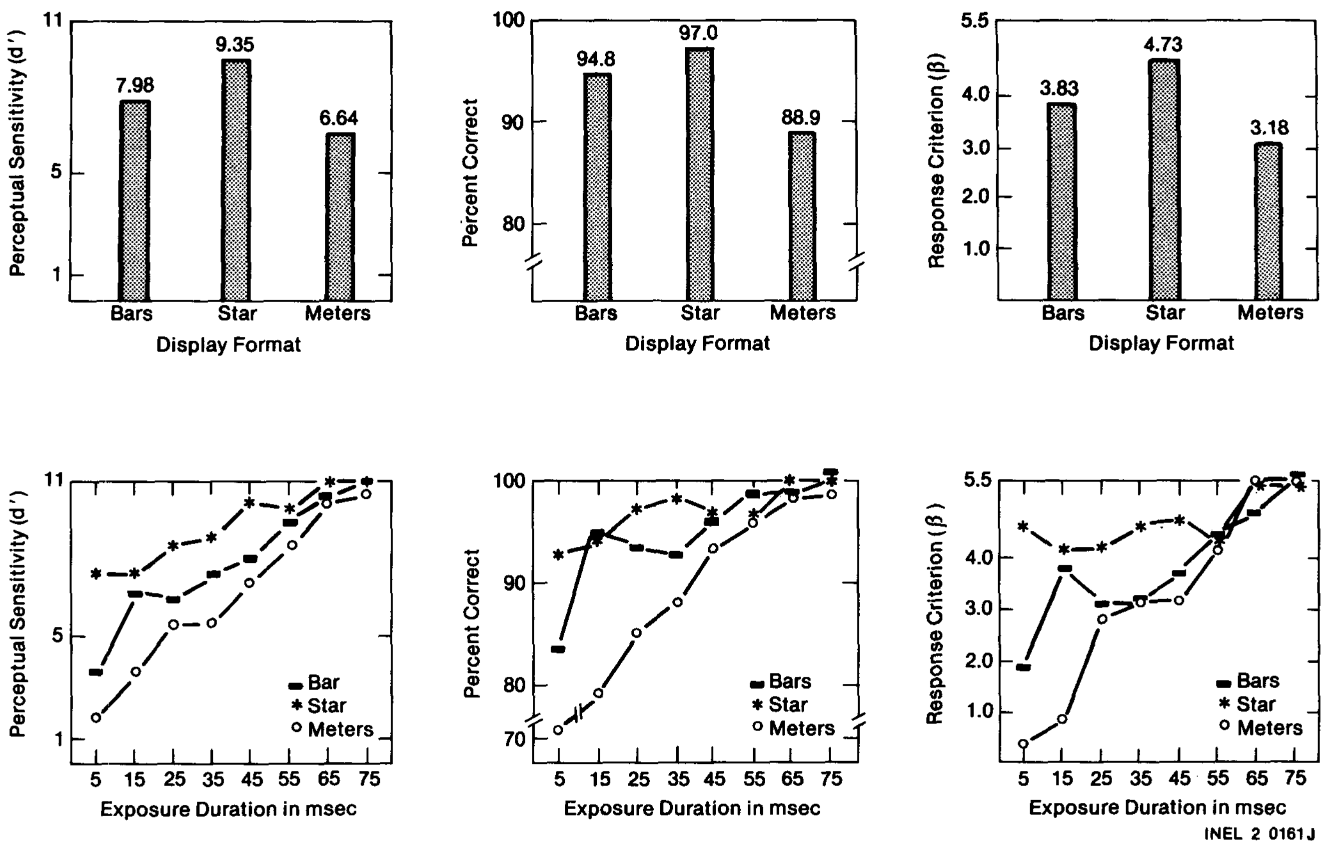

Figure 2. Data from Experiment 1 plotted as a function of display configuration and response duration for percent correct and signal detection measures.

duration (F(7,56) = 9.58, $p < 0.01$). No interactions were found to be significant that the display type variable is very close to being significant ($p < 0.01$).

Conclusions

Star and bar formats were shown to be better display configurations than the meter format for detection tasks. Although star and bar formats cannot be distinguished on a strictly statistical basis, visual examination of Figure 2a and 2b leads one to infer that the star format may hold some advantage over the bar format for this task. In addition, examination of Figure 2d and 2e, the data representing the significant interaction of display format and exposure duration for both perceptual sensitivity (d') and response accuracy (percent correct), reveals that the star format promotes consistantly better performance on this detection task. Therefore, the star format apparently transmits information concerning parameter conditions better than the meter format, which requires that the viewer have longer exposures (more information) to accurately assess the condition of the display. Interestingly, the rate of change in the subjects' ability to extract information seems greater with the meter format, perhaps due to a ceiling on the subjects' responses using the star format. Given that the order of presentation for exposure duration was fixed (5 to 75 ms), the subjects may have been engaging in more perceptual pattern learning from the meter display than from the other two formats.

Longer exposure times increased the measured sensitivities (d') of the subjects and produced more correct responses. This is not surprising, since the amount of information available for making a decision would usually increase with a longer exposure, and the more information available, the better the decision would be.

The background and experience of the subjects was not a significant variable in this task. Operators could not be distinguished from the engineers on the basis of performance. Therefore, we can assume that the detection task is purely perceptual in nature and is not influenced by the differences in training and experience between these two groups.

EXPERIMENT 2: SPATIAL LOCALIZATION

Experiment 2 was designed to examine the subjects ability to determine the locus of information on a safety parameter display. The question being investigated concerns the ability of subjects to locate particular information in the three SPDS configurations. The subject's task in this experiment is to record the location of abnormal parameters when a display is presented using the t-scope.

Method

The second experiment utilized essentially the same paradigm as Experiment 1. The following exceptions should be noted however.

1. The subjects task was changed from a detection task to a localization task. Now, instead of merely reporting the detection of an abnormal parameter, the subject must mark the location of all abnormal parameters on a data sheet.
2. The duration of the display presentation was set and held constant at 700 ms. This duration was derived during a short pilot study. The number of display presentations was limited to 27 trials as described in Experiment 1.
3. In addition to marking the data sheets with the location of the abnormal parameters, the subjects were asked to learn the names and locations of the parameters by labeling the locations during the intertrial interval. The goal of this task was to have the subject be able to correctly label each parameter by the end the 27 trials. In this way this experiment served as the training for Experiment 3.
4. Twelve adult volunteers served as subjects in this experiment. They were a mixture of operators and engineers all of whom fall within the characteristics used to describe the subjects in Experiment 1.
5. The intertrial mask was not used in this experiment. Instead, a dark screen was used.

Results

The data from Experiment 2 are shown in Figure 3. An analysis of variance of these data was conducted for each of the three dependent variables, perceptual sensitivity (d'), percent correct and response criterion (β). The perceptual sensitivity analysis (see Figure 3a) showed a significant effect for the type of display factor, $F(2,33) = 6.10$ $p < .01$, $MSe = 1.58$. However, none of the planned comparisons were significant. The analysis of the percent correct data revealed a significant effect for display type, $F(2,33) = 5.91$, $p < .01$, $MSe = .00778$. Again, none of the comparisons were significant. The response criterion (β) analysis showed a significant effect for the type of display factor, $F(2,33) = 9.52$, $p < .01$, $MSe = .468$. The planned comparison showed a significant effect for the comparison of meters versus bars and stars, $t(33) = 2.18$, $p < .05$.

Conclusion

The data from Experiment 2 suggest that the meter displays are easier to locate information on. Unfortunately, this conclusion must be hedged since most of the individual planned comparisons were not significant. The fact that the main

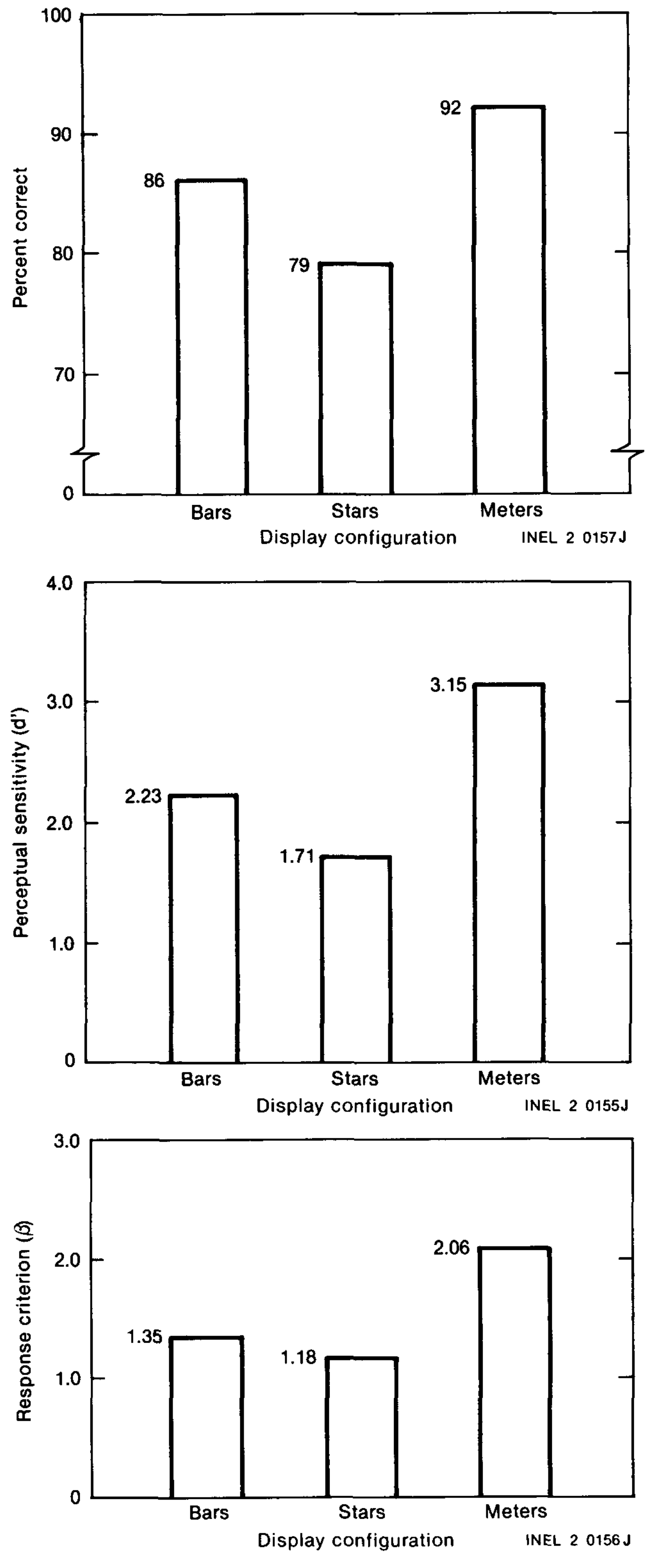

Figure 3. Data from Experiment 2 plotted as a function of display configuration for percent correct and signal detection measures.

effects for all three of the dependent variables were significant carries some weight in supporting the conclusion. At any rate, it seems clear that there is a trend, which indicates that meters are perceived better in this task. The superiority of the meter display for this task may be due to the spatial arrangement of the information on the display. The information on the meters display is spread out over the entire screen while the information on the star and bars displays is clustered in the center. In fact, it seems that the very aspect of the display which may have contributed poor performance in the detection task seems to be involved in the relatively superior performance in the this localization task. Additionally, it appears that the subjects change their strategy, response criterion (β), as a function of display type.

Experiment 3: Parameter Recognition

Given control over the information content of the displays, there are several methodologies which have been developed to address the preliminary stages of cognitive processing (perception). One prominent method, reaction time (Donders, 1868; Estes, 1975; Posner, 1975; Pachella, 1974; and others), requires the subject to view a display, make some sort of decision regarding the display, and make a response as quickly as possible. Information extraction is assumed to involve a complicated set of mental operations each of which requires a finite amount of time. It is reasoned that when more processing is required to make a response more time will also be required. The main premise is that more processing time will be required when information is more difficult to extract from a particular display configuration. Therefore, one way of assessing ease of information extraction would be to measure the time required to respond to the various displays. Of course, more complicated decisions or responses will also increase reaction time and must be controlled. Both the decision and the response related components of reaction time can be assumed to remain constant, if both the task and the response remain constant across the experimental conditions. Therefore, any change in response time can be attributed to a change in the amount of perceptual processing required to extract information from the different display configurations.

Method

Subjects. The same subjects used in Experiment 2 participated in this experiment.

Apparatus. In addition to the apparatus described in Experiment 1, an Apple II plus computer with 64 K memory, Supertalker (a voice digitizer), a real-time clock (Appleclock) and a modified I/O board to control display presentation and monitor subjects' button responses were used. The software included Apple DOS 3.3, and a control program written in BASIC. A movable button box 35.6 cm wide x 21.5 cm deep x 8.5 cm high placed on the table in front of the subject. Two mushroom buttons were mounted on the top of the box. The buttons protruded in 2.0 cm above the box surface. Each button was 34.0 mm in diameter with a throw of 7.0 mm. The button mounted on the right side of the box was black while the one on the left was red.

Instructions to Subjects. After the subject had attained the training criteria for the display type, they were informed that they would participate in a reaction time experiment. It was explained that they were to listen to the computer voice a particular parameter and find that parameter in the display. The subjects task then was to determine if the parameter was normal or abnormal. If the parameter was abnormal the subject pressed the left (red) button. If the parameter was normal the subject pressed the right (black) button. Feedback was supplied on a CRT located to the left of the subject. The subjects reaction time and the correctness of the response were shown on the screen. In addition, the subject initiated each trial by depressing a footswitch.

Stimuli. The same slides used in Experiment 1 were used in this experiment.

Procedure: Operator Training. Experiment 2 served as the training session for this experiment. At the conclusion of Experiment 2 the subjects were required to label each format data sheet with the names of the parameters in the proper locations.

Procedure: Testing. The test procedure described in Experiment 1 was followed with the following deviations.

1. The display was presented for a constant 2000 ms.
2. The accuracy and speed of the subjects responses were recorded. Any responses longer than 2500 ms were not recorded.
3. No mask was presented during the intertrial interval.
4. The warmup block consisted of twenty seven trials.
5. Only three test blocks were presented.

Design. A "within subjects" design was used in this experiment. Four independent variables were manipulated: Three fixed variables, display configurations, parameter type (Normal vs Abnormal), and test blocks, and one random variable, subjects. Two dependent measures were examined: response time and percent correct.

Results

The data from Experiment 3 are plotted in Figure 4. An analysis of variance of the reaction time data revealed a marginally significant main effect for display type, ($F(2,198) = 3.20$, $p < .05$, $MSe = 67689.46$) and no other significant main effects or interactions. In addition, none of the orthoganal planned comparisons were significant. An analysis of variance of the accuracy data found significant main effects for display type ($F(2,198) = 5.62$, $p < .01$, $MSe = 79.5$) and test blocks, replications ($F(2,198) = 5.90$, $p < .01$, $MSe = 79.5$). A significant

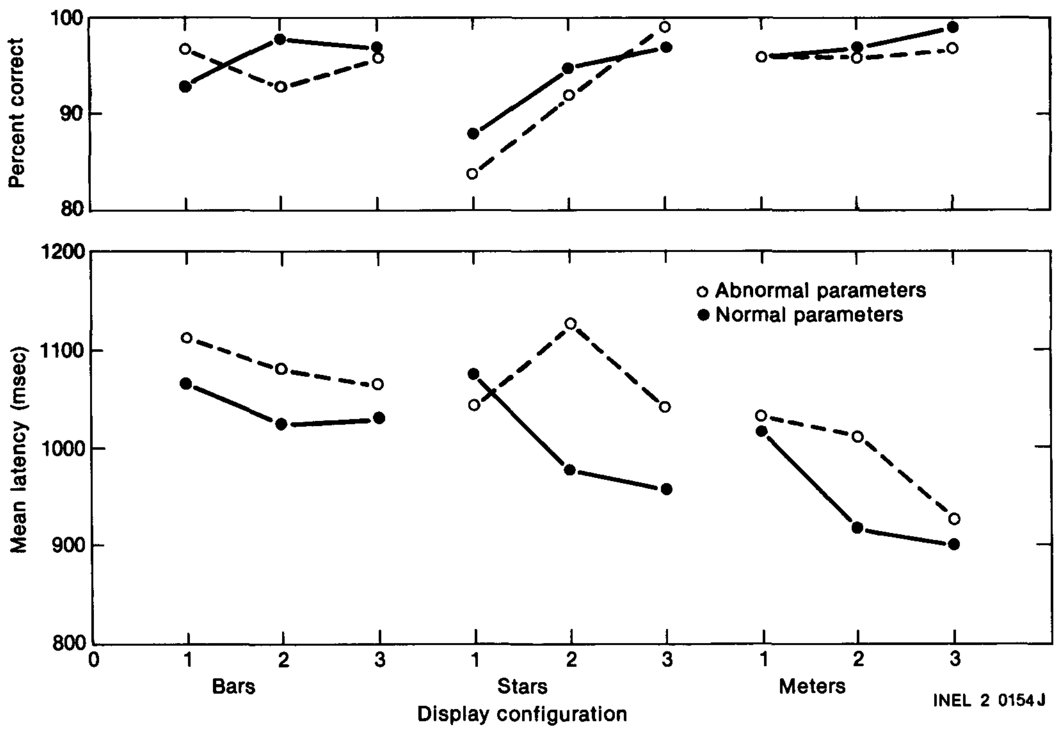

Figure 4. Reaction time and percent correct data from Experiment 3.

interaction was found for display type and test blocks (F(4,198) = 3.46, p < .01, MSe = 79.5). Again the planned comparisons shown no significant effects.

Conclusion

The results of this experiment are not conclusive in any sense of the word. There are indications, in both the reaction time data and the percent correct data, that display configuration influences performance. However, the exact form of that influence can not be determined statistically. Perhaps the strongest statement one would want to make on the basis of this data is that the trends look promising but the paridigm must be tightened to provide the control necessary to make diffinitive conclusions.

SUMMARY

The experiments described above represent a portion of the laboratory part of our evaluation effort. The data from these experiments and others will be correlated with data from rating scales, simulation studies and other data producing paridigms (check lists, field studies, etc.) to produce an overall measure of display adequacy. A major focus of this effort will be to produce a cost-effective methodology for evaluating displays.

REFERENCES

1. F. C. Donders, Over de snelheid van psychische processen. Onderzoekingen gedaan in bet Physiologisch Lo boratorium der Utrechtsche Hoogeschool, 1868-1869, Tweede reeks, II, 92-120. Translated by W. C. Koster in W. G. Koster (Ed.), Attention and Performance II, Acta Psychologica, 1969, 30, 412-431.
2. W. K. Estes, "The Locus of Inferential and Perceptual Processes in Letter Identification," Journal of Experimental Psychology: General, 1975, 104, 122-145.
3. D. M. Green and J. A. Swets, Signal Detection Theory and Psychophysics, New York; Wiley, 1966.
4. R. G. Pachella, The Information of Reaction Time in Information-Processing Research, in B. H. Kantowitz (Ed.), Human Information Processing; Tutorials in Performances and Cognition, Hillsdale NJ; Lawrence Erlbaum Associates, 1974.
5. M. I. Posner, Psychobiology of attention. In M. S. Gazzaniga & C. Blakemore (Eds.). Handbook of Psychology, New York; Academic Press, 1975.
6. W. P. Tanner and J. A. Swets, A Decision Making Theory of Visual Detection, Psychological Review, 61, 401-409, 1954.