Int. J. Human–Computer Studies (1996) **44,** 653–688

Inducing effective operator control through ecological interface design

WILLIAM S. PAWLAK† AND KIM J. VICENTE‡

Cognitive Engineering Laboratory, Department of Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

(Received 27 July 1994 and accepted in revised form 31 October 1995)

Ecological Interface Design (EID) is a theoretical framework for designing interfaces for complex human–machine systems. This article investigates the utility of EID in inducing effective real-time operator control performance during both normal and abnormal conditions. Two interfaces for a thermal-hydraulic process were compared, an EID interface based on physical and functional (P + F) system representations and a more traditional interface based solely on a physical (P) representation. Subjects were given 4 weeks of daily practice with one of the two interfaces before their performance on normal events and unfamiliar faults was evaluated. Under normal conditions, there was no performance difference between the P + F and P interfaces. However, dual task results indicate that the P interface loads more on verbal resources, whereas the P + F interface loads more on spatial resources during normal trials. Furthermore, a process tracing analysis of the fault trials showed that the P + F interface led to faster fault detection and more accurate fault diagnosis. Moreover, the P + F subjects exhibited a more sophisticated and effective set of fault management strategies that are similar to those observed in field studies of experienced operators in complex human–machine systems. In addition, a deficiency of the P + F interface was identified, suggesting a need for integrating historical information with emergent feature displays. Collectively, these findings have significant practical implications for the design of advanced computer interfaces for complex industrial systems.

1. Introduction

Recently, several authors have suggested that computer interfaces for complex human–machine systems should put greater emphasis on presenting *functional* information to operators (Vicente & Rasmussen, 1990, 1992; Woods, 1991; Bennett & Flach, 1992). The rationale for this recommendation can be best understood by examining the well-known limitations of traditional, single-sensor-single-indicator (SSSI) displays (Goodstein, 1981), which focus primarily on providing operators with elemental *physical* information describing the status of system components. While this type of information is certainly valuable, operators also frequently need

† William S. Pawlak is now with Sterling Software, NASA Ames Research Center, Moffett Field, CA, USA.

‡ Correspondence and requests for reprints should be addressed to Dr. Kim J. Vicente, Department of Industrial Engineering, University of Toronto, 4 Taddle Creek Road, Toronto, Ontario, Canada M5S 3G9. email: benfica@ie.utoronto.ca

1071-5819/96/050653 + 36$18.00/0

to answer questions at a higher level of abstraction (e.g. is the plant in a safe state?). To answer such questions with a SSSI display, operators must determine which data are relevant for the current task, gather the relevant data from individual instruments, and finally, integrate these data mentally to derive the higher-order properties of interest (e.g. mass inventory, heat transfer). These cognitive demands can impose a great burden on operators under the best of circumstances. The situation is even worse under fault conditions, as it may not be possible to derive higher-order properties from lower-level elements (Vicente & Rasmussen, 1990), making it very difficult and sometimes impossible to consistently diagnose the nature of the abnormality. By directly providing operators with functional information (as well as physical information) in the display, advanced approaches to interface design attempt to provide the information that operators need for the entire range of domain demands, thereby reducing the need for mental integration.

One exemplar of an advanced approach to interface design is the theoretical framework developed by Vicente and Rasmussen (1990, 1992), Ecological Interface Design (EID). EID suggests that the higher-order functional constraints that govern the process be made directly available to operators in a manner that would allow them to pick up that information using their powerful capabilities of perception. This approach results in an integrated interface that contains multiple system views, at physical and functional levels of abstraction. Despite the relatively wide-spread agreement that such an approach is needed to combat the limitations of the SSSI approach, there is virtually no empirical research comparing the performance of a multilevel interface based on physical and functional system models with a more traditional interface for a complex human–machine system under representative conditions. The present paper addresses this gap by evaluating the impact of an EID interface on operator performance in a dynamic process control system under both normal and abnormal conditions.

Why would one expect EID to aid performance? Very briefly, providing subjects with a visualization of the higher-order functional domain constraints should result in a 'transparent' interface that should make it easier for subjects to apprehend the current state of the system. In addition, the interface should also reveal to subjects how different system variables are related to each other. Basic research from ecological psychology (Gibson, 1991) suggests that if an interface provides such a rich source of information, then with experience subjects should gradually, but naturally, gravitate towards identifying, and attending to, highly diagnostic higher-order information that can be used to satisfy domain tasks. In other words, an EID interface should make it easier for subjects to adapt to the properties of the work environment in a functional, rather than dysfunctional, fashion. If this is the case, one would expect experienced subjects with an EID interface to inductively develop control strategies that are well tailored to the way in which the system in question actually operates, and therefore achieve a more proficient level of performance.

Before descibing the study that was conducted to evaluate these ideas, previous research on EID will be briefly reviewed. The first experimental evaluation of EID was conducted in the context of DURESS (Dual Reservoir System Simulation), a thermal-hydraulic process simulation (Vicente, Christoffersen, & Pereklita, 1995). A traditional interface containing only physical (P) information was compared to an interface based on EID, which contained both physical and functional (P + F) information. The goal of that study was to compare these two interfaces in terms of how well they support problem-solving behaviour. The experimental evidence indicated that the P + F interface provided better support for problem solving than the P interface. While this finding was encouraging, the study was limited in several ways. First, the subjects were either theoretical experts or novices at generic thermal-hydraulic principles, but neither group has any substantial experience with DURESS itself. Second, the experiment evaluated the subjects' ability to diagnose and remember the values from dynamic, "canned" scenarios. Therefore, subjects did not interactively control the system. While there were reasons for conducting the experiment under these restricted conditions (cf. Vicente, 1991), it remained to be seen whether the advantage of the P + F interface would also hold with subjects who have extensive experience controlling DURESS.

The study presented in this article was conducted to address this question. First, an updated real-time interactive simulator, DURESS II, was used as a research vehicle, thereby providing subjects with continuous closed-loop control of the system. Second, subjects were given extensive experience at controlling the system, thereby giving them a chance to adapt to system structure and dynamics. This allowed us to compare the performance of subjects who were experienced with the P + F or P interfaces, under both normal and abnormal operating conditions, in a more representative manner than in previous research.

The next section will describe the DURESS II system, as well as the P and P + F interfaces that were designed for DURESS II.

2. DURESS II: a research vehicle

DURESS II is similar in most respects to the original DURESS system used in earlier research (e.g. Bisantz & Vicente, 1994; Vicente, Christofferson, & Pereklita, 1995), but a few structural modifications have been made. The physical structure of DURESS II is illustrated in Figure 1. The system consists of two redundant feedwater streams (FWSs) that can be configured to supply water to either, both, or neither of the two reservoirs. Each reservoir has associated with it an externally determined demand for water that can change over time. The system purposes are twofold: to keep each of the reservoirs at a prescribed temperature (40°C and 20°C), and to satisfy the current mass (water) output demand rates. To accomplish these goals, the subject has control over eight valves (VA, VA1, VA2, VO1, VB, VB1, VB2, and VO2), two pumps (PA and PB), and two heaters (HTR1 and HTR2). All of these components are governed by first order lag dynamics, with a time constant of 15 s for the heaters and 5 s for the remaining components. The system input temperature (T0), reservoir output temperatures (T1 and T2), and the volumes for both reservoirs (V1 and V2) are also displayed in Figure 1.

The remainder of this section describes the two interfaces used in this research. These are updated versions of the P and P + F interfaces used in previous research on DURESS [see Vicente and Rasmussen (1990) for a detailed description of their design rationale].

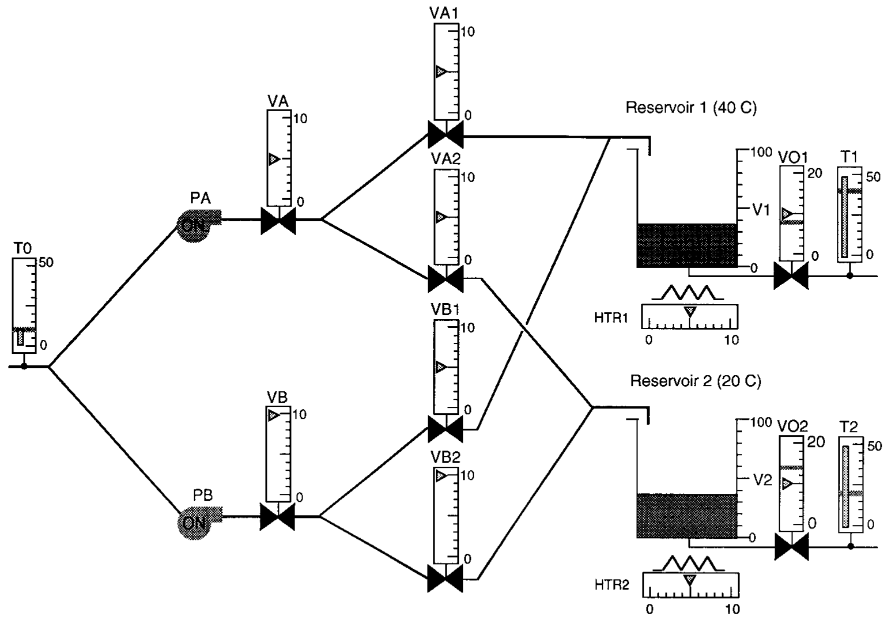

FIGURE 1. P interface for DURESS II.

2.1. P INTERFACE

The P interface, illustrated in Figure 1, provides a physical representation of DURESS II. It displays only the state of the physical components and the goal variables. The first meter on the extreme left of the display is a water thermometer (T0), which measures the temperature of the water entering the system. This thermometer displays temperature as a vertical bar (coloured red in the actual display) that increases in height as the water temperature increases. The normal inlet water temperature is 10°C, as indicated by the thin area on the T0 scale (coloured green in the display). After the thermometer, the input water stream splits and flows to two pumps (PA and PB). The maximum flowrate through each pump is 10 units/s. The pumps operate as discrete switches and are either on or off. The subject uses a mouse to click on the pump to change its state. The pumps are displayed in black (with white lettering) if they are off, and in light grey (with black lettering) if they are on. If either pump is turned on without any of the downstream valves being opened, the pump will fail after approximately 5 s. This error terminates the trial.

The next set of components are the primary valves (VA and VB), which have a continuous range from 0 to 10. The valve state is set using a mouse to either drag the yellow triangular pointer to the desired setting, or to simply click on the scale at the desired point. From these pirmary valves, each FWS splits into two secondary valves connecting each stream to both reservoirs. The secondary valves (VA1, VA2, VB1, and VB2) operate in the same manner as the primary valves. The water then flows to each of the two reservoirs, where it is heated and removed, through the use of the heaters (HTR1 and HTR2) and the output valves (VO1 and VO2), in order to meet the temperature and demand goals, respectively. The reservoirs have a maximum capacity of 100 units. Reservoir volume levels are indicated by a scale on the side of each reservoir and by the shaded area (blue in the display) depicting water in the reservoir. It is possible to overflow either of the reservoirs, if input flowrate is consistently greater than output flowrate. When reservoir volume exceeds the maximum capacity of 100 units, the trial ends automatically.

The heaters (HTR1 and HTR2) also have a continuous range of 0 to 10. The subject can either slide the triangular pointer (red in the display) on the heater scale to the desired setpoint, or click on the scale itself at the desired point. Heating an empty reservoir for an extended period will lead to a malfunction. Thus, if there is continued heat transfer to a reservoir without any water in it, then the system will eventually fail and the trail will end.

The water temperature in the reservoirs is displayed with thermometers (T1 and T2). The goal temperature is represented as a thin area (green in the display) on the temperature scale. There is a tolerance of ±2°C from the setpoints (40°C for Reservoir 1 and 20°C for Reservoir 2). If the water in the reservoir boils, the system fails and the trial ends.

Finally, the subjects also have control over the outlet valves (VO1 and VO2) that are used to meet the demand goals. These valves operate in the same manner as the other valves, except that their maximum setting is 20. The demand for each reservoir is indicated by a thin area (green in the display) on the outlet valve setting scale. This goal area, which is ±1 unit around the desired level, moves as a function of changes in the demand.

This interface design format was chosen because it is typical of how existing computer interfaces for process control systems have been designed. To some readers, it may seem like a straw man because it is sparse in content. However, this is not atypical of industrial systems. For example, the Three Mile Island control room did not display the flowrate from the emergency auxiliary feedwater system to the steam generators (Malone, Kirkpatrick, Mallory, Eike, Johnson, & Walker, 1980). Similarly, the Biblis plant in Germany did not have a pressure sensor for the isolation valve separating the primary system from the low pressure injection system, and as Becker (1991) points out, this led to the flow of radioactive water to the outside of the containment. Other sub-systems that are poorly instrumented and that occasionally give rise to complications can also be found in industrial systems (e.g. Vicente & Burns, 1995). Given that such examples exist, that it is precisely these subsystems that give operators problems in fault situations, and that virtually no prior research has been conducted on the effects of physical and functional information on operator performance, we feel that the P interface is not a straw man. Thus, it serves as a meaningful baseline condition.

2.2. P + F INTERFACE

The P + F interface, based on the principles of EID, is illustrated in Figure 2. A detailed explanation of how this interface was designed would be quite lengthy, and therefore cannot be provided here (the interested reader is referred to Vicente & Rasmussen, 1990). Only a description of the interface itself will be provided. The

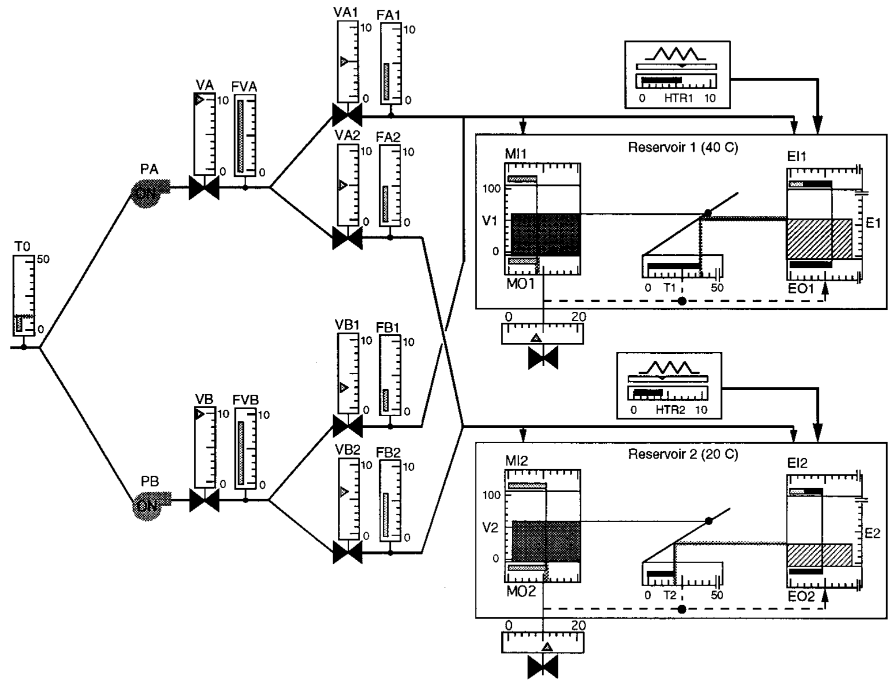

FIGURE 2. P + F interface for DURESS II. This Figure shows the system in steady state, and therefore provides a baseline against which to compare Figures 7 and 8.

input water thermometer, both pumps, and all of the valves operate in the same manner as in the P interface. However, the P + F interface also contains higher-order functional information identified through an abstraction hierarchy analysis of DURESS (see Vicente & Rasmussen, 1990) that describes the state of the functions that the physical components are intended to achieve. Thus, each valve also has a flow meter next to it (FVA, FVB, FA1, FA2, FB1, FB2, and MO1 and MO2 for the mass output flowrates). These flow meters have the same value range as their respective valves. The vertical bar in each meter is yellow, the colour used throughout the interface to indicate both valve settings and flowrate values.

The boxed group of graphics on the right of Figure 2 represent the most innovative portion of this interface, providing additional higher-order functional information in the form of first principles (i.e. mass & energy conservation laws). The rectangular graphic on the left represents the mass balance for the reservoir, and the rectangular graphic on the right represents the energy balance. The display in the centre with the diagonal line shows the relationship between volume, energy inventory, and temperature. Each of these three displays will now be described in turn.

Figure 3 provides a close-up view of the mass balance display for Reservoir 1. The various inputs are shown at the top of the graphics (MI1). The volume level is indicated by the scale on the left side of the graphic (V1) as well as the shaded area of the container, which is coloured blue. The output, MO1, is shown at the bottom of the graphic. The sloped line connecting the input and output relies on a funnel

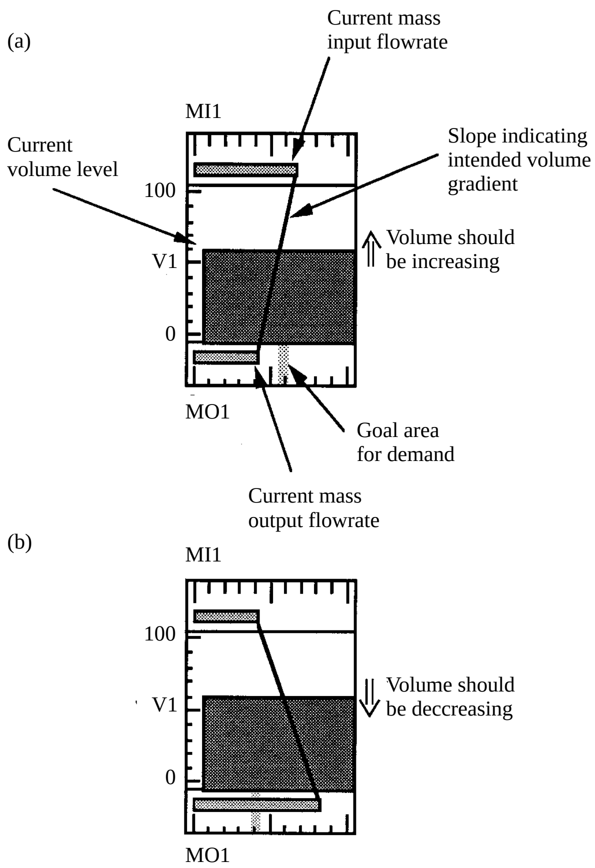

FIGURE 3. Close-up view of the mass balance display in the P + F interface. When input is greater than output (a), the slope is positive indicating that the volume should be increasing. When input is less than output (b), the slope is negative indicating that the volume should be decreasing.

metaphor. For example, if the top is wider than the bottom (i.e. input > output, as in Figure 3(a)), then it is easy to visualize the consequence, namely, that volume should be increasing. Conversely, if the bottom is wider than the top (i.e. output > input, as in Figure 3(b)), then volume should be decreasing. Thus, the slope of the line represents the rate at which the volume *should be* changing.

The next display, showing the energy balance, is illustrated in Figure 4. This display is located on the extreme right in the P + F interface (see Figure 2). The energy balance display functions in a very similar manner to the mass balance display. As shown in Figure 4(a), the primary difference is that the energy input to the reservoir (EI1) is partialled out according to its two contributors. The rate at which energy is added by the FWS (i.e. temperature of the water, in degrees C, multiplied by its flowrate) is shown as a lightly shaded bar coloured yellow in the display, and the rate at which energy is added by the heater is shown as a dark red bar. The energy inventory area in the interface (E1) is coloured orange. Again, a

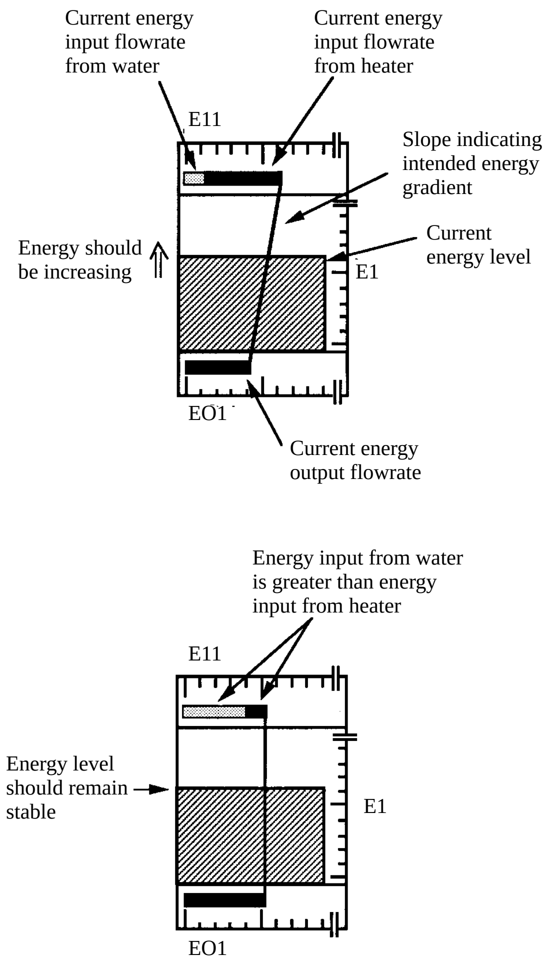

FIGURE 4. Close-up view of the energy balance display in the P + F interface. This display functions in the same manner as the mass balance display, except that there are two energy input sources, the water coming into the reservoir (lightly shaded bar) and the heater (darkly shaded bar). Figure (a) shows the case where the heater is adding energy at a greater rate than the incoming water. Also, in this case, the energy inventory should be increasing because the energy input is greater than the output, as indicated by the positive slope. Figure (b) shows the case where the incoming water is added energy at a greater rate than the heater. Also, in this case, the energy inventory should be stable because the energy input equals the energy output, indicated by the perpendicular line.

funnel metaphor is used, so that, for example, if the total energy input flowrate is greater than the output (as in Figure 4(a)), then energy inventory should be increasing. Figure 4(b) shows the same display in a different state. In this case, the incoming water is adding energy at a greater rate than the heater. Another difference is that the total energy input flowrate now equals the energy output flowrate, so the sloped line is perpendicular, indicating that the energy inventory should remain stable.

Now that these two displays have been thoroughly illustrated, the third display located between the mass and energy balance displays in Figure 2 can be described.

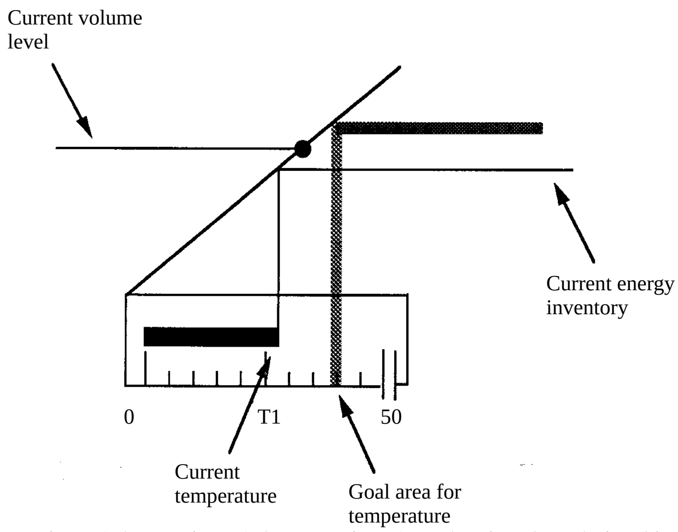

FIGURE 5. Close-up view of the portion of the P + F interface showing the relationship between volume level, energy inventory level, and reservoir temperature.

A close-up of this display, which shows the relationship between volume, energy, and temperature, is provided in Figure 5. The horizontal line with a ball on the end that emanates from the current volume level (V1) is coloured light blue. Change in the height of this line always accompanies any change in mass inventory (i.e. the bar will always be at the same height as the water level, V1). The diagonal line in the centre display rotates about its leftmost endpoint (connected to the top left of the T1 box) and is always tangent to the ball on the end of the horizontal line. The slope of the diagonal line represents the function that maps the relationship between volume and energy onto temperature. This mapping is indicated by the line emanating from the current energy inventory level (E1) that comes across and reflects off of the diagonal line at a right angle down onto the current value on the temperature scale (T1). The goal temperature is indicated by the thin shaded area (coloured green in the display) on the temperature scale. This goal area reflects up from the temperature scale, off of the diagonal line, and onto the energy inventory scale. As a result, the energy level required to achieve the goal temperature is directly visible.

Figure 6 shows how this display reveals changes in system state. In Figure 6(a), the energy inventory is being held constant while the volume is increased slightly. The increase in volume causes the horizontal line emanating from the volume level to go up. This change in vertical position, in turn, causes the diagonal line in the centre display to rotate counterclockwise towards the vertical, increasing the slope of the diagonal line. The energy inventory level is at the same height but it now intersects the diagonal line at a different point. As a result, the new temperature is now lower than the old temperature, which makes intuitive sense since the energy per unit volume is lower than before, due to the increase in volume. Note also that the energy level required to achieve the same goal temperature is now greater than before because of the increase in volume.

Figure 6(b) shows how this display changes when volume is held constant and

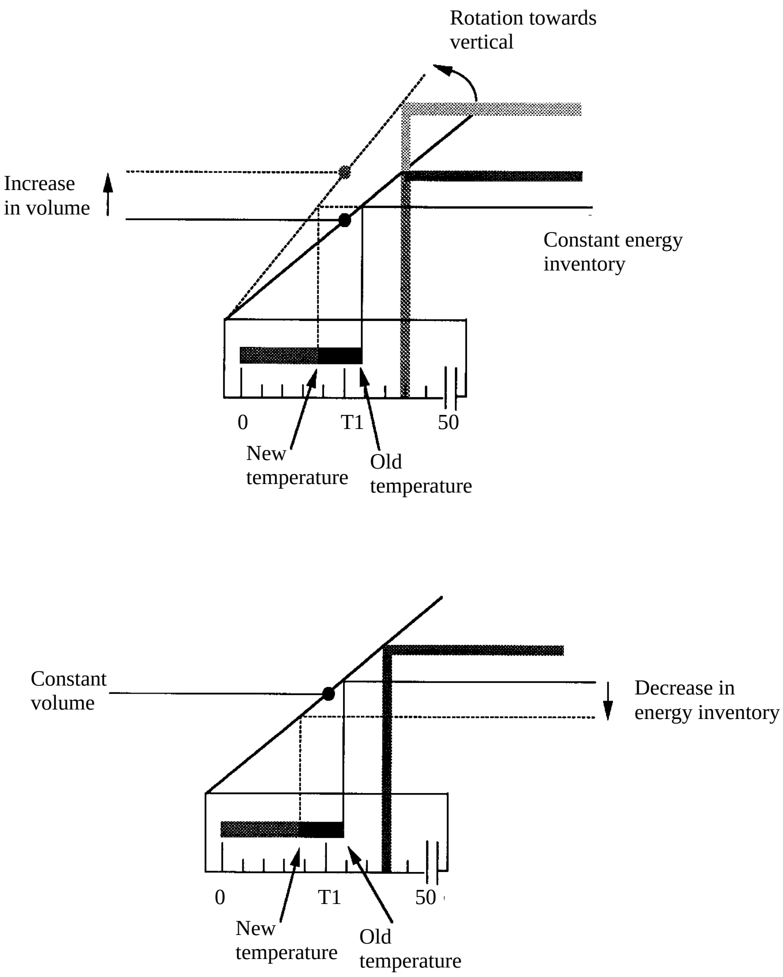

FIGURE 6. The dynamics of the display in Figure 5. Figure (a) reveals how the display changes when energy inventory is held constant while volume is increased, leading to a decrease in temperature. Figure (b) reveals how the display changes when volume is held constant while energy inventory is decreased, leading to a decrease in temperature.

energy inventory level is decreased. In this case, the change is much simpler. Because volume is constant, the height of the horizontal bar is fixed, which in turn, fixes the slope of the diagonal line in the centre. The only change that occurs when energy inventory is decreased is that the horizontal line emanating from the far right goes down, which means that it intersects the diagonal at a different point, indicating by the dashed line in Figure 6(b). Again, the new temperature is lower than the previous one because there is now less energy in the same amount of water.

The three displays described in isolation in Figures 3 to 5 are integrated in a coordinated fashion in the P + F interface in Figure 2 (for a detailed description of the *rationale* behind the design of the P + F interface, see Vicente and Rasmussen,

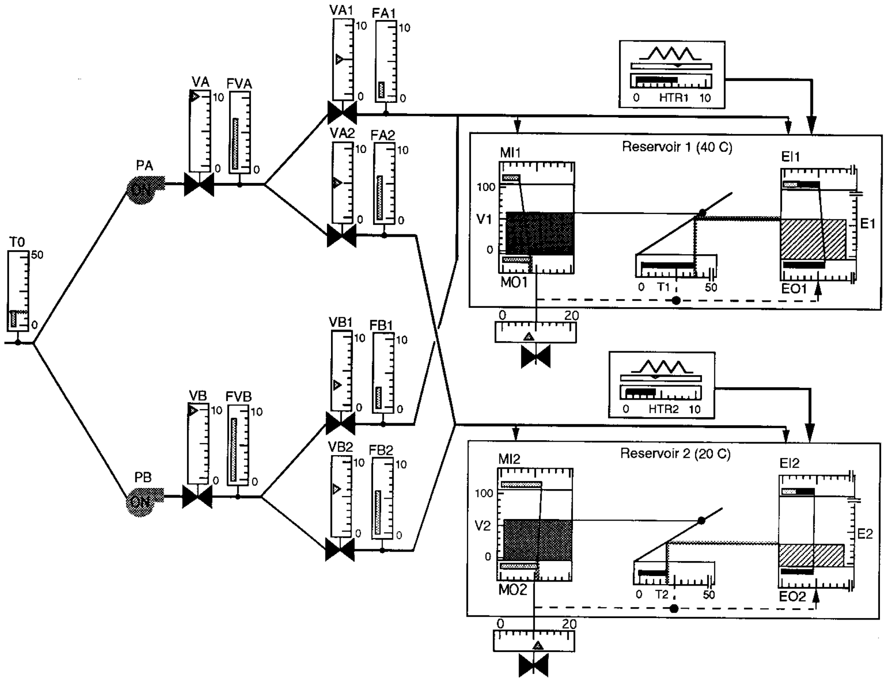

FIGURE 7. P + F interface during a transient showing the dynamic content of the interface. This snapshot was taken 5 s after the beginning of a block of valve VA1.

1990). Note that off-scale markers are represented in the output temperature scales and the energy input, inventory, and output scales as well. These were added to the interface by creating a gap in the scale at the off-scale point, thereby allowing subjects to discriminate the maximum value from off-scale (Mumaw, Woods & Eastman, 1992).

An impression of the dynamic content of the P + F interface can be obtained by viewing Figures 2, 7, and 8, which show the successive changes caused by a transient. The figures illustrate the progression of events before and after a block in valve VA1. In Figure 2, the system is at steady state. Figure 7 illustrates the state of the system 5 s after the onset of the fault, and shows a number of significant changes. FA1, the flow through the blocked valve, has started to decrease, as has FVA. This reduction in inflow, in turn, causes the mass and energy inputs to Reservoir 1 (MI1 and EI1, respectively) to decrease. As a result, the input flowrates for both mass and energy are now less than their respective output flowrates, thereby causing the two lines connecting the input and output flows to deviate from vertical. Using the funnel metaphor on which these balance displays are based, the bottom is now wider than the top, so the two slanted lines indicate an expected negative gradient which should cause the levels to decrease. The decrease in FA1 also causes a temporary surge in FA2 which propagates to a slight increase in MI2 and EI2. In this case, the balance lines are slightly tilted in the opposite direction (i.e. inputs slightly greater than outputs), indicating an expected small positive gradient. Figure 8 shows the state of the system 20 s after the fault, by which point VA1 has become completely

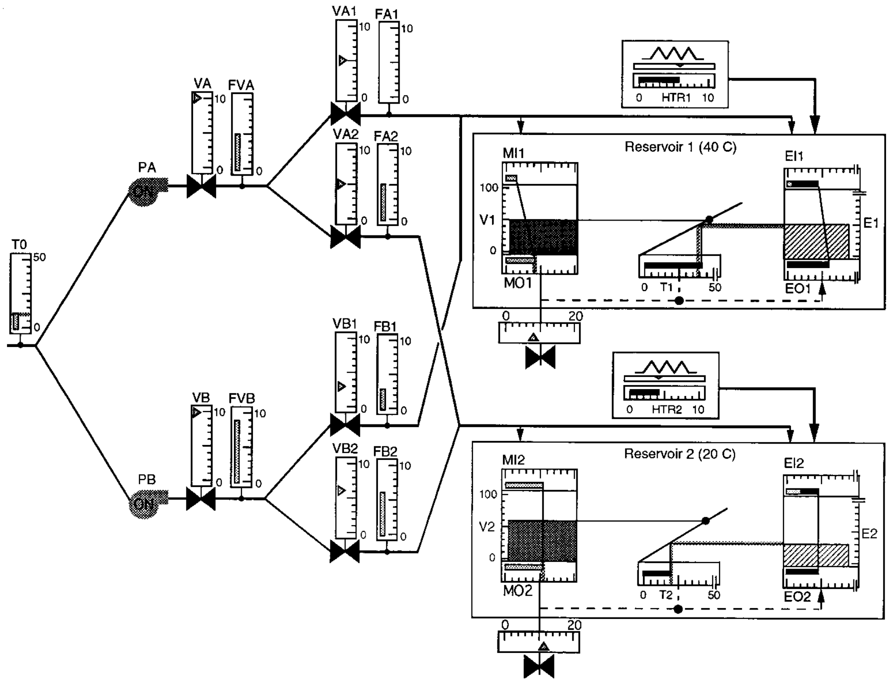

FIGURE 8. P + F interface during a transient showing the dynamic content of the interface. This snapshot was taken 20 s after the start of the fault, at which point valve VA1 had become completely blocked.

blocked. This fact is clearly shown by the fact that VA1 is open but there is no flow through the valve (i.e. FA1 = 0). As a result, FVA has decreased as well, now supplying only the rate being drawn through VA2. Because FWS A is no longer supplying water to Reservoir 1, MI1 and EI1 have decreased considerably. Examining EI1, the fact that the lightly shaded bar (representing the contribution of the incoming water) is much smaller than before and the dark bar (representing the contribution of the heater) is the same length as before indicates that the reduction in EI1 is caused by a reduction in the flow of water, not a reduction in the transfer of heat. This reduction in inflow causes the sloped lines for the mass and energy balances to be even more negative than in Figure 7. As expected, the volume (V1) and the energy inventory (E1) have decreased as a result. Furthermore, because there is no longer any cold water coming into Reservoir 1 and the volume is decreasing, the temperature (T1) has increased and is now outside of the goal region.

In summary, Figures 2, 7, and 8 give an idea of how the P + F interface reacts dynamically to a fault. A number of changes indicating propagation in time and space were illustrated. These changes are intended to aid performance, but it is equally possible that they can confuse subjects and thereby have no effect, or even a negative effect, on performance. Thus, it is important to empirically compare the P and P + F interfaces to evaluate their relative worth. More specifically, it is worthwhile investigating how subjects use a multilevel functional interface, such as the P + F, and how such an interface can affect operator strategies.

3. Method

This experiment compared the P and P + F interfaces in terms of their ability to support practiced subjects in dealing with both normal and abnormal scenarios. According to the ecological psychology literature, adaptation consists of training one's attention to search for and attune to goal-relevant higher-order invariants. If this is so, then making these invariants directly observable, as the P + F interface does, should constrain adaptation in a functional manner, thereby allowing subjects to induce more effective control strategies. In contrast, the P interface only presents a subset of the goal-relevant system constraints. Subjects are not provided with comprehensive feedback on how the system operates, and so adaptation is expected to suffer as a result, especially during fault situations (cf. Vicente & Rasmussen, 1990).

3.1. EXPERIMENTAL DESIGN

A repeated measures, between-subjects design, with the type of interface (P or P + F) as the primary manipulation, was adopted for the experiment. Subjects were assigned to one of the two interfaces, and participated for about an hour a day for a total of 27 days. Both thinking aloud and behavioural protocols were collected.

3.2. EXPERIMENTAL TASKS

During the experiment, subjects were asked to perform four different control tasks.

(1) *Start-up.* For this task, the subject was presented with a shut-down system and was asked to bring the system to steady-state, meeting pre-defined setpoints consisting of temperature and demand goals.

(2) *Tuning to new setpoints.* In this task, the subject needed to bring the system from an on-line, steady-state initial condition to a pair of new steady-state demand setpoints.

(3) *Shut-down.* During this task, the subject was required to bring the system from an on-line, steady-state condition to a shut-down condition.

(4) *Fault Management.* Once subjects had extensive practice at controlling the system, they were occasionally presented with trials during which a fault would occur. Subjects had to detect, diagnose, and compensate for the presented fault.

For all control tasks, steady-state was defined as maintaining both reservoirs in the goal areas (both temperature and output demand) for five consecutive minutes.

Twice during the experiment, subjects were also required to control the system while performing a secondary task, one verbal and the other spatial. Both types of secondary tasks were selected because of the relative independence of verbal and spatial information processing resources (Wickens, 1992). By intentionally loading one of these two resources with a secondary task, and examining the resulting effect on primary task performance, it is possible to infer which type of cognitive processing is being most heavily utilized by the primary task. For example, if one adds a spatial loading task to the primary task and performance is found to degrade significantly on the primary task, one could conclude that the primary task draws on spatial resources. This technique can also be used to compare interfaces in terms of the extent to which they comparatively load on either verbal or spatial resources.

Note, however, that comparison across secondary task types is not meaningful since the verbal and spatial secondary tasks are not of equal difficulty.

The specific secondary tasks chosen for this study have been frequently used in experimental psychology research and are known to load on either verbal or spatial resources. A variation of Brooks' (1968) mental imaginary task was used as the spatial loading task. This requires subjects to imagine that they were navigating around the perimeter of certain letters of the alphabet and state in which direction (right or left) they were turning each time they came to the end of a straight edge. A verbal repetition task used by Saariluoma (1992) was adopted for the verbal loading task. For this task, subjects were asked to continuously repeat a nonsense word, "tiikuri," while they performed the experimental task. An experimenter ensured that subjects repeated the word continuously. Both secondary tasks were administered as loading tasks (O'Donnell & Eggemeier, 1986), i.e. subjects were instructed to devote their attention to the secondary task, and only perform the primary task when they were able.

3.3. SUBJECTS

All subjects were undergraduate or graduate engineering students at the University of Toronto. While subjects' specific educational backgrounds varied, an effort was made to ensure that subjects with similar backgrounds were paired and assigned to different interfaces. Subjects using the P interface had taken an average of two university physics courses and 1.75 thermodynamics or thermal-hydraulics courses. Subjects using the P + F interface had taken an average of 2.33 university physics, and 1.5 thermodynamics or thermal-hydraulics courses. The average of the subjects was 23.5 years for the P interface users, and 22.5 years for the P + F interface users. There were six subjects assigned to each interface. The 12 subjects, all males, received $5.00 for each session, plus a bonus of $2.00 per session for completing the experiment.

3.4. APPARATUS

The DURESS II simulation runs on a Silicon Graphics IRIS Indigo R4000 computer workstation. The simulation code was written in C, while the two interfaces were designed using a graphical construction set called FORMS. All data collection and analyses were performed using both the IRIS computer and an Apple Macintosh IIvx computer. Verbal protocols were collected using a Sony CCD-TR81 Hi-8 Handycam.

3.5. TRIAL TYPES

Trials consisted of the four control tasks described earlier, performed either in isolation or in sequence in the same trial. The trials progressively increased in difficulty during the experiment. Initially, subjects were required to perform a start-up task alone. Then, they were required to start-up the system, and then shut it down after they achieved steady state. Next, subjects had to start-up the system to one set of demand setpoints, achieve steady state, and then tune the system to a second set of demand setpoints and achieve steady state again. The most involved trial type consisted of start-up, tuning, and shutdown tasks, all in the same trial. In addition, some later trials required subjects to deal with a fault while they were doing a start-up task. Regardless of its type, the trial ended when subjects fulfilled the task(s) required for that trial.

Setpoints for startup and tuning trials were varied to keep subjects from consistently adopting simplistic control strategies. New tuning setpoints were chosen to require the use of multiple or complex control strategies within a trial. Also, the demand pairs were selected to gradually increase in complexity over sessions (as determined by a cognitive work analysis [CWA] of DURESS II, Vicente & Pawlak, 1994). Faults were chosen from a number of possible faults available in the simulation program. They occurred after some fixed time within a session. Only one fault occurred within a particular session.

3.6. PROCEDURE

The entire experiment consisted of one introductory session, 18 skill-acquisition sessions (composed of 33 trials), and eight data-collection sessions (nine trials) for a total of 27 sessions (42 trials). Each session took place on a separate day.

3.6.1. Introductory session

The introductory session was the same for both groups. Subjects were presented with a description of the experiment, were asked to complete an informed consent form and a demographic questionnaire, and were provided with a technical description of DURESS II. Results reveal that no significant difference existed between the two groups ($t(5) < 1$, n.s.), indicating that they did not differ in terms of thermal-hydraulics background knowledge. Subjects were then given a complete list of the system variable names so that they could become familiar with the labels. A pre-test questionnaire consisting of 20 questions (see Vicente, 1991) was then administered to assess subjects' prior knowledge of thermal-hydraulics. Thirty minutes were allotted for completing the questionnaire.

Subjects were then administered a spatial task, based on studies conducted by Brooks (1968) and Saariluoma (1992), to be used as a control condition for the spatial component of the dual task section of the experiment. This test required subjects to listen to an audio-tape of the experimenter reading off two sets of 25 pairs of upper-case block letters, presented at 15 s intervals. Subjects were to imagine standing in the lower left-hand corner of each letter and to "walk" around the letter. Subjects were to walk in a clockwise direction for the first letter, and a counter-clockwise direction for the second letter. At each turn, they were to state whether they were making a right or left turn. Before the test, subjects were given a sheet of paper with these block letters illustrated (N, L, H, T, K, Z, F, E, X, A, and M) so that they would understand the precise shape of the letters. They had 1 min to study the formation of these letters. The individual letters were randomly combined to create the letter pairs. Because the task was paced, performance could only be evaluated by the percentage of letters completed. For each letter, subjects receive one full point for a complete, correct letter, one-half of a point for an incomplete but correct letter, and zero points whenever an incorrect turn was made.

Next, subjects were given information on the verbal protocol procedure that was to be used periodically throughout the experiment. Each subject was then given a description of the interface they would be using for the experiment, as well as the tasks they would be performing. Subjects were told that in the case of a fault, they were responsible for detecting, diagnosing, and compensating for that fault in order to continue to meet the prescribed demands. Subjects were not told what types of faults could occur, nor when they would occur, nor how frequently. Moreover, they were also not told that the system would fail if they performed certain actions or put the system in certain states (see previous section). Faults were not mentioned after the introductory session.

3.6.2. Skill-acquisition sessions

The skill-acquisition process consisted of an incremental progression of 33 trials. Each class of trials was designed to be slightly more complex than the previous class. The same trials were presented in the same order to all subjects. No fault trials were presented in this phase of the experiment. Knowledge of results was not given to allow subjects to adopt their own performance criteria. However, a clock showing the elapsed time for each trial was always present in the upper left corner of the screen in both interfaces. The duration of the skill acquisition phase was estimated based on informal pre-testing. The data show that trial times on the last three consecutive trials were within, at the most, 1.3 S.D.s for each subject.

The first session in each class of skill-acquisition sessions was videotaped so that verbal protocols could be collected. These protocols were not analysed. However, they gave subjects an opportunity to become adept at verbalizing their thought processes as they controlled the system. This practice ensured that the verbal protocols for the subsequent data collection sessions would be as non-intrusive as possible.

3.6.3. Data-collection sessions

There were eight data collection sessions, consisting of nine trials. Only data from these trials were analysed. All subjects received the same nine trials in the same order. The first of these was a control condition for the dual task trials. Subjects were required to perform a start-up task followed by a tuning task. The next two trials required subjects to perform the same control tasks while performing a verbal or spatial loading task. The order of these loading tasks was counterbalanced. These three trials were of comparable complexity (as determined by a CWA of DURESS II, Vicente & Pawlak, 1994), so as to allow for meaningful comparisons between the dual task trials and the control condition trial.

Three of the remaining data collection trials required subjects to perform a start-up task in the presence of a fault. The faults were as follows:

(1) valve VA1 became blocked at 4 min;
(2) Reservoir 2 developed a leak of 4 units/s at 1 min;
(3) valve VA1 became blocked at 6 min.

The last fault scenario was of the same type as the first to see if subjects would perform better, or differently, the second time they encountered a fault of the same time. Verbal protocols were collected during all three fault trials.

The other three data collection trials required subjects to perform start-up, tuning, and shut-down tasks in isolation to obtain an independent and unconfounded evaluation of performance with each trial type. Verbal protocols were collected during all three trials.

3.7. PERFORMANCE MEASURES

There were two primary sources of data in the experiment, time-stamped data logs and verbal protocols. With regard to the former, the simulation logged all of the subjects' control inputs, as well as the time and the current values of all of the system variables. As for the latter, every data-collection trial (except the dual-task trials) required subjects to give verbal protocols while controlling the system. This verbal protocol was then used as an additional data source to allow us to interpret and understand subjects' behaviour, both qualitatively and quantitatively.

In addition, a trial reply program was used to provide us with an informal, intuitive understanding of subjects' behaviour. A program, Dplayer, was developed which would allow the experimenters to replay each subject's recorded log file through the DURESS II simulator. Dplayer allowed experimenters to fast-forward and rewind through each file as desired, while viewing the actual screen that subjects used for the simulation. It also interpolates the values of system variables between time-stamped control inputs to allow for a continuous replay of the trial. Each logged control input is highlighted during playback by a large green arrow appearing on the display where the mouse pointer was located when the control input was recorded. This allowed the experimenter to plainly see the control inputs performed by subjects, the magnitude and direction of those inputs, and the state of the system, all on the very same interface that subjects used, thereby reconstructing the system context at the time of subjects' behaviour.

4. Results

This section is divided into four parts describing results from the fault management trials, dual tasks, startup trial, and shutdown & tuning trials (for a more detailed presentation of the results from this experiment, see Pawlak, 1994). Before these results are presented, however, the general approach to statistical analysis that was adopted in this study will be briefly discussed.

In this experiment, we tried to strike a balance between external validity and statistical reliability. On the one hand, the principle of *representative design* (Brunswik, 1956) was rigorously applied to improve the generalizability of experimental results to operational settings. This choice greatly increased the complexity and duration of the experiment. As a result, only a relatively small number of subjects could be included in each group in the time available. While this makes it more difficult to obtain statistically significant effects, we tried to compensate by analysing each individual subject's data in great detail. Where appropriate, statistical tests were performed. However, due to the small sample size, the large degree of variability between subjects, and (in some cases) missing data points, we frequently had to rely on non-parametric statistics (Siegel, 1956).

4.1. FAULT MANAGEMENT TRIALS

As already mentioned, the primary goal of this study was to investigate the relative impact of the P and P + F interfaces on operator strategies and performance. The most important and unique results were obtained from the fault management trials, so these results will be presented first. Recall that no faults were presented in the skill acquisition phase of the experiment, so these trials represent the first time that subjects encounter faults. As a result, these data provide an opportunity to observe each group learning how to manage faults with their respective interfaces.

Analysis of the fault trials began with a process tracing analysis of each fault scenario using both verbal protocol analysis and the Dplayer analysis tool. The decision making categories in Rasmussen's (1976) decision ladder were used to classify subjects' verbalizations and actions. This resulted in eight categories: activation, observation, identification, interpretation, evaluate, define task, formulate procedure, and execution. In order to simplify the analysis, these eight categories were then aggregated into three general phases of fault management behaviour: detection, diagnosis, and compensation (cf. Roth, Woods & Pople, 1992). The data from each trial were then represented in the form of a timeline-based data table, which allowed for a relatively structured and meaningful interpretation of the trials. An example of such a table, along with a summary of the coding scheme is presented in the Appendix. A total of 36 trials were analysed (three fault trials for each of the 12 subjects). Fault analysis tables for each of these trials can be found in Pawlak (1994).

From these data tables, information about each interface group's performance on each fault trial was extracted. The goal of this qualitative analysis was to develop a coherent account of the performance of the interface groups on each fault trial. This qualitative analysis was supplemented by quantitative analyses of detection and diagnosis performance. Time of detection was measured from the subjects' verbal reports. Randomization tests were performed to see if there were differences between interface groups on the time to detect a fault. Diagnosis was scored according to a 0 to 3 point rating scheme (cf. Sicard & Siebert, 1988). The scoring criteria were as follows.

0 Subject says nothing relative to the fault, or says nothing at all.

1 Subject gives a vague but correct description of the effects of the fault (e.g. "Reservoir 1 is going down · · · ").

2 The subject correctly states the symptoms of the fault, but at a more specific and functional level (e.g. "I don't seem to be getting any flow from this line here · · · ").

3 The subject correctly localizes the faulty component (e.g. "VA1 is blocked").

In addition to the diagnosis scores, the total time to complete the trial was also measured. This provided a rough measure of compensation. It is important to note, however, that this measure is confounded by detection and diagnosis times. This is unavoidable because, in closed loop control, it is impossible to decouple compensation from other decision activities. Some of the subjects ended the trial with a system failure, and these failures are identified in place of the total trial time. The system failures observed in this study are described according to the following abbreviations.

- R1 HE – Reservoir 1 (2) was heated empty. The heater was transferring heat to an empty reservoir.
- PA BLOW UP – Pump A failed due to the fact that the pump was in operation, with the downstream valves being closed.
- R1 OVERFLOW – Reservoir 1 was filled to greater than 100 units.

The findings for each fault trial are first presented individually, followed by an

Table 1
Detection, diagnosis, and compensation results for fault 1 (VA1 blocks at 4 min)

		Detection after fault (s)	Diagnosis score	Correct diagnosis (s)	Difference (Dia–Det) (s)	Compensation (s)
	AM	25	2	—	—	PA BLOW UP
	CN	4	3	43	39	625
P + F	EC	5	1	—	—	646
	NJ	22	3	54	32	681
	PL	6	3	83	77	691
	RL	19	0	—	—	PA BLOW UP
	AT	—	0	—	—	779
	DC	35	1	—	—	PA BLOW UP
P	HS	51	3	103	52	707
	MC	112	0	—	—	PA BLOW UP
	SA	116	3	143	27	746
	WN	—	0	—	—	PA BLOW UP

integrative summary of the performance of each interface group across all fault trials.

4.1.1. Fault 1

In fault 1, valve VA1 became blocked 4 min after the start of the trial. There was no difference between the two groups in terms of the number of people who had stabilized the system when the fault occurred (four for each group). The detection times, diagnosis scores, and compensation times for each interface group are presented in Table 1. A randomization test indicates that the P + F interface group had significantly faster detection times than the P interface group ($p = 0.006$, one-tailed).

After fault detection, all but one of the P + F subjects attempted to compensate for the fault by initially increasing the setting for valve VA1. Due to the fault, however, this valve was no longer functioning. Because this initial compensation effort failed, subjects then attempted a diagnosis of the problem. Once fault diagnosis was complete (correct root cause diagnosis by subjects CN, NJ, and PL; partial diagnosis by EC), the subjects were able to correctly formulate and execute a compensation plan. Two subjects in the P + F interface group, AM and RL, ended the trial with a system break; neither had diagnosed the problem correctly.

To understand how the P group managed this particular fault, it is helpful to consider the typical control behaviour of these subjects under normal system conditions. When a subject increases an input valve setting, he usually observes the system to see if the action he performed had the intended effect on the system. He uses this information to determine if another control action (possibly another increase in the same valve setting) is necessary. Since the subject only has a general idea of the magnitude and direction of volume change which corresponds to his control input, it is likely that another control input may indeed be necessary. This process continues until the desired state is achieved. In the event of a blocked valve, the process is essentially the same. The subject increases the valve setting and waits to see the corresponding change. Because there is a system fault, however, the control input will not have any effect on the system. Therefore, the subject will usually input another control action, presuming that his initial input was not great enough to have a significant effect. This continues for several iterations until the subject finally realizes that the system is not responding to his control actions as expected, and is therefore behaving abnormally. Consequently, with the P interface, it is difficult for subjects to discriminate between the onset of a fault and a normal situation requiring fine tuning.

Note that two subjects in the P group never even realized that a fault had occurred. One of these subjects was still able to successfully compensate for the fault. This is presumably because the subject was blindly reacting to the feedback given by the interface, without considering whether the observed system behaviour was consistent with what he would expect, given his control actions. Of equal importance is the fact that only two P subjects were able to correctly diagnose the root cause of the fault. They both did so by closing the outlet valve VO1 and by then watching the system to see what would happen. Since the volume in Reservoir 1 stopped decreasing and no changes occurred in Reservoir 2, the subjects were able to deduce that VA1 was the faulty component. The only subjects in this interface group who were able to compensate for this fault were the two who correctly diagnosed the problem, and one of the subjects who did not detect the fault.

4.1.2. Fault 2

In the second fault trial, Reservoir 2 developed a leak of 4 units/s 1 min after the start of the trial. The data for each group are presented in Table 2. A randomization test failed to reveal any significant difference between the interface groups in terms of detection times. Because the fault occurs early in the trial, no subject had stabilized the system when the fault occurred. This is an important point because it means that subjects were still filling up the reservoirs when the fault occurred.

TABLE 2

Detection, diagnosis, and compensation results for fault 2 (*leak of* 4 *units/s in Reservoir* 2 *at* 1 *min*)

		Detection after fault (s)	Diagnosis score	Correct diagnosis (s)	Difference (Dia–Det) (s)	Compensation (s)
	AM	169	2	—	—	1055
	CN	267	2	—	—	R2 HE
P + F	EC	189	2	—	—	976
	NJ	104	1	—	—	1140
	PL	53	3	54	1	498
	RL	219	3	313	94	646
	AT	—	0	—	—	608
	DC	228	0	—	—	586
P	HS	146	1	—	—	716
	MC	701	0	—	—	529
	SA	103	1	—	—	902
	WN	—	0	—	—	745

Because the leak was a relatively small one, the reservoir volumes kept on increasing after the fault occurred (although at a slower rate than they would without the leak). As a result, the effects of this fault were masked until the subjects completed their initial start-up actions and tried to stabilize the volume. At this point, most subjects realized that the system was not behaving normally because they could not stabilize the system as they normally would. As shown in Table 2, this masking effect caused the detection times for both interface groups to be generally longer than the detection times on the other two fault trials, for some subjects considerably so.

One would think that the P + F subjects would be at an advantage for this fault since the P + F interface directly displays the intended rate of change of volume in the salient form of a sloped line in the mass balance graphic (see Figure 3). Indeed, verbal protocol analysis of normal trials revealed that P + F subjects regularly stabilized reservoir volumes using this graphical indicator of the intended rate of change of volume. In this fault trial as well, the P + F interface displays a vertical line when subjects equate the input and output mass flows, indicating that the intended volume gradient is zero. When subjects see that the volume is in fact going down, the mismatch between the intended and observed volume gradients uniquely specifies a reservoir leak. In comparison, one would expect that the P subjects would find this trial more difficult because they do not have the reference provided by the sloped line in the mass balance graphic.

Interestingly, the results revealed that the P + F subjects did not react as expected. Although the sloped line allowed all six subjects to detect the fault, after detection four of the subjects persisted in trying to achieve a "vertical line" in the mass balance graphic. These subjects initially set the input mass flow greater than output flow, until the desired reservoir volume level was reached. The subjects then set input flowrate equal to output flowrate. However, because there was a leak, this resulted in a decreasing volume level. The subjects then increased the input flowrate until volume increased to a higher level. Instead of leaving the input flowrate at an increased setting to offset the leak, subjects would then reduce the input flowrate to its previous value so that it equaled the output flowrate, which of course resulted in a decreasing volume once again. This process continued until subjects realized that they could no longer use the sloped line to stabilize volume, as they normally did. At this point, subjects set the input flowrate to be consistently greater than output flowrate, ignoring the sloped mass balance line. Doing so finally resulted in a stable reservoir volume level.

One of the subjects in the P + F interface group never overcame his fixation, and ended the trial with a system failure. The other five subjects were able to effectively compensate for the fault after they overcame their fixations. It is interesting to note that three of the five P + F subjects who successfully compensated for the fault encountered the same type of problem when they tried to stabilize the temperature. Since the reservoir leak also caused a loss of energy, there was a mismatch between the intended rate of change of energy (displayed by the energy balance slope) and the actual rate of change of energy (displayed by changes in energy inventory level over time). Of the three subjects who encountered this fixation, two of them used the output temperature indicator as a referent for guiding heater control actions, ignoring the sloped energy balance gradient line. The remaining subject cleverly matched the slope of the energy balance line with that of the mass balance line to stabilize temperature. This shows an enhanced understanding of system dynamics. The rate of energy loss due to the leak is proportionally the same as the mass loss. Therefore, making the energy balance line parallel to the mass balance slope that is required to stabilize volume in the presence of a leak leads to a stable temperature.

Subjects with the P interface exhibited a very different pattern of behaviour. As with the previous fault trial, two subjects never even detected that a fault had occurred. Only one of the remaining subjects made an attempt at diagnosis. It is interesting to note that even though this subject's diagnosis was wrong, and the rest of the subjects never diagnosed the problem at all, they were all still able to compensate using a trial and error control strategy. That is, they would iteratively increase the valve settings to increase the flowrate until volume roughly stabilized, despite the leak. Finally, because the P interface does not display the same perceptual information as the P + F interface (in terms of the mass and energy balance graphics for the reservoirs), the P subjects never encountered a fixation problem. As a result, the P subjects exhibited faster compensation times than most, but not all, of the P + F subjects.

4.1.3. Fault 3

For the third fault trial, valve VA1 once again became blocked. This occurred at 6 min after the start of the trial, so all of the subjects had stabilized the system before the fault occurred. The data for this fault are presented in Table 3. As in fault 1, a randomization test revealed that the P + F subjects had significantly faster detection times than the P subjects ($p = 0.001$, one-tailed). Originally, this fault trial was intended to be the same as fault 1, so that cross-trial comparisons could be made. The idea was to determine whether there was a difference in performance between interfaces when subjects viewed a similar fault for the second time. As it turned out, the trials were unexpectedly quite different from each other. The important difference was that the demand for Reservoir 1 was 4 units/s in the first

TABLE 3
Detection, diagnosis, and compensation results for fault 3 (*VA*1 *blocks at 6 min*)

		Detection after fault (s)	Diagnosis score	Correct diagnosis (s)	Difference (Dia–Det) (s)	Compensation (s)
	AM	6	1	—	—	PA BLOW UP
	CN	6	3	58	52	840
P + F	EC	7	3	131	124	1079
	NJ	2	3	9	7	868
	PL	16	3	96	80	R2 OVERFLOW
	RL	4	3	39	35	1110
	AT	311	0	—	—	R1 HE
	DC	70	3	78	8	R1 HE
P	HS	32	0	—	—	PA BLOW UP
	MC	45	2	—	—	PA BLOW UP
	SA	20	3	81	61	934
	WN	17	3	135	118	PA BLOW UP

fault trial, and 7 units/s in this trial. Because the demand was higher, the output valve VO1 was set at a higher value (7), so when valve VA1 became blocked, the reservoir volume decreased at a faster rate. This increased rate of volume loss had profound effects on performance since operators' thinking was not able to keep up with the pace of the event.

After detection, four P + F subjects initially reacted to the fault by trying to increase the flow through valve VA1, as in the first fault. One of these subjects also closed valve VA2, presumably to force more flow through valve VA1. However, shutting valve VA2 means that pump PA was running with both flow lines downstream closed. Eventually, this caused the pump to break, which ended the trial. Two subjects immediately compensated for the fault by acting to ensure system safety (i.e. keep the reservoirs from going empty, or the temperature from getting excessively hot), thereby buying them time to reflect on the situation. This type of compensation effort is different from a compensation directed at maintaining the system in the goal state. In trying to ensure system safety, subjects completely stopped production by closing the outlet valve (VO1) and turning off the heater (HTR1). If the subjects had compensated to keep the system within the goal state, the outlet valve setting would most likely have been decreased and the heater setting reduced, but not turned off completely. Once the system was safe (e.g. the reservoirs were not going to empty or overflow), the subjects had time to attempt a diagnosis. Five P + F subjects correctly diagnosed the fault. Once this fault diagnosis was completed, two subjects again stopped output production (one subject even turned off the pumps), in order to allow time to develop a compensation plan. Once this plan was formulated, the system was brought back on-line, and the plan was executed. Four of the five subjects that formulated a correct compensation plan completed the trial. The other subject accidentally overflowed Reservoir 2 because he was concentrating on Reservoir 1, ignoring Reservoir 2.

As in the other faults, the P subjects exhibited a trial and error control strategy because the feedback they received made it difficult for them to discriminate the onset of a fault from a normal situation requiring fine tuning. Five out of the six subjects were only able to detect the fault after inputting control actions upon the system and realizing that the feedback they were receiving was inconsistent with what they were expecting. Only three out of the six subjects fully diagnosed the fault. Although they correctly diagnosed the fault, only one of these three was able to correctly compensate for it. Since the system was changing very quickly when this fault occurred, most subjects using the P interface could not complete the trial. Five out of six subjects blew up the system.

4.1.4. General summary

Fault management performance is affected by the type and magnitude of the fault, the system demands within that fault trial, and the system state when the fault occurs. As a result, it is difficult to make any broad generalizations. The descriptions below summarize the performance of each group, given this caveat.

Overall, the richer feedback provided by the P + F interface provides users with more opportunities to detect system faults. Also, subjects can perceive and understand the dynamics of the system more effectively, allowing them to diagnose the faults more accurately. To determine if there was a difference between the interface groups in terms of fault diagnosis, a two-factor, between-subjects ANOVA was performed. Interface (P or P + F) and Fault (1, 2 or 3) were the independent variables with diagnosis score (0, 1, 2, or 3) as the dependent variable. Results show that the P + F interface users were more accurate at fault diagnosis than the P interface users ($F(1, 10) = 10.80, p = 0.008$). Finally, although no significant differences in compensation times were found, it is generally true that subjects who diagnosed the problem had more success at compensating for the fault.

Although the functional information presented in the P + F interface proved to be useful to the operators for early detection and correct diagnosis, two problems were uncovered. First, it was difficult for subjects to determine initially that the volume gradient had changed during the reservoir leak. Because of this masking effect, fault detection was delayed. This problem was also observed with the P interface. Second, in the reservoir leak fault, subjects had to learn to pay less attention to the slope of the mass and energy balance lines and focus upon the volume level and temperature indicators, respectively, to effectively compensate for the fault. Interestingly, this is precisely the strategy that the P subjects had been forced to adopt all along, since they did not have the enhanced information to guide their actions. The fact that P + F subjects were fixated upon the perceptual properties of the interface, and seemed to interpret the line as representing the actual rather than intended volume gradient, led to generally slower compensation times for most P + F subjects for this fault. Whereas the first problem described above is a deficiency in each interface, the second problem seems to result from a learning process, since subjects did not have any experience using the P + F interface in fault trials.

As for the P interface, the lack of higher-order functional information made it difficult for subjects to detect a fault. Usually, detection occurred only because the subject was inputting control actions and the system was not responding as expected. In some cases, subjects using the P interface did not even detect that a fault had occurred. Also, the feedback given by the system during these faults is relatively indiscriminable from normal system feedback, so subjects had a comparatively more difficult time diagnosing the root cause of the faults. While fault compensation was possible, sometimes without detection or correct diagnosis, it was only because the subjects used a trial and error strategy for control. Subjects could not determine if the actions they were performing on the system were useful or detrimental to the system, due to the lack of comprehensive information regarding system state in the P interface. Therefore, subjects using the P interface were more likely to input control actions that were detrimental to the system, resulting in a higher number of system failures overall.

4.2. DUAL TASK PERFORMANCE

This subsection describes the results from the spatial and verbal loading tasks that subjects had to perform during normal (i.e. non-fault) trials.

4.2.1. Spatial loading task

An initial analysis of the controls for the dual task indicated that the results were consistent with the assumptions of the loading task paradigm, and therefore allowed us to meaningfully interpret the effects of the loading task on primary task performance [see Pawlak (1994) for a presentation of these results]. For the primary

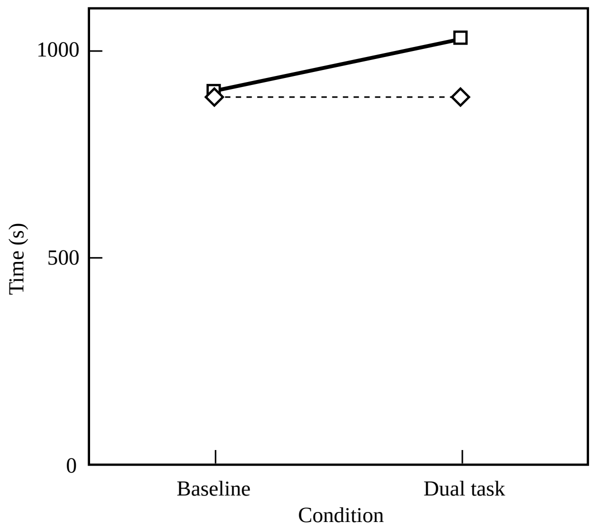

FIGURE 9. Effect of verbal loading task on total time to complete control task, by interface group. —□—: P; ---◇---: P + F.

task, five out of six P + F subjects ended the trial with a system failure, while only one out of six P subjects caused a system failure. A Fisher exact probability test indicates that the number of system failures in the P + F condition was significantly greater than in the P condition ($p = 0.04$, one-tailed). This finding suggests that the P + F interface loads more on spatial resources than the P interface under normal conditions.

4.2.2. *Verbal loading task*

There was no measure of performance on the verbal secondary task, since the task was only to continuously repeat a nonsense word. Thus, no control condition was necssary. The effect of the verbal loading task on the two interface groups is presented in Figure 9. A randomization test for matched pairs shows that there was a significant increase in time to reach steady state in the dual task condition for the P interface ($p = 0.047$, one-tailed) when compared to the baseline condition. In contrast, there was no significant difference in steady-state times between the baseline and dual task conditions for the P + F interface. Thus, unlike the spatial task results, these data suggest that, under normal conditions, the P interface loads more on verbal resources than the P + F interface.

4.3. START-UP TRIAL

Analysis of the start-up trials began with an investigation of aggregate measures of overall performance for the whole trial, such as total time and total number of control actions. These analyses failed to reveal any differences in mean performance between interface groups (Pawlak, 1994). As a result, more specific measures of performance were investigated, with the hope that these would be more sensitive. A CWA of DURESS II revealed that one of the system properties that can make start-up tasks difficult is the joint effect of throughput and volume on heater sensitivity (Vicente & Pawlak, 1994). A low mass throughput and low reservoir volume make the heater more sensitive, which in turn, makes it more difficult to control the temperature. Intuitively, if one has a low throughput, and only a drop of water in the reservoir, turning on the heater will increase the temperature very quickly, thereby overshooting the setpoint temperature. However, if the volume is larger, it will take longer to achieve the same steady-state temperature, thereby reducing the chances of overshooting the setpoint temperature.

All of the subjects adopted a strategy of stabilizing volume (i.e. making mass input flowrate and output flowrate equal) *before* trying to stabilize temperature. With this strategy, the mass throughput is fixed by the external demand setpoint. However, it is still up to the subjects to decide how much volume to have in the reservoir, in order to facilitate heater control. For example, if the demand is low, then a low volume results in the temperature being more difficult to control. If the operator uses a high volume instead, then the heater is not nearly as sensitive, providing more leeway in keeping temperature in the goal area. This "heater gain problem" can be observed, in part, by the relationship between volume, setpoint temperature, and energy inventory.

Figure 10 shows that this relationship is explicitly represented by the P + F interface. As a result, it is possible to explain this aspect of the heater gain problem by referring to the surface features of the P + F interface. Recall that the tolerance around the temperature is shown as a green area on the T1 and T2 thermometer

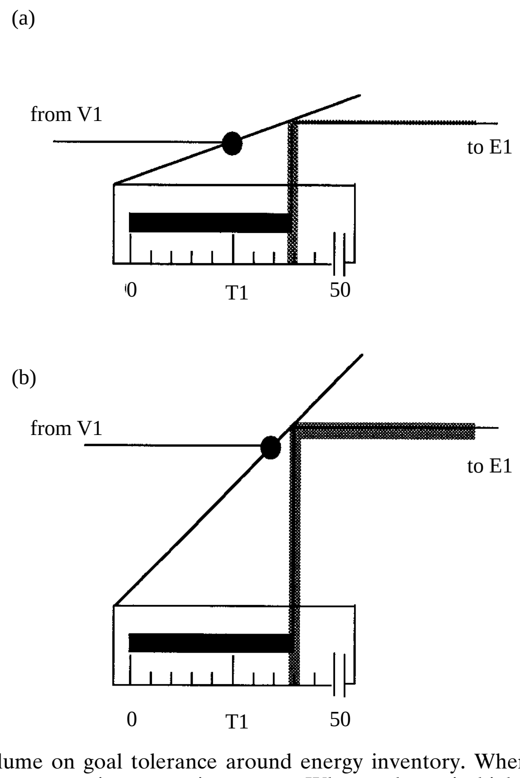

FIGURE 10. Impact of volume on goal tolerance around energy inventory. When volume is low (a), the goal tolerance around the energy inventory is narrow. When volume is high (b), the goal tolerance around the energy inventory is wider.

scales, which then reflects off of the diagonal to show the corresponding setpoint area on the energy inventory scales (E1 or E2). This tolerance on temperature is fixed at ±2°C, but the corresponding tolerance on the energy inventory is *not* fixed as it is mediated by volume. As shown in Figure 10, this mediating effect of volume is represented in the P + F interface by the diagonal line which relates energy and temperature. With a low volume (V1), the slope is small which means that the green goal area on the temperature scale (T1) when reflected onto the energy inventory scale (E1) is very thin (Figure 10(a)). This makes it difficult to stabilize the temperature at the setpoint. In contrast, when volume is larger, the slope is steep which means that the tolerance on the energy scale is now much wider (Figure 10(b)), thereby facilitating control. We hypothesized that the P + F subjects would learn that low volumes are problematic, and adjust their strategy accordingly. The P subjects, on the other hand, would probably not learn about this system property since it is not directly visible in that interface.

For the start-up trial, the demand for Reservoir 1 was 2 units/s and the demand for Reservoir 2 was 17 units/s. Thus, one would expect that those subjects with a low steady-state volume for Reservoir 1 would have difficulty reaching the temperature goals, requiring a greater number of heater actions. Using the data in the log files, the steady-state volume level for each reservoir was recorded. Since the trial ended when steady-state was reached, the values of the final volume levels for each reservoir at the end of the trial were used for analysis. To determine if the steady-state volume levels differed between the interface groups, a Fisher exact probability test was performed (see Table 4). An arbitrary cutoff point of one third total capacity was used as an operational definition of low volume. Results from the test showed that subjects' volumes (independent of reservoir) with the P interface were more often in the lower one-third of the total reservoir capacity ($p = 0.05$, one-tailed). This result indicates that subjects using the P interface were more likely to have low reservoir volumes. This increases the possibility that they would have trouble controlling the temperatures, resulting in an increased number of heater actions.

To determine if steady-state reservoir volume was indeed related to the number of heater actions, each log file was examined to first identify the point at which subjects stopped concentrating on stabilizing mass and started focusing on stabilizing temperature. This "switching point" between mass stabilization and temperature stabilization was determined by using Dplayer. Each subject's trial was examined and the switching point was defined as the last control input on any valve that ultimately resulted in the input flow being approximately equal to the output flow

TABLE 4
Steady-state volume level, by interface

		P interface	**P + F interface**
Volume Level	Above 1/3 of total capacity	7	11
	Below 1/3 of total capacity	5	1

(thereby stabilizing the mass in the reservoir) for each reservoir. A correlation analysis was then performed in order to determine if there was a relationship between the steady-state volume and the number of heater control actions after the switching point. Results from this analysis indicate a significant negative relationship between steady-state volume in Reservoir 1 and the number of heater control actions performed on heater 1 ($r(4) = -0.65, p = 0.02$). These results indicate that when throughput and demand are low, a lower reservoir volume does in fact result in an increased number of heater actions, after volume was stabilized. This relationship did not exist, however, for Reservoir 2 because of the higher demand throughput (D2 = 17).

Additional analyses revealed no significant differences between interfaces in number of heater actions after the switching point nor in total time to achieve steady state (Pawlak, 1994).

4.4. SHUTDOWN AND TUNING TRIALS

The shutdown and tuning trials for each subject were analysed in great detail using a variety of methods (Pawlak, 1994). However, no consistent differences between interface groups were observed.

5. Discussion

The goal of this experiment was to investigate, under representative conditions, the effects of a multilevel interface containing physical and functional levels of information on operator strategies and control in a dynamic process control system. The experimental findings indicate that there were no significant qualitative or quantitative performance differences between the interface groups for start-up, shut-down, and tuning trials. There was evidence of a strategy difference on start-up trials, with the P + F subjects showing a more refined adaptation, avoiding the problematic low volume areas. However, it seems that the P group, while entering this problem area more often, were able to achieve the same level of performance. Thus, after 4 weeks of daily practice, performance for the two groups did not differ under normal system conditions. This is not an uncommon finding. It is well known that, if given sufficient practice, operators can adapt to overcome the deficiencies of an interface (Rasmussen, 1968; Crossman, Cooke & Beishon, 1964/1974). It is only under abnormal situations that differences between operators tend to reveal themselves (Spencer, 1962/1974; Rasmussen, 1969).

Just because the P and P + F subjects achieved the same level of performance on normal trials does not necessarily mean that they were doing so in the same manner. The dual task results indicate that the P + F interface loads more on spatial resources than the P interface. Conversely, the P interface loads more on verbal resources than the P + F interface. Thus, there is no clear superiority. The important question is, Would one rather have more spare verbal or spatial resources available? We cannot definitively answer this question, but there are several reasons for suggesting that having more verbal resources would be most beneficial in an operational setting. First, there are various tasks that require verbal resources, such as communicating with other operators and reading procedures. The added load of these verbal tasks may cause problems with an interface similar to the P. Second, it seems likely that when faced with an abnormal situation, the type of activity operators would need to engage in (i.e. propositional reasoning) would load on verbal rather than spatial resources. The superior performance of the P + F interface during fault management tasks is consistent with this conjecture. It remains to be seen whether the differences in resource utilization between interfaces observed here change with experience.

For abnormal system conditions, the results indicate that the P + F interface is better for both fault detection and diagnosis. This finding has importance implications for the safe and efficient control of complex systems. With respect to detection, as mentioned before, subjects using the P interface generally detected a fault by an iterative cycle of slight adjustments to the system followed by observation of the results. However, for two of the three trials, two of these subjects did not even detect that a fault had occurred! In addition, the results indicate that this method of fault detection takes a considerable amount of time to complete, leading to long detection times. In complex industrial systems, this delay may lead to costly or irreparable damage being done to the system.

As for diagnosis, the P + F interface exhibited a clear superiority in all fault trials. Even in the reservoir leak fault, where the P + F subjects encountered a fixation problem, their diagnosis was still superior to that of the P group. Furthermore, in some cases, the lack of accurate diagnosis on the part of the P subjects caused them to take compensatory actions which wound up violating the integrity of system components. This finding also has important implications for the control of industrial systems. If operators are not aware of the actual state of the system, they may take actions that are inappropriate, thereby damaging equipment in the process. In summary, the P + F interface based on the EID framework has performance advantages that have important practical implications for both the safety and efficiency of plant operation.

Despite these advantages, the P + F interface is not without its limitations. The reservoir leak fault revealed that subjects find it difficult to compare the actual rate of change of volume, indicated by movement in the height of the volume indicator, with the intended rate of change, indicated by the sloped mass balance line. The slope of the line actually does specify in precise, quantitative terms how fast the volume level indicator should be changing, but this association is perceptually difficult to pick up and use to guide action (except in an ordinal fashion). Consequently, subjects find it difficult to detect that volume is increasing, but at a slower rate than it should, when a leak occurs. Thus, the reservoir leak is masked because subjects were increasing the reservoir volume at the time of the leak. The same problem was observed with the P interface. This deficiency seems to indicate a need for history information, which could make it easier to determine if the actual rate of change of volume was indeed the same as the intended rate of change of volume. Note that this is a problem with the P + F interface (and the P as well), not the EID framework. The latter requires that goal-relevant system constraints be made visible, but the P + F interface does not make some temporal constraints visible because it was difficult to integrate emergent feature displays with history information.

The reservoir leak also revealed that subjects encountered a perceptual fixation problem with the P + F interface. Subjects fixated on getting the volume gradient line vertical to stabilize reservoir level because this cue had been used so frequently and effectively in the past. However, in order to compensate for the leak, subjects had to pay less attention to the slope and use the volume level as information for simple proportional control of the input flowrate. It took several iterations before the P + F subjects figured this out, and in fact, one of them never did. The perceptual fixation resulted in longer total trial times for most, but not all, of the P + F subjects. Ironically, P subjects did not have a problem since they had been using a proportional feedback strategy all along.

There are two ways to interpret this fixation phenomenon. First, one could argue that this reveals a deficiency in the P + F interface. Second, it is possible that this result was observed because subjects were not experienced at using P + F interface under fault conditions. Under this interpretation, the fixation behaviour was a product of learning how to effectively use the P + F interface, not of poor design. There are several reasons to suggest that this second interpretation is the correct one. For example, the perceptual fixation problem was overcome by five subjects, and all of these correctly diagnosed the fault. Moreover, in a second study (Christoffersen, Hunter, & Vicente, 1994), the perceptual fixation problem was replicated but disappeared with subsequent experience with using the P + F interface under fault conditions.

6. Conclusions

The experiment presented in this article evaluated the EID framework under a more representative set of conditions than previous studies of EID. Subjects were allowed to control a real-time interactive simulation of DURESS II, and were given unusually extensive practice at doing so. The primary question of interest was whether, after extended practice, the principles of EID would induce more effective strategies and control than the P interface for normal, and especially, abnormal events.

Under normal conditions, there was no performance difference between an interface based on the principles of EID and a more traditional interface based on a physical representation of the system. However, the P + F subjects did reveal a more refined adaptation to system dynamics than the P subjects, as evidenced by the fact that the former adopted a reservoir volume control strategy that allowed them to avoid system states that are more difficult to control. In addition, the dual task results indicate that the P interface loads more on verbal resources, whereas the P + F interface loads more on spatial resources for normal trials. This provides the P + F interface with important advantages because it allows operators to have more free verbal resources that can be used for reading procedures, communicating with other operators, and propositional reasoning.

The most interesting findings to emerge from this study were that the P + F interface led to faster fault detection and more accurate fault diagnosis performance. These results are particularly noteworthy since they were accompanied by a qualitative difference in strategies. Whereas the P group generally tended to follow a trial and error fault compensation strategy, the P + F group exhibited much more sophisticated and more effective strategies that illustrated a better awareness and understanding of system behaviour and structure. Moreover, some of the strategies adopted by the P + F subjects are very similar to those exhibited by experienced operators in field studies of complex human–machine systems. For example, the observation that subjects sacrificed production and acted to ensure system safety in order to buy time for diagnosing the situation, and for planning a response is very similar to the findings of Amalberti and Deblon (1992), among others. Therefore, in addition to broadening the empirical support for EID, this experiment also shows that multilevel functional interfaces can actually induce very different strategies which, in some sense, raise subjects' level of expertise and, as a result, lead to superior performance. This is the single most important result to emerge from the experiment.

This study was also beneficial in that it uncovered a limitation of the P + F interface. The masking problem experienced by P + F subjects during the second fault trial was probably due, in part, to the fact that there was no history information available for reservoir volume. While one could add simple trend plots, the challenge lies in integrating historical information with the existing emergent features of the display.

6.1. LIMITATIONS

While this empirical evaluation of the EID framework has provided some encouraging results, it is evident that there are several limitations to this work. For example, we still do not know to what extent the benefits of the P + F interface are due to the added levels of information (content), or to the way in which that information is presented (form), except that both are important (Vicente, 1991). Also, the fault management portion of this evaluation only consisted of three fault trials. Thus, this experiment does not does address how well subjects who are experienced with fault conditions perform with an EID interface. Another limitation is that following the EID principles alone will not guarantee effective performance in large-scale systems (Vicente & Rasmussen, 1992). Presenting the set of goal-relevant constraints in a manner that is easily perceivable may succeed in inducing effective control in a small system like DURESS II, but in a more complex system, operators will need additional mechanisms for selecting among the vast amounts of information potentially available to them. EID must be complemented by principles of context sensitivity (Woods, 1991) and visual momentum (Woods, 1984) to address this problem of information extraction.

6.2. FUTURE RESEARCH

There are several outstanding issues pertinent to EID. First, it is important to study the effect that an EID interface has on long-term skill acquisition and on subjects' learning. In particular, it would be useful to know if the P + F interface used in this study enhances subjects' knowledge of the system over long periods of time. The results of such a longitudinal study would complement the findings obtained here. Second, it would also be useful to investigate the interaction between training an interface design. In this study, subjects were not provided with any training whatsoever, and therefore had to engage in discovery learning. It would be interesting to determine whether the interface differences observed here would change, quantitatively or qualitatively, if subjects were exposed to training programs of different types.

This research was sponsored in part by a contract with AECL Research (Jean-Yves Fiset, contract monitor), and research and equipment grants from the Natural Sciences and Engineering Research Council of Canada. The first author was also supported by University of Toronto Open Fellowship awards. We are deeply indebted to Dimitrios Nastos for programming the DURESS II simulation environment. Thank you also to Nuno da Silva and JoAnne Wang for their contributions to this research, and to David Woods and three anonymous reviewers for their comments on the manuscript.

References

AMALBERTI, R. & DEBLON, F. (1992). Cognitive modelling of fighter aircraft process control: a step towards an intelligent onboard assistance system. *International Journal of Man–Machine Studies,* **36,** 637–671.

BECKER, G. (1991). Analysis of human behaviour during NPP incidents: a case study. In *Balancing Automation and Human Action in Nuclear Power Plants,* pp. 517–526. Vienna: IAEA.

BENNETT, K. B. & FLACH, J. M. (1992). Graphical displays: implications for divided attention, focused attention, and problem solving. *Human Factors,* **34,** 513–533.

BISANTZ, A. M. & VICENTE, K. J. (1994). Making the abstraction hierarchy concrete. *International Journal of Human-Computer Studies,* **40,** 83–117.

BROOKS, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology,* **22,** 349–368.

BRUNSWIK, E. (1956). *The Representative Design of Psychological Experiments* (2nd edn.). Berkeley, CA: University of California Press.

CHRISTOFFERSEN, K. HUNTER, C. N. & VICENTE, K. J. (1994). *Factors influencing human cognitive behaviour I.* Technical Report CEL 94-05, Cognitive Engineering Laboratory, University of Toronto, Toronto, Canada.

CROSSMAN, E. R. F. W., COOKE, J. E. & BEISHON, R. J. (1964/1974) Visual attention and the sampling of displayed information in process control. In E. EDWARDS & F. P. LEES, Eds. *The Human Operator in Process Control,* pp. 25–50. London: Taylor and Francis.

GIBSON, E. J. (1991). *An Odyssey in Learning and Perception.* Cambridge, MA: MIT Press.

GOODSTEIN, L. P. (1981). Discriminative display support for process operators. In J. RASMUSSEN & W. B. ROUSE, Eds. *Human Detection and Diagnosis of System Failures,* pp. 433–449. New York, NY: Plenum.

MALONE, T. B., KIRKPATRICK, M., MALLORY, K., EIKE, D., JOHNSON, J. H. & WALKER, R. W. (1980). *Human factors evaluation of control room design and operator performance at Three Mile Island-2.* NUREG/CR-1270. Washington, DC: USNRC.

MUMAW, R. J., WOODS, D. D. & EASTMAN, M. C. (1992). *Interim report on techniques and principles for computer-based display of data.* Technical Report STC 92-1SJ4-CHICR-R1, Westinghouse Science and Technology Center, Pittsburgh, PA.

O'DONNELL, R. D. & EGGEMEIER, F. T. (1986). Workload assessment methodology. In K. R. BOFF, L. KAUFMAN & J. P. THOMAS, Eds. *Handbook of Perception and Human Performance,* Vol. 2, pp. 42-1–42-49. New York, NY: Wiley.

PAWLAK, W. S. (1994). *Inducing effective operator control strategies through ecological interface design.* Technical Report CEL 94-04, Cognitive Engineering Laboratory, University of Toronto, Toronto, Canada.

RASMUSSEN, J. (1968). *On the communication between operators and instrumentation in automatic process control plants.* Technical Report Risø-M-686, Research Establishment Risø, Danish Atomic Energy Commission, Roskilde, Denmark.

Rasmussen, J. (1969). *Man–machine communication in the light of accident records.* Technical Report S-1-69, Research Establishment Risø, Danish Atomic Energy Commission, Roskilde, Denmark.

Rasmussen, J. (1976). Outlines of a hybrid model of the process plant operator. In T. B. Sheridan & G. Johannsen, Eds. *Monitoring Behaviour and Supervisory Control,* pp. 371–383. New York, NY: Plenum.

Roth, E. M., Woods, D. D. & Pople, H. E. (1992). Cognitive simulation as a tool for cognitive task analysis. *Ergonomics,* **35,** 1163–1198.

Saariluoma, P. (1992). Visuospatial and articulatory interference in chess players' information intake. *Applied Cognitive Psychology* **6,** 77–89.

Sicard, Y. & Siebert, H. T. (1988). Control interfaces evaluation: incident tests on graphic displays. IAEA-CN-49/63P. In *Man–Machine Interface in the Nuclear Industry,* pp. 401–416. Vienna: International Atomic Energy Association.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioural Sciences.* New York, NY: McGraw-Hill.

Spencer, J. (1962/1974). An investigation of process control skill. In E. Edwards & F. P. Lees, Eds. *The Human Operator in Process Control,* pp. 67–78. London: Taylor and Francis.

Vicente, K. J. (1991). *Supporting knowledge-based behaviour through ecological interface design.* Ph.D. Thesis, University of Illinois at Urbana-Champaign, Department of Mechanical and Industrial Engineering, Urbana, IL, USA.

Vicente, K. J. & Burns, C. M. (1995). *A field study of operator cognitive monitoring at Pickering nuclear generating station—B.* Technical Report CEL 95-04, Cognitive Engineering Laboratory, University of Toronto, Toronto, Canada.

Vicente, K. J. & Pawlak, W. S. (1994). *Cognitive work analysis of the DURESS II system.* Technical Report CEL 94–03, Cognitive Engineering Laboratory, University of Toronto, Toronto, Canada.

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.

Vicente, K. J. & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics,* **SMC-22,** 589–606.

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

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd edn.). New York, NY: Harper Collins.

Woods, D. D. (1991). The cognitive engineering of problem representations. In J. Alty & G. Weir, Eds. *Human–Computer Interaction and Complex Systems,* pp. 169–188. London: Academic Press.

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

Paper accepted for publication by Associate Editor, D. Woods.

Appendix

This appendix provides a description of the coding scheme that was used to conduct the process tracing analysis of fault management behaviour. An example of the resulting fault data tables that were produced is provided in Table 5. One of these tables was produced for each subjects for each fault trial, for a total of 36 data tables (see Pawlak, 1994). Experimenter notes are in italics.

A short description of the fault, including when it occurred within the trial, was listed at the top of each data table. Also listed were the system output demands for

TABLE 5

Example of process tracing analysis data table

Fault: No. 2–1 min after start, Reservoir 2 develops a leak of 4 units/s. D1 = 4; D2 = 11 **Interface:** P + F **Subject:** PL

Initial Conditions: At time of fault, subject had low volumes in both reservoirs 1 and 2 (mass was approximately stable, but temperature was not). Both heaters were on.

Time after fault (min:s)	Detection activation	Diagnosis [O]bservation [I]dentification	Compensation [D]efine Task [F]ormulate a procedure [E]xecute [I]nterpret*
0:53	"OK · · · "		
0:54		I " · · · there's a leak in the reservoir."	
1:03			E $\uparrow$VO2
1:08			E $\uparrow$VO1
1:10			D "Going to have to investigate this problem · · · "
1:30			E $\downarrow$VA1
1:37		I "I think there's a leak in reservoir two."	
1:41			F " · · · I need to supply water to Reservoir 2 to compensate for the leak." *Subject was already doing this, he was just explaining what his rationale was.*
2:01			E $\uparrow$HTR2
3:11			E $\uparrow$HTR2
3:17			E $\downarrow$HTR2
3:38			E $\uparrow$HTR2
4:06			I* "Due to the leak in Reservoir 2, the energy balance is no longer reliable · · · " *Points to slope of* E_2
4:18			" · · · so I had to adjust the heater · · · "
4:25			" · · · by using the temperature level as a reference."
4:40		O "It seems that the setting of the heater is just right to maintain a steady temperature in reservoir two."	
4:40–7:17			E *Fine Tuning the system to maintain within goal areas.*
7:17			Steady-state reached

Before detection, the subject had attained a vertical line in Reservoir 2. He observed what was happening in the system and at (0:49) began to increase flow into R2 through valve VA2, and closed valve VO2. After detection, this subject immediately diagnosed the correct fault. Once this was accomplised, he formulated a compensation plan that enabled him to reach steady-state fairly quickly. This subject did not encounter any problems concerning perceptual fixation on display features. Note that at 4:06, the subject interpreted the consequences of what had happened in the system.

that trial, as well as each subject's initials and interface group. Below this, a short description of the initial conditions were recorded. System state, and any other relevant subject control activity around the time that the fault occurred were described. This was used to provide the context for the information contained in the data tables. Additional information that may have been relevant to the discussion of each subject's performance was included as needed.

In the interest of brevity, the data tables only contain actions, verbal protocol, and experimenter notes which were relevant to the detection, diagnosis, and compensation of the fault. In addition, this information only began from the point at which the subject detects the fault. The only data that could be used for analysis were the subject's actions and verbalizations. Therefore, if the subject did not verbalize all of his thoughts, then the data collected would be an incomplete and conservative picture of the subject's cognitive activity.

The first column of each data table contains the time at which each control action or verbalization occurred. This time was defined with respect to the onset of each fault so that comparisons can be made, not only across interface groups, but also across fault trials. Time is listed in minutes and seconds.

The second column, labelled detection, contains the subject's verbal protocol or experimenter notes that correspond to the subject's detection of the fault. Detection was defined using one of the following criteria.

- The subject says something which indicates that something abnormal is occurring within the system (e.g. "What was that?").
- The subject explicitly states that something specific about the system has changed in an abnormal manner (e.g. "There's a sudden drop of the water supply", "The water is strangely decreasing").
- The subject doesn't explicitly state that there is any sort of problem, but based on the control actions being made, it is easily identifiable that the subject is attempting to compensate for the abnormal system behaviour. This is identified by the subject acting upon the components of the system that are functionally affected by the occurrence of the fault. For detection to be scored in this manner, subjects must also verbally state at a later time in the trial that they had detected the fault and were attempting diagnosis and/or compensation.

The third column, labelled diagnosis, used two nodes of the decision ladder, observation and identification, for analysis. The data tables identify this information, which is always in the form of verbal protocol, with a letter "O" or "I", respectively.

The last column, labeled compensation, is broken down into three nodes of the decision ladder: define a task, formulate a procedure, and execute a control action. These are identified by the letters "D," "F," and "E," respectively. "Defining a task" and "formulating a procedure" are typically accompanied by the subject's verbal protocol in the data tables. However, sometimes experimenter notes are listed in place of the protocol, especially if the subject provided a great deal of redundant protocol. Control actions relevant to fault compensation are identified by the letter "E." Again, if many redundant control actions occurred, such as the repeated manipulation of the heaters to maintain the goal temperatures, experimenter notes are presented in place of each individual action.

It is important to discuss the process of mapping the verbal protocol onto the categories of the decision ladder. Each item of verbal protocol was examined and, based on the definitions given below, mapped onto the relevant information processing nodes of the ladder. The examples below illustrate typical statements made by subjects during the trials and how these statements were mapped onto the decision ladder.

- ACTIVATION. The subject says something that indicates some sort of acknowledgement that the system has significantly changed (e.g. "There must be some problem").
- OBSERVE. The subject describes either a current system parameter or indicates that a change in a general system parameter has taken place (e.g. "Reservoir 1 is going down").
- IDENTIFY. The subject makes a clear statement about the current system state. This usually is identified as a statement about some higher-level system functional relationship (e.g. "I don't seem to be getting any flow from this line here").
- INTERPRET. The subject describes the effects that a control input or a procedure has had, or could have, upon the current and/or future state of the system (e.g. "Due to the leak in Reservoir 2, the energy balance is no longer reliable").
- EVALUATE. This information processing activity was not observed in this experiment.
- DEFINE TASK. The subject states what needs to be done to the system to reach some goal (e.g. "First I have to prevent this from overflowing (R2) and prevent this (R1) from running dry").
- FORMULATE PROCEDURE. The subject identifies a set of actions that need to be performed in order to accomplish a task (e.g. "The water level is up, so I will open the outlet valve").
- EXECUTE. This is usually only a control input, but it is sometimes accompanied by a verbalization (e.g. "I'm going to increase this (VA1)").

The summary paragraph presented at the end of each data table is a qualitative interpretation of the entire trial. Subject control actions both before and during the fault are described, especially if these affect the interpretation of the data within the table. Any additional relevant information about the trial is also presented, including any additional verbal protocol the subject provided after the trial ended.