Putting It All Together: Improving Display Integration in Ecological Displays

Catherine M. Burns, University of Toronto, Toronto, Canada

Computer displays are being designed for increasingly larger industrial systems. As the application domain scales up, maintaining integration across different kinds of views becomes more challenging. This paper presents the results of a study of three different approaches to integration based on the spatial and temporal proximity of related information objects. The domain used for evaluation was a simulation of an industry-scale conventional power plant. All three displays were ecological displays developed using an abstraction hierarchy analysis. Views were integrated in a high-space/low-time, low-space/high-time, and high-space/high-time integration of means-end related objects. During a fault detection and diagnosis task, it was found that a low level of integration, high-space/low-time, provided the fastest fault detection time. However, the most integrated condition, high-space/high-time, resulted in the fastest and most accurate fault diagnosis performance. Actual or potential applications of this research include computer displays for large-scale systems such as network management or process control, for which problem solving is critical and integration must be maintained.

INTRODUCTION

The evolution of modern computer technology has brought many benefits. Computer technology can manage vast quantities of information and permit flexible and graphical display of that information. In the large-scale industrial situation, it might be possible to replace large hard-wired control rooms with much smaller operator stations with a few CRTs.

Although this is initially appealing, evidence from the literature suggests that it might pose new challenges for the human operator. The operator would be faced with a vast virtual information space behind every CRT. Authors have already noted problems with operators "getting lost," not using all the available information, and losing a sense of the plant as a whole (Elm & Woods, 1985; Roth, Mumaw & Stubler, 1993; Woods, Roth, Stubler, & Mumaw, 1990). Woods (1984), in particular, likened this situation to looking through a keyhole: The operator has access to only a small view of a very large information space.

These problems of lack of integration result when the information space is vast and different parts of it are disjointed from other areas. Simply defined, integration means "keeping related things together." An operator who has access to a small view of an integrated and interrelated information space might be able to maintain an awareness, similar to situation awareness (Endsley, 1995), of the entire plant. If that small view is disjointed and unconnected, however, the operator has little chance of keeping any sense of overall plant changes. The next sections expand on integration, exploring the ideas of "related things" and "keeping things together."

Related Things

There are two approaches for determining which things are related to each other. In the

proximity compatibility principle (PCP; Wickens & Carswell, 1995), things are related to each other when they must be used sequentially during a task. PCP argues that interfaces should be designed with the goal of keeping task-related items together. In contrast, the ecological approach (Rasmussen & Vicente, 1990; Vicente & Rasmussen, 1990) defines things as related if they have means-end relations. In the means-end relation, one piece of information is related to another if it can answer either how something is accomplished (a means) or why something is needed (an end). The ecological perspective is best suited for fault-management and troubleshooting activities for which operators must reason from the perspective of a system's purpose in order to find the problem with its components. Field studies of problem solving have shown that people reason in a means-end fashion (Rasmussen, 1985; Rasmussen & Jensen, 1974). Studies of ecological displays versus nonecological displays have confirmed this approach (Christoffersen, Hunter, & Vicente, 1996; Pawlak & Vicente, 1996; Vicente, Christoffersen, & Pereklita, 1995).

A broader view of these two approaches suggests that in the problem-solving situation they do not differ. According to the field study results of problem solving (Rasmussen, 1985; Rasmussen & Jensen, 1974), people must access means-end information sequentially. The accessing of means-end information, therefore, is critical to the task of problem solving. Thus, following the principles of PCP, means-end information should be kept together.

How to Keep Things Together in an Interface

An interface designer has control over several aspects of designing a graphical interface. Foremost, the designer must decide where the information will go on the display and when the information will be seen in complex displays. PCP (Wickens & Carswell, 1995) concentrates on *spatial proximity* – that is, which pieces of information should be closest together spatially. In complex displays with a large amount of information, it becomes critical to determine when information should be seen

(the *temporal proximity* of information). Although Wickens and Carswell (1995) did not discuss temporal proximity directly, this seems to be a reasonable extension of PCP.

In order to adhere to the definition of integration as "keeping related things together," a designer should provide means-end information in close spatial and temporal proximity to provide integration in displays that would improve problem-solving. This idea led to the following hypothesis: High-spatial and high-temporal proximity of means-end related information will improve operator performance on fault diagnosis when compared with displays that do not keep means-end related information together in high-spatial/high-temporal proximity.

The objective of this research was to investigate the hypothesis that means-end information should be kept in close spatial and temporal proximity. The general benefits of the ecological approach over nonecological approaches to display design have been shown in other work (Pawlak & Vicente, 1996; Vicente & Rasmussen, 1990).

THEORY-BASED APPROACH

To test this hypothesis in the context of a large system, I used a simulation of a conventional power plant provided by Asea Brown Boveri (ABB) Corporate Research (Heidelberg, Germany). I developed an abstraction hierarchy model of the simulated plant (Burns, 1998) in order to determine means-end information. The abstraction hierarchy model provided means-end information organized by level of abstraction (means-end level) and level of decomposition (whole-part level). Based on the model, views of each abstraction level were designed, thereby creating views of information at the same means-end level. These views were then combined into plant displays using different combinations of spatial and temporal proximity. In particular, a high-spatial/high-temporal (HH) display, a high-spatial/low-temporal (HL) display, and a low-spatial/high-temporal display (LH) were developed. A low-spatial/low-temporal display was not developed because it was unreasonable from a design perspective.

The Simulated Plant

ABB prepared a simulation of a coal-fired power plant. The simulation covered primarily the water cycle of the plant, so this was modeled and then implemented in displays. The simulation controlled 402 plant variables. The simulator ran in continuous real time but was not interactive; that is, participants could monitor but not control the plant.

Abstraction Hierarchy Model of the Plant

An abstraction hierarchy model identifies means-end related information within a two-dimensional information space that spans the gap between the intended purpose, or ultimate end, of the plant (functional purpose) and its physical implementation, or lowest-level means (physical function and physical form). The intermediate levels contain constraints that must hold for the plant to achieve its intended purpose. In the case of a power plant, the intended purpose is to produce electrical energy. To achieve that, the plant must convert and transport mass and energy (abstract function). This is typically achieved through temperature and pressure changes of a transport medium such as water (generalized function). These changes at generalized function are achieved by heaters, boilers, and turbines (physical function). The location, appearance, and condition of that equipment are physical form information.

In large systems, the abstraction hierarchy is worked out at different levels ranging from whole to part (decomposition). In this case I looked at the entire plant, major subsystems, redundant parallel lines of components (called trains), and individual components. This gave a complete model of 10 cells of information covering four abstraction levels and four decomposition levels (Figure 1). There were two regions of the abstraction hierarchy that were not modeled. The first was the physical form level. Because this was a simulated plant, no physical form information was available. Physical form information would typically be provided as an in-plant video. Also, because the extreme regions of the space are not used in problem solving, they were not modeled (Rasmussen, 1985).
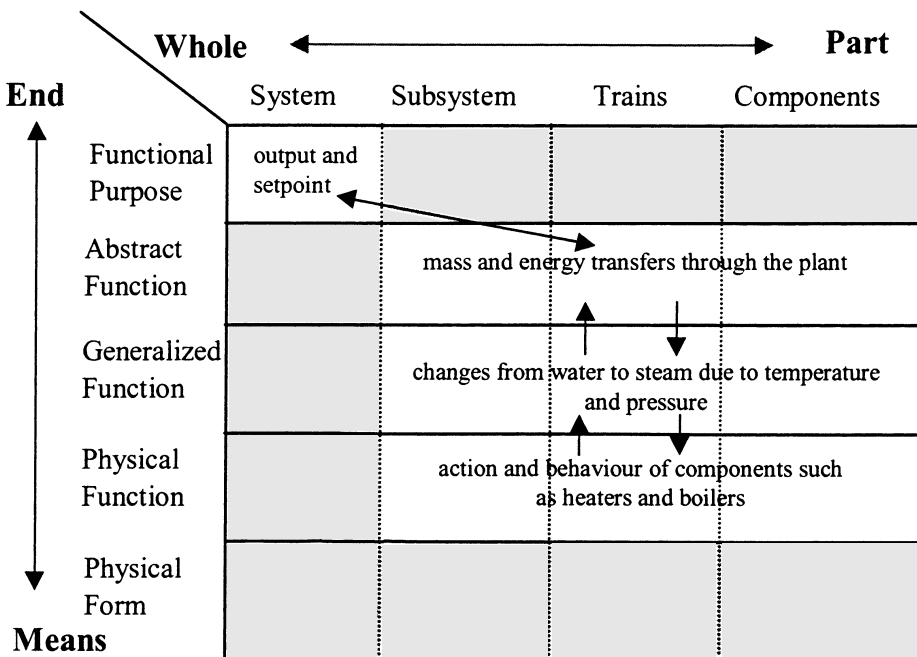

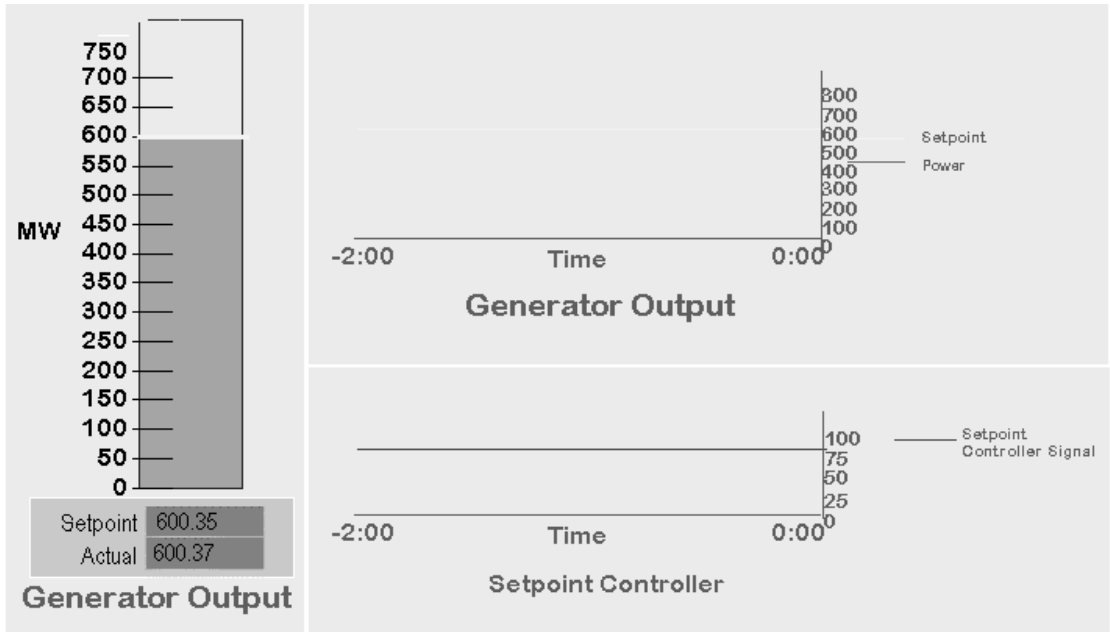*Figure 1.* Regions of the abstraction hierarchy that were modeled.

Figure 2. Functional purpose view.

Implementation of the Abstraction Hierarchy

Each cell of the abstraction hierarchy (Figure 1) was implemented in a separate view. The functional purpose view (Figure 2) showed plant output and plant set point. The abstract function views, shown in Figure 3, showed mass and energy levels throughout the plant. This view showed mass and energy levels in bar graphs, in context with nearby components to show changes, mapped against a background of normal plant values. Generalized function views (Figure 4) showed temperature, pressure, and entropy levels throughout the plant in a graph called the Rankine cycle. This particular display has been evaluated successfully (Vicente et al., 1996). This display is a plot of temperature, pressure, and entropy values against the saturation curve of water; this enabled users to perceptually determine the state of H_2O (liquid, vapor, or a mixture) at various points in the process. The physical function view (Figure 5) showed plant components and settings. Figures 2, 3, 4, and 5 show the least aggregated views of that abstraction level. All views provided access to trended dis-plays of the variables in order to show changes over time.

Integration of Views

The 10 views were integrated in three different ways to test the hypothesis. Thus the individual views were the same in all conditions, and the conditions could be compared solely in terms of their integration of means-end information. For that reason the views were integrated vertically (not horizontally) along the abstraction hierarchy. Figure 6 shows a screen from the high-space/low-time display, in which only one level of means-end information was visible at a time. The matrix in the lower left of the screen showed the available views, and the light cell identified the current cell. Participants were free to change from a cell to any other cell. Figure 7 shows the low-space/high-time display, in which all four abstraction levels were visible in parallel but separate windows. Participants could change the level of detail of the abstraction views independently – that is, they could have different levels of detail in different windows. Figure 8 shows the high-space/high-time display, in which all four abstraction levels were visible
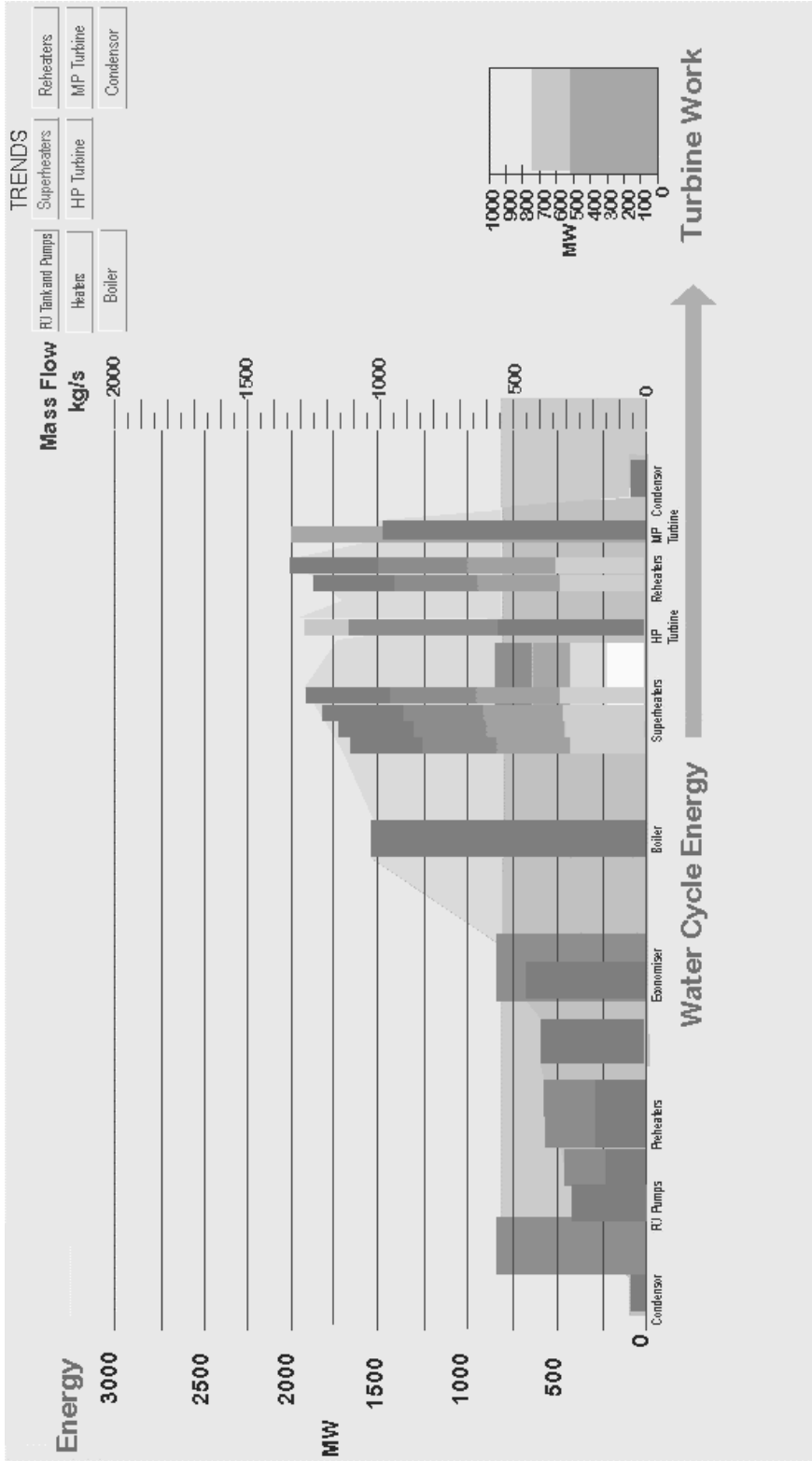
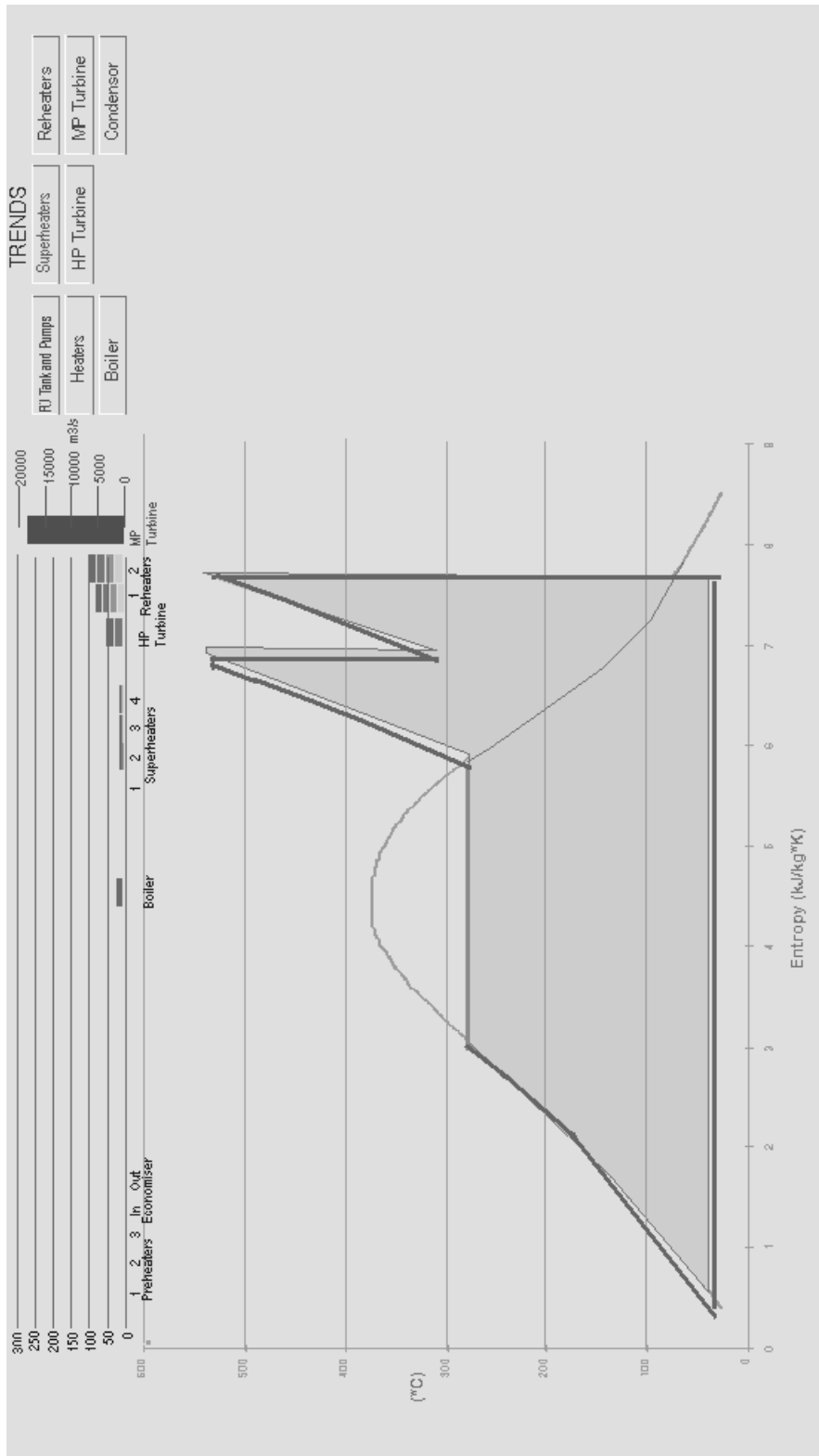*Figure 3.* Abstract function view.
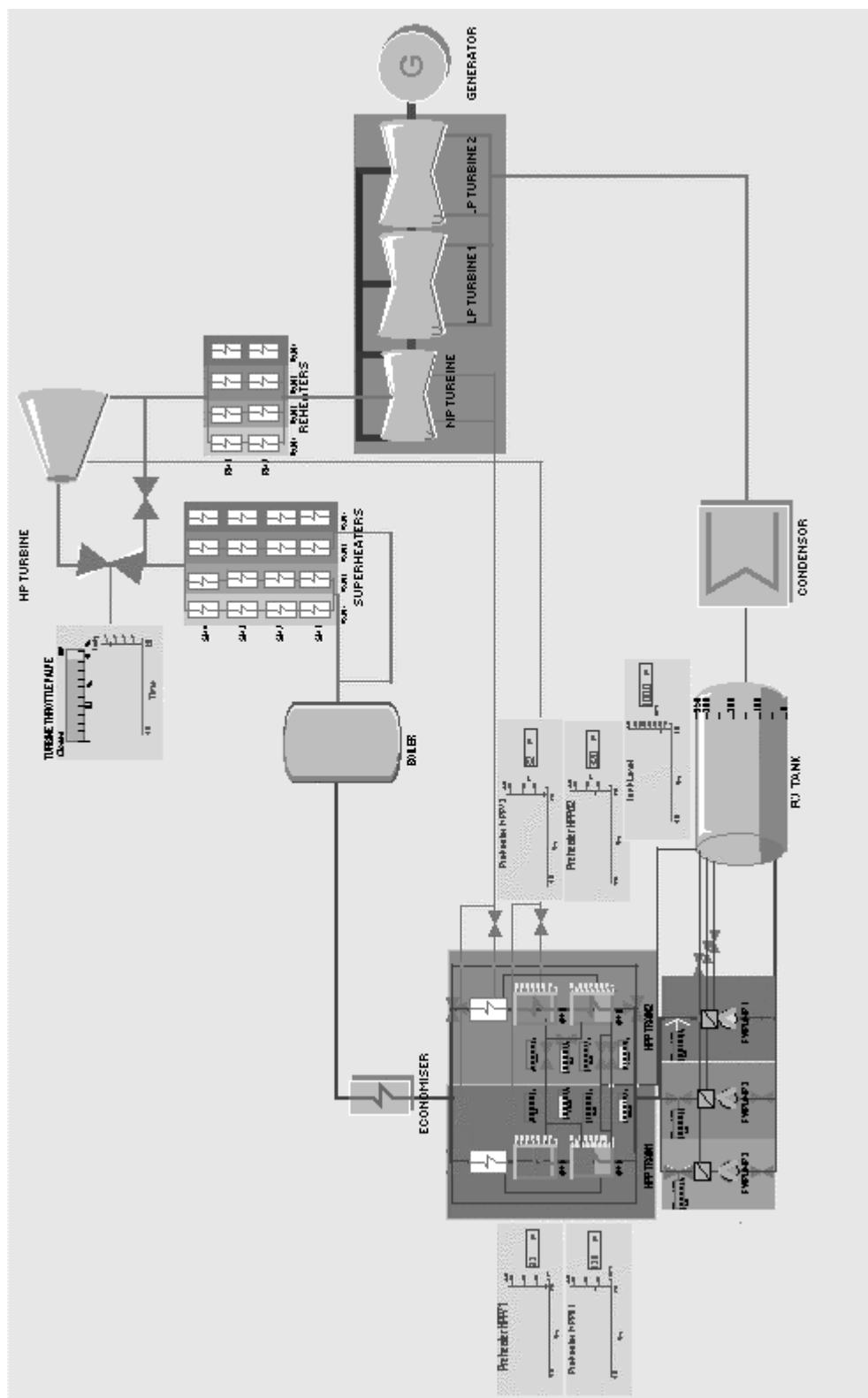
Figure 4. Generalized function view.

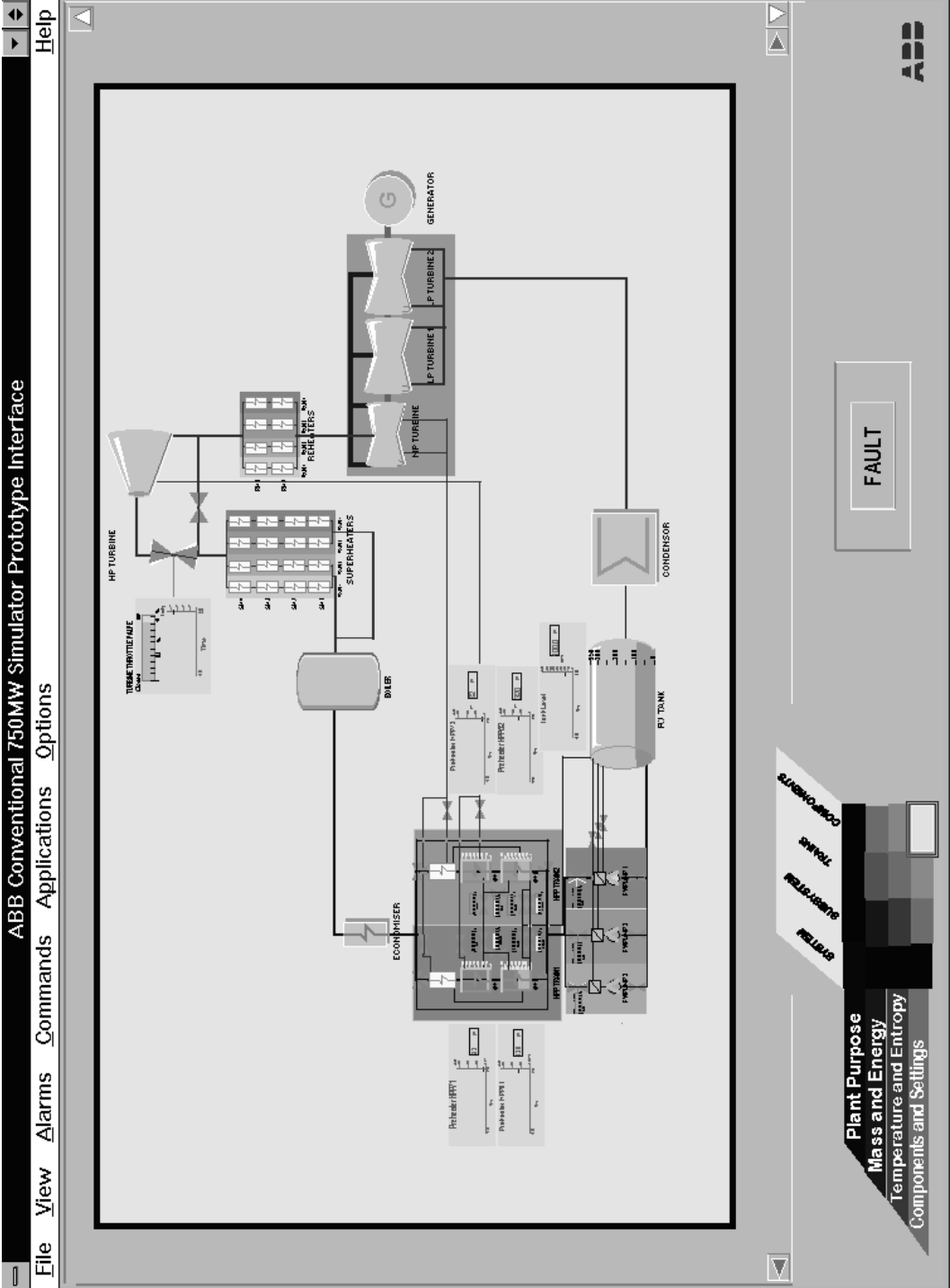*Figure 5.* Physical function view.
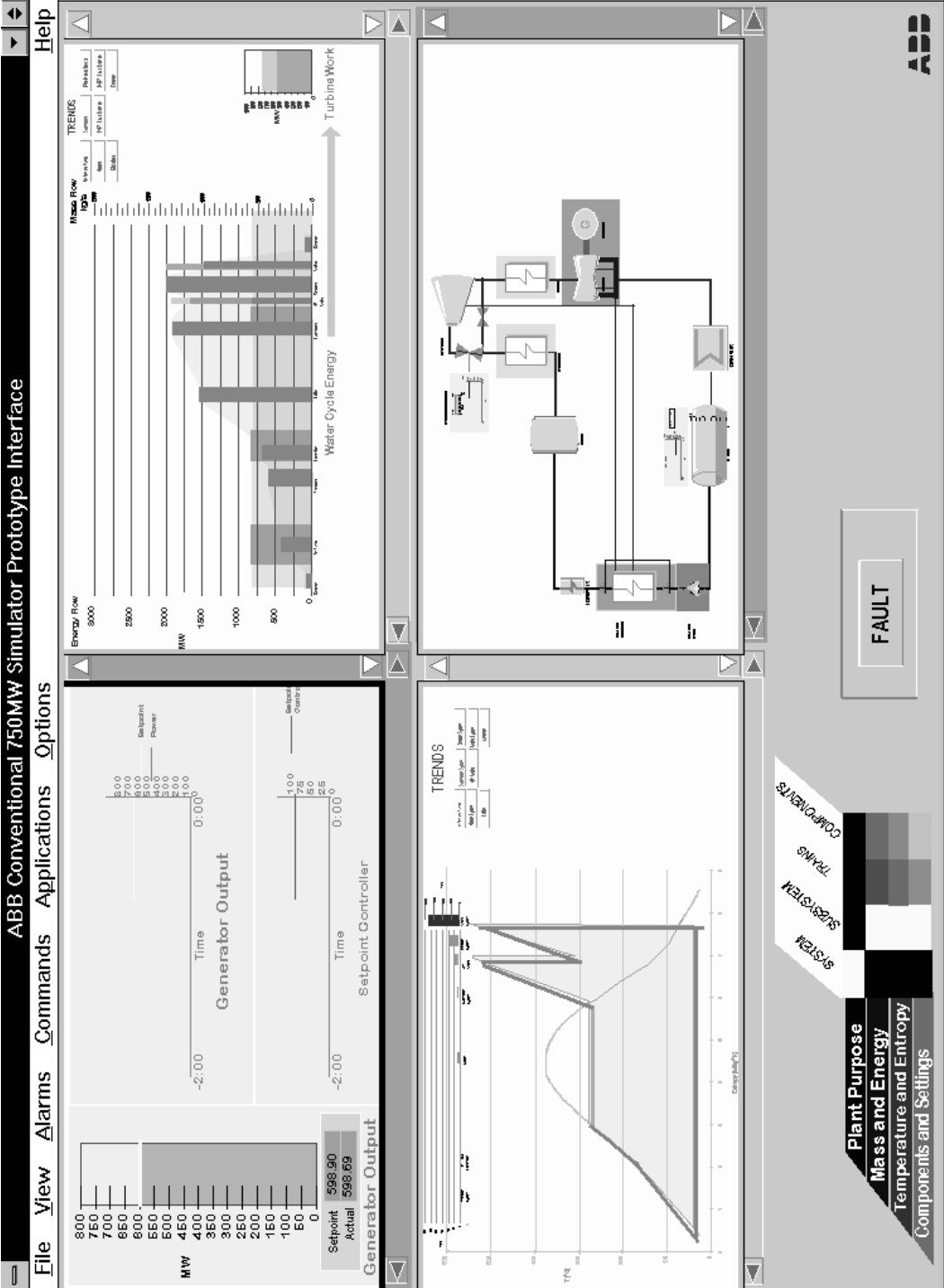
Figure 6. High-space/low-time display.
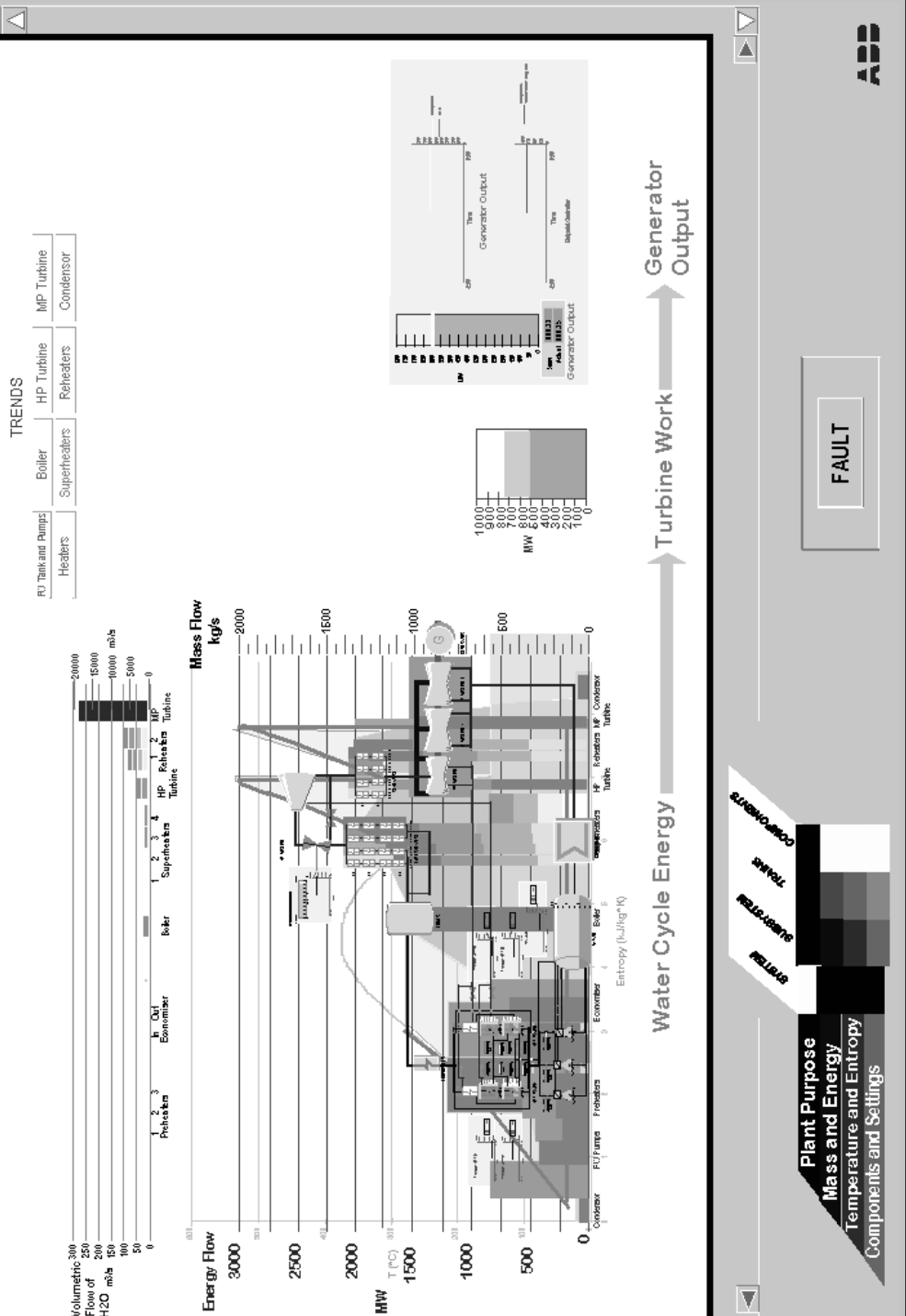
Figure 7. Low-space/high-time display.

Figure 8. High-space/high-time display

but tightly integrated in space and time. For example, participants saw the boiler physical information, boiler temperature and pressure, and boiler energy information in close space and at the same time.

Designing for High-Space/ High-Time Integration

In order to design the high-space/high-time display, I used the Rankine cycle view as an organizing spatial template for the other views. The spatial form of this view was constrained, whereas the spatial form of the other views was less constrained. This organization was retained in all three display conditions. Where possible, occlusion was minimized by placing smaller objects over larger objects and dynamic objects over static objects.

METHOD

In order to evaluate the displays under conditions that were as realistic as possible, I chose to use a fault detection and diagnosis task. Fault scenarios representative of real plant problems were provided by ABB. It should be noted that there was no form of alarming or cueing on the displays; all detections or diagnoses were made on the basis of abnormal plant behavior.

Participants

Participants were solicited from the third-year class of mechanical and industrial engineering students at the University of Toronto and were compensated for their participation. The 18 participants who completed the experiment (6 women and 12 men) ranged in age from 19 to 24 years with a mean age of 20.9 years ($SD = 1.2$). All participants had taken exactly one course in thermodynamics ($SD = 0$) and had been in the same class. Participants were tested on the Spy Ring History Test, and participant groups were balanced based on their holism scores. This was because tendencies toward holism have been found to interact positively with performance on ecological displays (Torenvliet, Jamieson, & Vicente, 1998). After the criterion of balancing the groups was applied, the participants were randomly assigned to display groups of 6 participants each; this was a between-subjects design.

Procedure and Measures

The experiment took 12 h per participant and was spread over 6 days in sessions of 2 h per day. The first four sessions were primarily training and background sessions. The final two sessions were scenario diagnosis sessions and constituted the main evaluative portion of the experiment.

In consideration of the novice nature of the participant pool, there was a substantial investment of time in training the participants. Participants received four different stages of training: training on plant equipment, training on user interface features, an on-line training session with the display, and a practice session consisting of watching a subset of five scenarios. Each stage was concluded with a test on which a performance rate of 80% or better was required in order for a participant to progress further in the experiment. The goal of the training was to ensure that participants understood the plant and its portrayal on the display. The practice session gave participants experience with indicating fault detection and making a diagnosis.

The scenario diagnosis task took place over 2 days in order to keep the length of the sessions within 2 h. Participants observed a total of 17 scenarios, 5 of which they had seen in the fourth training session. Scenarios consisted of a mixture of faults, equipment and sensor failures, and normal scenarios. Scenarios lasted either 5 or 15 min depending on the scenario. The order of the scenarios was randomized once and then presented to each participant in the same order; random presentation for each participant with a relatively small participant pool risked creating order effects. All scenarios started with a randomly determined period of normal plant behavior before the fault was triggered. The fault continued until the end of the scenario. The fault did not progressively worsen, but as in a real plant, plant conditions worsened over time with realistic dynamics if that would be the effect of the fault. Participants used on-screen buttons to indicate the time of detection or diagnosis to permit the timing of their responses. Fault detection time was measured from the onset of the fault to activation of the fault indicator button. Failure

to respond was recorded as a missed detection. Participants were encouraged to speak their thoughts aloud, and all verbalizations were recorded on audiotape. Screen actions were videotaped using a scan converter. The data were checked for relations between performance and time on task, and no such relation was found.

RESULTS

Results are presented on the following measures: fault detection time, missed detections, fault diagnosis time, and diagnosis accuracy. In the tables in the following section, the names of the displays have been abbreviated as such: high-space/low-time (HL), low-space/high-time (LH), and high-space/high-time (HH).

Fault Detection Time

The time to detect a fault was calculated by measuring the time from the start of the fault to the clicking of the fault detection button.

Trials for which the participant exceeded the maximum time for the trial before registering a detection were removed, given that no fault detection was registered. These were considered to be missed detections.

The numbers of data points contributing to each mean after removing the missed detections were HL = 79, LH = 81, and HH = 80 out of a possible total of 90 for HL and LH and 88 for HH. There were fewer points in HH because in two instances the simulator failed to trigger the fault. These trials were removed from the analysis. The total possible points were calculated as (17 trials – 2 normal trials) × 6 participants for each display group.

An analysis of variance (ANOVA) was considered for the analysis of this data. Several factors, however, meant that these data did not conform to the standard conditions for an ANOVA. First, this experimental design was a between-subjects, mixed-plot design, not a within-subjects design. Second, the variance was an unknown function of trial and there-

TABLE 1: Fault Detection Time × Display, Showing Fastest Display

| | HL | | LH | | HH | | Fastest Time |
Trial	Mean	SD	Mean	SD	Mean	SD	Display
1	0:52	0:23	1:26	1:49	3:03	4:03	HL
2	0:55	0:33	0:25	0:34	0:45	0:40	LH
3	1:04	1:25	2:07	2:59	1:32	0:57	HL
5	0:50	0:50	0:21	0:20	0:27	0:17	LH
6	2:07	0:59	3:25	1:20	2:54	1:32	HL
7	1:34	0:48	1:41	1:41	1:07	0:55	HH
8	0:38	0:44	0:50	1:33	0:23	0:25	HH
9	2:22	0:52	5:29	1:55	3:45	1:44	HL
10	0:54	0:19	1:31	0:37	2:06	0:45	HL
11	0:39	0:15	4:05	5:51	2:11	1:07	HL
12	2:42	1:54	3:56	1:57	3:42	1:20	HL
13	3:19	0:50	3:27	2:00	2:23	1:38	HH
15	0:51	0:51	3:29	4:43	2:04	1:41	HL
16	1:49	1:25	3:18	0:54	5:34	4:02	HL
17	0:33	0:24	1:11	1:01	1:38	0:28	HL

fore could not be stabilized. Instead, an analysis of the ordering effects of display on mean detection times on fault trials was performed because this analysis was immune to the aforementioned conditions, had the added advantage of the ability to compare ordering, and thus was more informative. Ordering effects were analyzed using a multinomial test. Table 1 shows the means and standard deviations of Detection Time × Display and Detection Time × Trial.

The high-space/low-time display generated the fastest detection time 10 out of 15 times. The probability of this occurring, based on a multinomial test and assuming that a low detection time was equiprobable for all three displays, was $p = .005$.

A multinomial test was also conducted on the ordering of detection time for normal trials, but the results were not statistically significant.

Detection Accuracy

The number of hits, misses, false alarms, and correct rejections were recorded and counted for each display. Table 2 shows the detection responses for all three displays.

A contingency table analysis was used to examine differences between display types.

The frequency of misses across display was not significant, $\chi(2)^2 = 0.68$, $p < .75$. The frequency of false alarms across display was not found to be significantly different, $\chi(2)^2 = 0.56$, $p < .9$. A signal detection analysis showed that participants in all display conditions responded well, with B close to $B_{optimal}$, and there was no notable difference between levels of beta. Sensitivity (d') was slightly lower in the HL display. There is no evidence that this difference was significant. A receiver operating-characteristic curve was plotted and is shown in Figure 9.

Fault Diagnosis Time

The time to diagnose a fault was calculated from the start of the fault to the click of the diagnose button. Again, the trials for which the participant exceeded the maximum amount of time were removed. Also, only those times on trials in which the participant diagnosed the fault completely correctly were counted, because a fast time to achieve an incorrect diagnosis was considered meaningless.

As with detection time, diagnosis time was examined for ordering effects of the means across scenarios. These data are shown in
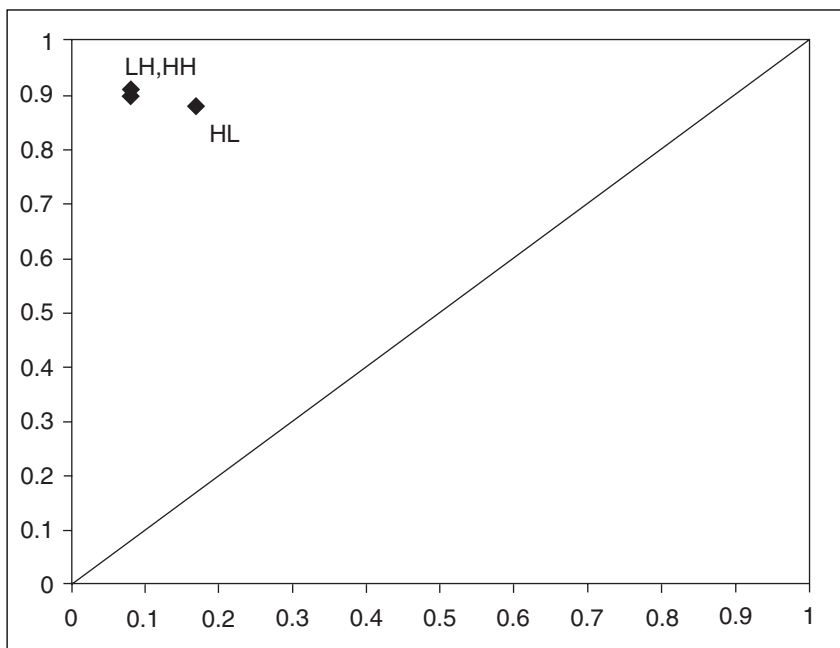

Figure 9. Receiver operating characteristic curve of performance on the three display types.

TABLE 2: Detection Results × Display

	LH		HL		HH	
Response	Fault	Normal	Fault	Normal	Fault	Normal
Fault	81 (.90)	1 (.08)	79 (.88)	2 (.17)	80 (.91)	1 (.08)
Normal	9 (.10)	11 (.92)	11 (.12)	10 (.83)	8 (.09)	11 (.92)

Table 3 and were analyzed using a multinomial test. Again, only correctly diagnosed trials were included. Two trials were poorly handled and had no correct responses. Therefore, the multinomial test was conducted only on those trials that had some correct responses. The high-space/high-time display generated the fastest diagnosis time on 9 of 13 scenarios ($p < .01$).

Diagnosis Accuracy

Diagnosis accuracy was evaluated separately for fault and normal trials. Diagnoses were scored according to the 4-point ordinal scale used by Pawlak and Vicente (1996). This scale gives a score of 0 to a completely missed diagnosis, 1 to a diagnosis referring to symptoms

TABLE 3: Diagnosis Times for all Three Displays

Trial	HL		LH		HH		Fastest Time Display
	Mean	SD	Mean	SD	Mean	SD	
1	4:47	3:40	9:23	3:56	5:17	4:02	HL
2	4:34	0:54	4:47	0:16	4:27		removed
3	4:26	2:08	8:34	3:54	4:37	3:21	HL
5	3:45	1:06	4:06	0:40	1:51	0:54	HH
6	6:43	5:18	10:47	3:49	4:32	2:04	HH
7	3:56	1:46	2:43	1:47	2:19	1:30	HH
8	3:51	2:00	4:29	1:05	2:48	2:07	HH
9	8:24	4:35	8:56	3:14	6:05	4:36	HH
10	7:31	3:28	5:42	2:49	7:54	4:11	LH
11	6:45	5:09	7:20	4:56	4:28	3:53	HH
12	8:00	5:29	8:01	2:52	5:07	3:10	HH
13	4:51		4:31		3:06	1:46	removed
15	8:49	5:30	8:09	4:11	5:10	3:08	HH
16	7:18	5:37	6:56	1:58	7:17	3:08	LH
17	3:47	1:57	4:21	1:05	3:29	1:34	HH

only, 2 to a correct but vague diagnosis, and 3 to a completely correct diagnosis. For a diagnosis to be considered completely correct, participants had to localize the fault to the broken piece of equipment. It is unclear whether or not the 4-point ordinal scale is an equal interval scale. However, the categories of the scale can be applied consistently. Therefore, these data have been analyzed as categorical data. Table 4 shows the frequency of occurrence of diagnosis score across the three displays.

The high-space/high-time display generated more correct diagnoses than did the other two displays (49 correct diagnoses vs. 38 and 30 for the other two displays). This result was statistically significant, $\chi(6)^2 = 14.6$, $p < .025$, suggesting that diagnosis scores of 0, 1, 2, and 3 were distributed differently across the three displays. It appears that the high-space/high-time display generated completely correct diagnoses more frequently than did the other two displays.

DISCUSSION

These findings converge into three major results that relate to the original hypothesis. The results are discussed in two sections in terms of the original definition of "keeping related things together."

Keeping Things Together

In terms of fault detection, the high-space/low-time display generated the fastest detection times. For fault diagnosis, the high-space/high-time display generated the fastest results. Both of these displays benefited from spatial proximity. The one spatially separated display, low-space/high-time, was suboptimal in both tasks. Temporal proximity by itself was not useful, but temporal proximity in conjunction with spatial proximity could have improved performance.

Keeping Related Things Together

The difference between the fault detection task and the fault diagnosis task was that the diagnosis task was a problem-solving task. The detection task was not a problem-solving task in that it did not require reasoning. Although the high-space/high-time display had longer fault detection times, it clearly improved fault diagnosis in both speed and accuracy. This is consistent with the hypothesis that maximal integration in space and time along means-end links should improve problem solving. In addition, given that diagnosis time was measured from the start of the fault, the actual amount of time spent in explicit diagnosis activities with this display was much shorter than the time spent in this regard with the other displays.

The other result of interest was that the difference between the displays was limited to fault trials and did not occur on normal trials. This again suggests that the main effect of the integration was an interaction with the user's problem-solving abilities, given that normal trials required no problem solving.

Interestingly, in terms of data, the low-space/high-time and the high-space/high-time displays showed the same amount of data. In fact, in comparison, the high-space/high-time display was a very cluttered and confusing display. It was not easy to predict that users would perform better with this display. Tufte

TABLE 4: Diagnosis Accuracy × Display

Diagnosis Accuracy	HL	LH	HH	Total × Score
0	8	15	8	31
1	23	19	12	54
2	29	18	19	66
3	30	38	49	117
Total × Display	90	90	88	268

(1983), however, has argued that more information, when displayed correctly, can be better than small amounts of information. In this case, the reason that the high-space/high-time display supported performance better is not that it presented more data or presented it better. The reason is that this display showed more than data: It showed the data in relation to one anther in a meaningful way. This portrayal of relations, beyond merely showing data, is the crux of designing for integration.

ACKNOWLEDGMENTS

REFERENCES

Burns, C. M. (1998). *The effects of spatial and temporal proximity of means-end information in ecological display design for an industrial simulation.* Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

Christoffersen, K., Hunter, C. N., & Vicente, K. J. (1996). A longitudinal study of the effects of ecological interface design on skill acquisition. *Human Factors, 38,* 523–541.

Elm, W. C., & Woods, D. D. (1985). Getting lost: A case study in interface design. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 927–931). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37,* 32–64.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing effective operator control through ecological interface design. *International Journal of Human-Computer Studies, 44,* 653–688.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics, 15,* 234–243.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic trouble shooting. *Ergonomics, 17,* 293–307.

Rasmussen, J., & Vicente, K. J. (1990). Ecological interfaces: A technological imperative in high-tech systems? *International Journal of Human-Computer Interaction, 2,* 93–110.

Roth, E. M., Mumaw, R. J., & Stubler, W. F. (1993). Human factors evaluation issues for advanced control rooms: A research agenda. In *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics* (pp. 254–259). New York: Institute of Electrical and Electronics Engineers.

Torenvliet, G. L., Jamieson, G. A., & Vicente, K. J. (1998). Making the most of ecological interface design: The role of cognitive style. In *Proceedings of the Fourth Symposium on Human-Interaction in Complex Systems* (pp. 214–225). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics Press.

Vicente, K. J., Christoffersen, K., & Pereklita, A. (1995). Supporting operator problem solving through ecological interface design. *IEEE Transactions on Systems, Man, and Cybernetics, 25,* 529–545.

Vicente, K. J., Moray, N., Lee, J. D., Rasmussen, J., Jones, B. G., Brock, R., & Djemil, T. (1996). Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis. *Human Factors, 38,* 506–521.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems: II. Mediating "direct perception" in complex work domains. *Ecological Psychology, 2,* 207–249.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37,* 473–494.

Woods, D. D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies, 21,* 229–244.

Woods, D. D., Roth, E. M., Stubler, W. F., & Mumaw, R. J. (1990). Navigating through large display networks in dynamic control applications. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 396–399). Santa Monica, CA: Human Factors and Ergonomics Society.

Catherine M. Burns is an assistant professor in the Department of Systems Design Engineering at the University of Waterloo, Waterloo, Canada, where she directs the Advanced Interface Design Lab. She received her Ph.D. in mechanical and industrial engineering from the University of Toronto in 1998.