183184

P. 7

N94-24189

AUTOMATED SYSTEM FUNCTION ALLOCATION AND DISPLAY FORMAT: TASK INFORMATION PROCESSING REQUIREMENTS

Mary P. Czerwinski
Lockheed Engineering and Sciences Company

Research directed by Marianne Rudisill, Manager, Human Computer Interaction Lab, NASA JSC.

INTRODUCTION

Questions relevant to the Human Factors community attempting to design the display of information presented by an intelligent system are many: What information does the user need? What does the user have to do with the data? What functions should be allocated to the machine versus the user? Currently, Johnson Space Center is the test site for an intelligent Thermal Control System (TCS), TEXSYS, being tested for use with Space Station Freedom. The implementation of TEXSYS' user interface provided the Human-Computer Interaction Laboratory with an opportunity to investigate some of the perceptual and cognitive issues underlying a human's interaction with an intelligent system.

An important consideration when designing the interface to an intelligent system concerns function allocation between the system and the user. The display of information could be held constant, or "fixed," leaving the user with the task of searching through all of the available information, integrating it, and classifying the data into a known system state. On the other hand, the system, based on its own intelligent diagnosis, could display only relevant information in order to reduce the user's search set. The user would still be left the task of perceiving and integrating the data and classifying it into the appropriate system state. Finally, the system could display the *patterns* of data. In this scenario, the task of integrating the data is carried out by the system, and the user's information processing load is reduced, leaving only the tasks of perception and classification of the patterns of data. Humans are especially adept at this form of display processing [1, 2, 11, and 12].

Although others have examined the relative effectiveness of alphanumeric and graphical display formats [7], it is interesting to reexamine this issue together with the function allocation problem. Expert TCS engineers, as well as novices, were asked to classify several displays of TEXSYS data into various system states (including nominal and anomalous states). Three different display formats were used: *fixed* (the TEXSYS "System Status at a Glance"), *subset* (a relevant subset of the TEXSYS "System Status at a Glance"), and *graphical*. These three formats were chosen due to previous research showing the relevant advantages and disadvantages of graphical versus alphanumeric displays (see Sanderson et al., 1989 for a review), and because of the vast amount of literature on the beneficial effects of reducing display size during visual search in cognitive psychology (see Shiffrin and Schneider, 1977; Schneider and Shiffrin, 1977). The hypothesis tested was that the graphical displays would provide for fewer errors and faster classification times by both experts and novices, regardless of the *kind* of system state represented within the display [11]. The subset displays were hypothesized to be the second most effective display format/function allocation condition, based on the fact that the search set is reduced in these displays [5, 6]. Both the subset and the graphic display conditions were hypothesized to be processed more efficiently than the fixed display condition, which corresponds to the "System Status at a Glance" display currently used in TEXSYS.

METHOD

SUBJECTS

Four frequent users of TEXSYS, thermal control engineers at JSC, participated in the experiment. The subjects had an average of

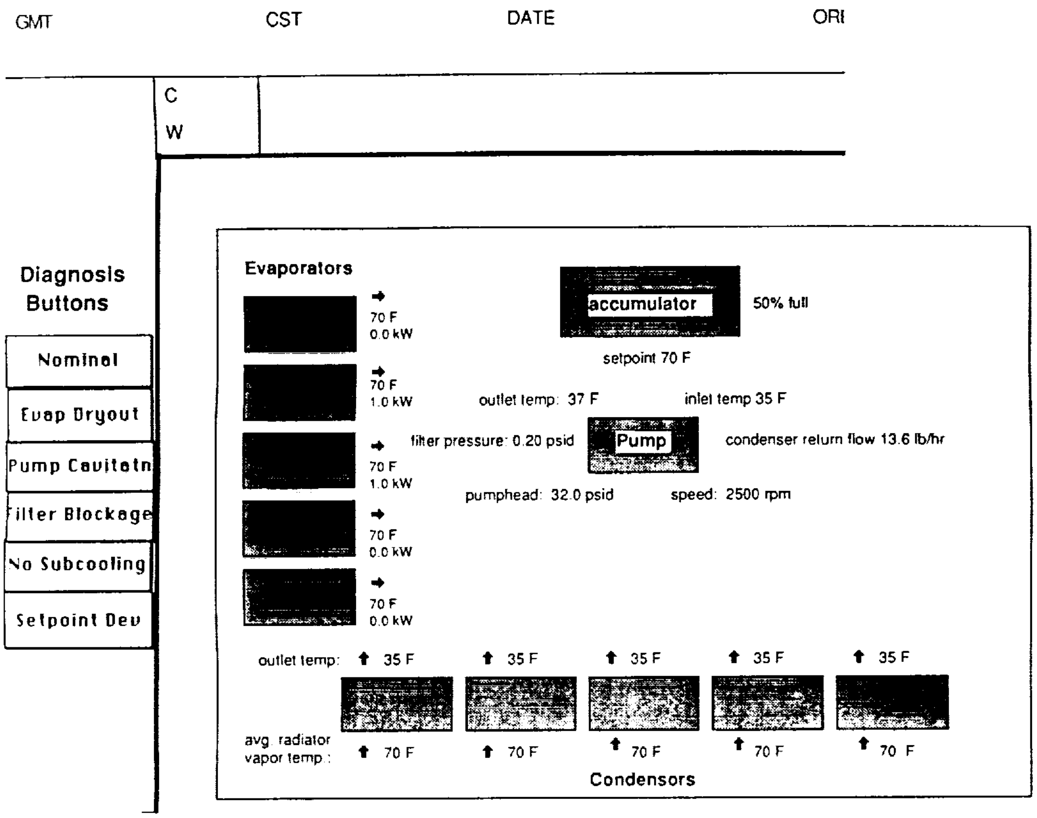

Figure 1. The "fixed" display.

eight years experience. Six novices, all engineers, also participated in the experiment. None of the novice subjects was familiar with the two-phase thermal bus system used in the TEXSYS project, nor with thermal control systems in general. All subjects were experienced users of Macintosh computers, and all had normal or corrected-to-normal vision.

STIMULI AND MATERIALS

The design, presentation, and collection of all stimulus materials and data were carried out on a Macintosh IIx computer using SuperCard and SuperTalk. A mouse was used for all subject inputs. Examples of the fixed, subset, and graphical display formats can be seen in Figures 1, 2, and 3, respectively. Note that, while the fixed and graphical displays both contain information about all of the major system components, the subset displays only show a subset of the system data.

System Faults. Five different system anomalies could occur during the experiment: evaporator dryout, filter blockage, pump cavitation, loss of subcooling and setpoint deviation.

MATCHING NOMINAL AND ANOMALOUS DISPLAYS

Nominal displays were matched with anomalous displays for two reasons. First, designing the experiment in this manner avoids biasing the subjects toward responding "fault" or "no fault." The second reason is related to a peculiarity in the subset display condition. In these displays, subjects were told that the expert system had made a reasonable guess as to the critical system state, and only information concerning that state was shown. In nominal conditions, in order to control for the amount of information displayed to the subject, the same component subsets were shown as in the fault conditions. However,

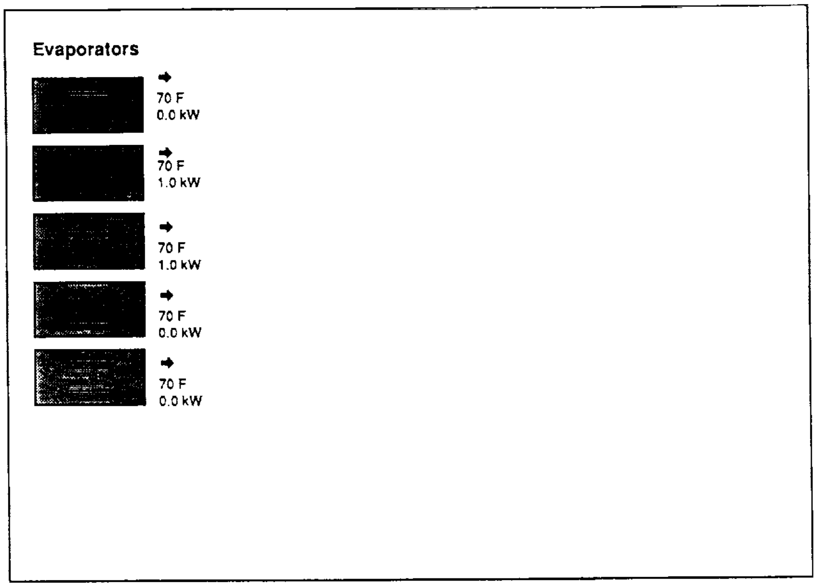

Figure 2. The "adaptive" display.

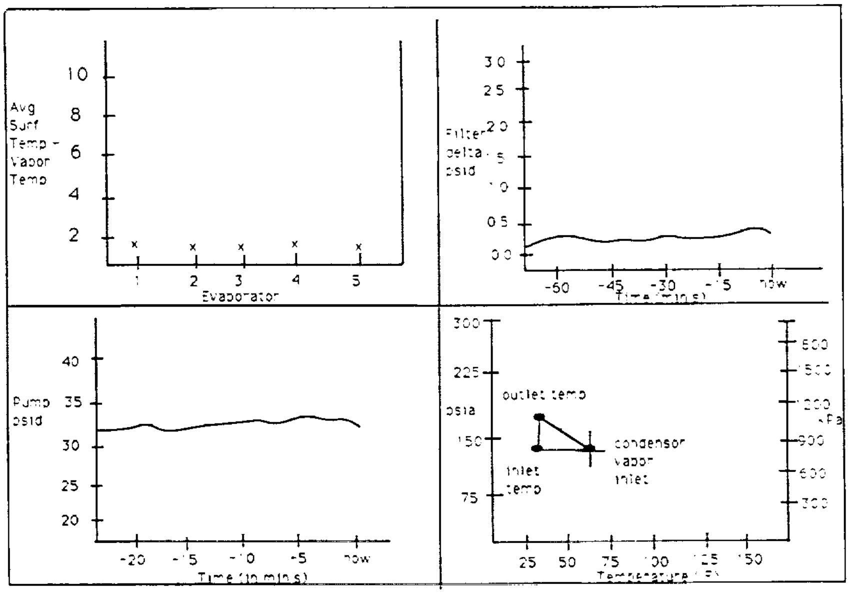

Figure 3. The "graphic" display.

since the displays were nominal, the displayed data values were never aberrant. The matching of displays simply involved replicating the no-fault displays and then changing particular component values to off-nominal for the fault displays.

DESIGN

The experimental design was a 3 x 2 x 5 x 2 factorial, with three different display formats (fixed, subset, and graphic), both nominal and anomalous display instances, five different state instances, and two repetitions per condition. Note that this design implies that a system fault occurred on 50% of the trials. There were two groups of subjects run in the experiment: experts and novices. The novices were given two sessions of training, which added an extra factor (session) to their design. All variables were run within subjects, but experts and novices were analyzed separately. The three different display formats were blocked, such that there were three blocks of 20 trials (including the repetitions) in each experimental session. The order in which each subject received the three display formats was counterbalanced. All of the other factors were randomized within a display condition block. The dependent measures collected were reaction time and percent correct.

PROCEDURES

Experts. During an orientation, prior to actual data collection, the experts were shown a table of nominal data values (as well as the acceptable ranges of deviation for those values) for the major components of the system.

Novices. The same materials that were used for orientation of the experts were used to train the novices. Unlike the expert subjects, the novices studied the nominal operations table for approximately 50 minutes[1]. During this time, they were informed about the patterns of data which might occur for each of the five system faults[2].

Both expert and novice subjects were instructed to monitor the displays presented to them for one of the six system states. They were instructed to search the system display quickly, without making errors, for system status information. Once the displayed data had been categorized by the subject, s/he was instructed to indicate which system state had occurred via a button-click with the mouse input device.

All subjects were run through a practice experiment, in which an example of each Display Format x System State combination was included. Feedback in the case of an error was provided for the subjects as a computer beep.

The diagnosis buttons were located to the far left of the display, as can be seen in Figure 1. The CONTINUE button (on the intertrial screen) was located in the center of the position previously occupied by the six diagnosis buttons. This button placement was used in order to reduce the motor movement time involved in selecting any of the six diagnosis buttons. Trials were self-paced, and subjects were encouraged to take a short break between blocks. The experimental session lasted approximately one hour.

RESULTS AND DISCUSSION

ERRORS

Experts. Overall, the experts operated at an accuracy level of 93% correct. A separate analysis of variance (ANOVA) with repeated measures was run on the error data for both

---

[1] This was the average amount of time needed to train each individual subject, although each subject's time varied slightly due to the number of questions they asked.

[2] Novice subjects were run through the experiment for two reasons: there were too few experts available to participate in the experiment, and the experts were extremely well-practiced at diagnosing the System Status-at-Glance displays. Both problems might have biased results. The extra novice session was to ensure that novice subjects had a chance to attain near-expert levels of performance in this task.

TABLE 1.

Average Logged Reaction Times for Diagnosing the Six System States in Each Display Format for Expert and Novice Subjects.

	Experts			Novices		
State	Fixed	Adaptive	Graphic	Fixed	Adaptive	Graphic
Nominal	9.3	8.5	9.4	8.5	8.0	8.1
Evap Dryout[3]	9.4	9.1	9.8	8.3	7.6	7.9
Filter Block[4]	9.4	8.7	8.9	8.9	8.3	8.0
Pump Cav[5]	9.1	8.6	8.8	8.5	8.0	7.7
No Subcooling[6]	9.5	9.3	9.7	8.5	8.0	8.6
Setpoint Dev[7]	9.3	8.2	8.9	8.9	8.6	8.7
Average	9.3	8.7	9.3	8.6	8.1	8.2

experts and novices. For experts, the ANOVA was a 3 x 2 x 5 x 2, representing the factors of display (fixed, subset, and graphic), fault or no fault, type of fault, and repetition. The analysis revealed a significantly larger number of errors with nominal displays, $F(1,3)= 22.09$, $p < .02$. No other effects were significant for the experts.

Novices. On the average, the novice subjects performed at an accuracy level of 91.2% correct in session 1, and 93% correct during session 2. For novice subjects, a 2 x 3 x 2 x 5 x 2 ANOVA with repeated measures was carried out on the error data. The first variable corresponds to the two sessions of training that novice subjects received during the experiment; all other factors are identical to those used in the expert subject's ANOVA. There was a significantly larger number of errors in the nominal display condition, $F(1,5) = 20.05$, $p < .01$. No other effects were significant.

REACTION TIMES

A *t*-test was performed between the overall average reaction times of the experts and the overall average (across two sessions) of the novices. No significant difference was found between the two groups[8], $t(8) = 1.61$, $p > .05$.

Experts. The pattern of results for the expert subjects can be seen in Table 1. The ANOVA revealed significant main effects of display condition, $F(2,6) = 7.9$, $p < .05$, with subset displays processed the most quickly, followed by the graphical displays. No other main effects were significant for the expert subjects. However, there was a significant interaction between whether or not a fault was present and which *type* of fault had to be diagnosed, $F(4,12) = 3.27$, $p < .05$. This interaction reflected the fact that there were larger response time differences within the anomalous display instances than within the nominal displays, although planned comparisons did not reveal any significant differences between the anomalous display instances (all p's > .05).

---

[3] Evaporator Dryout

[4] Filter Blockage

[5] Pump Cavitation

[6] Loss of Subcooling

[7] Setpoint Deviation

[8] No significant difference was found in the error data, as well.

Novices. The pattern of results for the novice subjects is shown in Table 1. The ANOVA revealed significant main effects of session, $F(1,5) = 38.33, p < .01$; display condition, $F(2,10) = 14.04, p < .01$; and type of fault being diagnosed, $F(4,20) = 13.51, p < .001$. Session 2 was faster than session 1, and, again, the subset displays were processed most quickly. A significant interaction occurred between display condition and the *type* of fault being diagnosed, $F(8, 40) = 2.76, p < .05$. This interaction was not observed for the expert subjects, and reveals a pattern of data whereby certain faults are processed more quickly in particular formats. Finally, there was a significant interaction between whether or not a fault was occurring and the type of fault to be diagnosed, $F(4,20) = 3.98, p < .05$. This interaction is similar to that observed in the expert data. This interaction reflected the fact that, for nominal conditions, none of the display instances were processed significantly faster than the average of the others, as determined by planned comparisons (all p's $> .05$). However, in the fault condition, the evaporator dryout fault was processed significantly faster than the average of the other faults, $t(9) = -1.88, p < .05$, and the setpoint deviation fault was processed significantly slower than the average of the other faults, $t(9) = 2.13, p < .05$.

Finally, it should be noted that for both the experts and the novices there was probably a speed-accuracy trade-off operating on the reaction times within the no-fault condition. Specifically, errors increased significantly in the nominal condition, while reaction times were no different than those in the fault displays. This may have masked any significant effects occurring in the no-fault display conditions.

EXPERIMENT 2

Experiment 1 demonstrated the benefit of showing only relevant information to the subject. It was also shown that novices appear to diagnose certain faults better in a subset, alphanumeric format, while other fault diagnoses benefit from a graphical display format. However, one problem with interpreting this result has to do with the fact that the amount of information was not controlled between the subset alphanumeric and the graphical display conditions. In other words, there was no subset, graphic display condition. Experiment 2 equated more fully the two conditions and it was a means by which to explore the issue that a graphical format would always be a better representation when only the relevant state information is displayed.

It was also hypothesized in Experiment 2 that the kind of information processing required while diagnosing a display could affect performance. This was because one subset of the Experiment 1 faults (evaporator dryout and loss of subcooling) could be described as requiring a serial scan of the data followed by one memory comparison in all of the format conditions (the one memory comparison refers to the comparison of the displayed data value with a memorized nominal value for that system component). All other faults required the identification of one or more data values, the same sort of mental comparison with a nominal value, and then a further comparison with other component values. This extra comparison step could be argued to add load to working memory, and perhaps a graphical format is better in these conditions [11]. These ideas were tested in Experiment 2 as well.

For this experiment, one of the subset displays (relevant to the evaporator dryout fault) was used throughout the entire experiment. In one half of the experiment, subjects simply scanned evaporators to detect off-nominal surface temperatures in both graphical and alphanumeric display formats. In another half of the experiment, an extra comparison step was required in order to diagnose the data displayed in both formats.

METHOD

SUBJECTS

Seventeen Lockheed Engineering and Sciences engineers voluntarily participated in the experiment. All subjects were naive concerning the operation of the automated Thermal Control System being simulated.

STIMULI AND MATERIALS

For the "scanning" level of the decision-making variable, the alphanumeric displays from the subset condition in Experiment 1 were used for this experiment. The graphical display was modified from Experiment 1 for this condition, so that a bar graph format was used. For the "scan + compare" condition, pump information was added to each of these display formats. Essentially, a pump outlet temperature was added to the displays for comparison with the evaporator information.

DESIGN

The experiment was a 2 x 2 x 2 factorial design, with two levels of the kind of decision-making steps required to diagnose a fault (scan, and scan + compare), both alphanumeric and graphical display formats, and nominal vs. anomalous display instances. Nested within the anomalous display instances, and only within the scan + compare conditions, was another factor — type of anomalous fault. This variable could not be added to the nomalous displays because nomalous displays do not fall into subcategories in this system. However, we did vary the particular data values within the nomalous displays so that the nomalous and anomalous displays were balanced in the number of *unique* system instances presented to any given subject during a session. This was because more faults were available for diagnosis when pump information was present in the display. Specifically, during the scan + compare trials, the subject had to distinguish four different system states: nominal, evaporator dryout, pump cavitation, or setpoint deviation. Note that in the scan only condition nominal and anomalous trials are equated, while in the scan + compare condition the subject received three times as many anomalous trials as nominal. Both the decision-making and the format variables were blocked, and the order in which subjects received the decision-making conditions was counterbalanced. However, if a subject randomly received the scan only (or scan + compare) decision-making condition first, that subject always received both display format conditions (in a random order) prior to diagnosing the scan + compare (scan only) blocks of the experiment. The magnitude and pattern of the faults within the displays were controlled across the graphic and alphanumeric display formats.

PROCEDURE

The procedure for running this experiment was identical to that for Experiment 1, although only novice subjects were run for a single session.

RESULTS AND DISCUSSION

ERRORS

The errors were submitted to an ANOVA, including the variables of decision-making steps, display format, and type of response (nominal or anomalous). There was no significant pattern of errors.

REACTION TIMES

The reaction time results are shown in Figure 4. The reaction times were submitted to an overall ANOVA, including the variables of decision-making steps, display format, and type of response (nominal or anomalous). The analysis revealed significant main effects of decision-making condition, $F(1,16) = 89.85$, $p < .001$, and display format condition, $F(1,16) = 34.72$, $p < .001$. The scanning only condition was diagnosed more quickly than the scanning and comparing condition, while the graphical format was processed more quickly than the alphanumeric display format. The interaction of decision-making condition and display format was not significant, $F(1,16) = 1.3$, $p = .2$. However, the interaction of display format condition and system state (nominal vs. anomalous) was significant, $F(1,16) = 7.37$, $p < .05$. Finally, a significant three-way interaction was observed between decision-making condition, display format, and system state, $F(1,16) = 9.16$, $p < .01$. The higher-level interactions

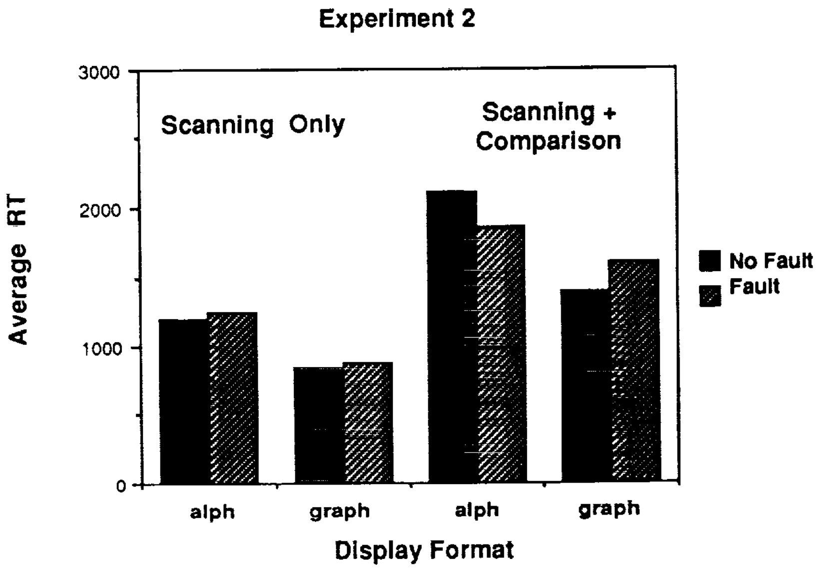

Figure 4. Average reaction time data as a function of decision-making condition and display format in Experiment 2. (alph = alphanumeric, graph = bar graph format).

reflect the fact that nominal (no fault) conditions were detected more readily than faults in all conditions except with the alphanumeric display format involving both scanning and comparing.

The results observed in Experiment 2 showed that diagnosing a subset graphical display took less time than diagnosing a subset alphanumeric display. The scanning only versus scanning and comparison manipulation could be argued to have increased the subjects' processing requirements, since diagnosis times were significantly longer in that condition. However, this increase in processing load did not lead to the interaction between display format and fault type observed in Experiment 1. It may be that the bar graph is a better way of representing data than the graphical representations used in Experiment 1. Several researchers have reported the integral processing benefits of a bar graph representation [3, 4, and 9]. Subjects may have been capitalizing on the configural [8] properties inherent in the bar graph representation in both decision-making conditions. This may be especially important when processing load is high. Some data to suggest that the bar graph representation is beneficial during heavy processing load conditions was observed in the three-way interaction reported in Experiment 2. The pattern of data showed that in the scanning and comparing condition subjects were faster at diagnosing faults in the alphanumeric displays (although still slower than in the graphic displays). Perhaps subjects were reverting to a serial search through the data in the former conditions, due to the high cognitive demands of the task. An obvious test of this notion would be to vary the number of system components showing aberrant data values for this task, in both alphanumeric and bar graph display formats. (In Experiments 1 and 2, only one system component was ever showing off-nominal data values within a display). If subjects revert to scanning in either of the display format conditions due to heavy cognitive task demands, diagnosis times should be shorter, on the average, the greater the number of off-nominal system components [10]. This experiment is currently being run in our laboratory.

ACKNOWLEDGEMENTS

This research was funded by the National Aeronautics and Space Administration, Office of Aeronautics, Exploration and Technology (OAET), and was performed at the NASA/ Johnson Space Center Human-Computer Interaction Laboratory. The author thanks Kevin O'Brien and Kim Donner for valuable contributions during the design phase of this work. The author also thanks Nancy Cooke, Shannon Halgren and Marianne Rudisill for helpful discussions and comments on earlier drafts of this paper.

REFERENCES

1. Casey, E. J. (1986). Visual display representation of multidimensional systems: The effect of information correlation and display integrality. In Proceedings of the Human Factors Society 30th Annual Meeting (pp. 430-434). Santa Monica, CA: Human Factors Society.
2. Carswell, C. M. and Wickens, C. D. (1984). Stimulus integrality in displays of system input-output relationships: a failure detection study. In *Proceedings of the Human Factors Society*, 28th Annual Meeting.
3. Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association, 68*, 361-368.
4. Coury, B. G., Boulette, M. D. and Smith, R. A. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. *Human Factors, 31*, 551-569.
5. Donderi, D. C. and Zelnicker, D. (1969). Parallel processing in visual same-different decisions. *Perception & Psychophysics, 5*, 197-200.
6. Egeth, H., Jonides, J., and Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3*, 674-698.
7. Keiras, D. E. (1988). Diagrammatic displays for engineered systems: Effects on human performance in interacting with malfunctioning systems. *NASA Technical Report No. 29,* July 29, 1988.
8. Pomerantz, J. R., Pristach, E. A., and Carson, C. E. (1989). Attention and object perception. In B. E. Shepp and S. Ballesteros (Eds.), *Object perception: Structure & process.* Hillsdale, NJ: Lawrence Erlbaum Associates.
9. Sanderson, P. M., Flach, J. M., Buttigieg, M. A. and Casey, E. J. (1989). Object displays do not always support integrated task performance. *Human Factors, 31*, 183-198.
10. Sternberg, (1966). High-speed scanning in human memory. *Science, 153*, 652-654.
11. Wickens, C. D. (1984). *Engineering psychology and human performance.* Columbus, OH: Charles E. Merrill.
12. Woods, D. D. & Roth, E. M. (1988). Cognitive systems engineering. In M. Helander (Ed.), *Handbook of human-computer interaction*, North-Holland: Elsevier Science Publishers.