ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Int. J. Human-Computer Studies 59 (2003) 837–873

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

Using eye movement parameters for evaluating human–machine interface frameworks under normal control operation and fault detection situations

Y. Lin, W.J. Zhang*, L.G. Watson

Advanced Engineering Design Laboratory, Department of Mechanical Engineering, University of Saskatchewan, 57 Campus Dr., Saskatoon SK, Canada S7N 5A9

Received 1 January 2003; received in revised form 12 June 2003; accepted 18 June 2003

Abstract

A human–machine interface framework provides general guidelines for what information should be put on an interface display screen. The framework is thus a first step towards the design of an effective and efficient interface. This paper reports on an experimental study of two proposed frameworks: the ecological interface design framework and the function–behaviour–state framework. In order to provide an unbiased comparative evaluation for both interfaces, the same application problem is used. The interfaces, based on each of the two frameworks, are implemented with as similar look-and-feel forms as possible in the presentation of information contents. Only the normal control operation and fault detection situations are considered at this stage of the study. In addition, in this study three categories of measures are used, namely: the performance measure; the physiological measure (the eye movement measure: the eye fixation and the pupil diameter change, in particular); and the subjective (or the user-rated) measure. The major results obtained from the study includes the following: (1) the information called the abstract function in the ecological interface design framework may not positively correlate to the performance improvement yet may increase the mental workload, (2) the function–behaviour–state framework seems more agreeable with the operator's mental model, and (3) operators may perform equally well with a function–behaviour–state interface but with a reduced mental workload. It is also found that the eye fixation measure is highly consistent with the performance measure and the subjective measure. The pupil diameter measure is found not to be significantly sensitive to the mental

---

*Corresponding author. Tel.: +1-306-966-5478; fax: +1-306-966-5427.
E-mail addresses: yil063@mail.usask.ca (Y. Lin), wjz485@engr.usask.ca (W.J. Zhang).

doi:10.1016/S1071-5819(03)00122-8

workload information; however, it appears sensitive to the mental workload information among individual participants and shows a consistent result with the other measures used.

---

1. Introduction

With the rapid development of computer technology, more and more human–computer interface systems have been introduced into conventional human–machine systems. For example, a few computers could represent a huge control room with numerous gauges, buttons, and handles in a nuclear power plant, or an autopilot or pilot-aid system in the cockpit of an aircraft. Gradually, conventional hard-wired physical human–machine systems with numerous panels are replaced by human–computer interface systems (only a few keyboards and a few screens). This means that today human-machine interfaces can often be considered to be human–computer interfaces (interfaces for short). With plant systems being more and more complex, interfaces are becoming more complex as well. The design of such a complex interface is no longer a straightforward task. Furthermore, when a plant system becomes more complex, its dynamics are also more complex. The complexity of dynamics can lead to difficulties in monitoring/controlling such a plant system, which puts extra demands on the design of the interface.

Just as in the design of any complex system, a design theory and methodology is highly needed for interfaces. The first issue in designing an interface for a complex plant is what information should be displayed. In this paper, the term *framework* is used for any theory that provides knowledge about what information should be displayed. The other two issues related to interface design are how and when to display identified information contents. It is noted that in the literature, a framework was sometimes used to cover all three issues, i.e., what to display, how to display, and when to display.

There have been a few frameworks developed in recent years, such as the direct manipulation interface design (Hutchins et al., 1986), the cognitive layout (Norman et al., 1986), the ecological interface design framework (Vicente and Rasmussen, 1990), and the function–behaviour–state framework (Lin, 2000). The ecological interface design framework has shown some success (Vicente, 1999, 2000). However, a few controversial issues have arisen since this framework was proposed, and they will be discussed later in this paper. In addition, it was noticed that none of these published studies on the ecological interface design framework seems to address the issue of human operator mental workload. However, understanding of the operator mental workload is extremely important because an inappropriate level of human mental workload is likely to be a leading cause of errors in normal control operation and fault detection situations (Wickens, 1992).

One of the important motivations for the present study was to follow the idea, suggested by Janzen and Vicente (1998) and Moray and Rottenberg (1989), for using eye movement parameters to perform experiments on the ecological interface design framework. An eyegaze facility allows the tracking of the eye gazepoint positions on an interface display screen with a high resolution such that small interface widgets (or variables) can be differentiated. Another motivation of the present study was to meet the demand from the Atomic Energy of Canada Limited that set up with the University of Saskatchewan an industrial research chair program (1995–2000). Several meetings were held between the research team of the university and the company during 1998 and 1999, which came to the conclusion that a careful evaluation of the ecological interface design framework in the context of nuclear power plant control rooms was needed.

The initial study with the company soon raised questions regarding the ecological interface design framework. These questions were then formulated into hypotheses that needed to be tested. In the meantime, an alternative to the ecological interface design framework based on two thoughts was initiated. *First*, it was asserted that the operator's mental model should agree with the designer's mental model of a plant with which they both are associated, and an interface should facilitate the achievement of this agreement. *Second*, it was believed that the most generic design model for a plant system is the one which views the system in three fundamental aspects: the function, the behaviour, and the state (Umenda et al., 1990). This alternative framework for interface design will be briefly discussed in the next section. More details about this framework can be found in Lin et al. (2001). A refined version of this framework is to be reported separately. The results of the study are primarily documented in Lin (2000) and Lin et al. (2002).

The objective of this paper is to present our work, especially the experimental study, more comprehensively and completely. The salient points of the experimental study can be summarized as follows: (1) Both the performance and mental workload measures were considered. (2) Eye movement parameters were used for the particular problem of evaluating an interface framework, which at least provided an alternative way to those measures primarily based on video and verbal protocol (Vicente et al., 1995; Janzen and Vicente, 1998). (3) Two interfaces, which were based on two frameworks, respectively, were implemented by having a similar look-and-feel presentation. Consequently, the experimental results were more comparable.

The remainder of this paper is organized as follows. Section 2 provides background information (including two frameworks) and hypotheses to be tested. Section 3 gives full details of the methods of the experimental study, followed by results and discussions (Section 4). Section 5 discusses more implications of the study with special reference to related studies. Section 6 gives conclusions, including a discussion of the limitations of the present study, further work, and contributions. In the following discussion, the ecological interface design is abbreviated as EID, and the function–behaviour–state framework is abbreviated as FBS.

2. Background

2.1. Research vehicle

A thermal-hydraulic process plant system called the dual reservoir system simulation (DURESS) was taken as the example. DURESS was initially prototyped

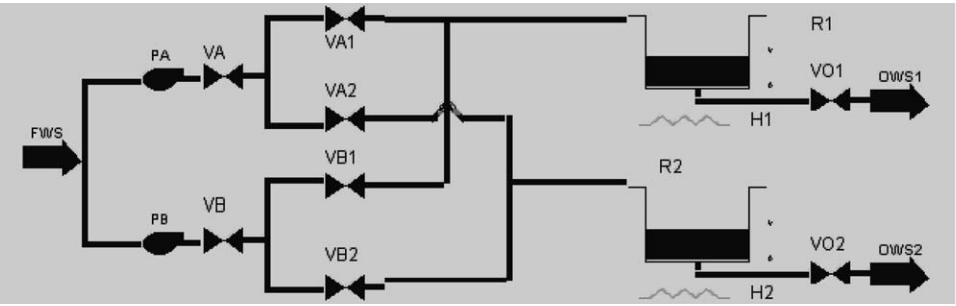

Fig. 1. DURESS plant system (the state network diagram).

by Vicente (1991) for illustrating and validating the ecological interface design framework. The structure of the DURESS plant system is shown by a state network diagram (see Fig. 1). The state network diagram is a representation of plant state relationships (note that a plant can be represented by a set of state variables, and thus, states). This system consists of the following components: VA and VB stand for input valve control, PA and PB for pump control, R for reservoir, VO for output valve control, H for heater control, T for temperature indicator, FWS for the feedwater stream, and OWS for output water stream. The meter beside each valve gives a reading of the flow rate.

The goal of the system is to maintain the desired temperatures (DT1 = 40°C, DT2 = 20°C) and desired flow rates (DV1 = 8 kg/s, and DV2 = 2 kg/s) of the water out of the two reservoirs at OWS1 and OWS2, respectively; see Fig. 1. To achieve these goals, one needs to control two pumps (PA, PB), eight valves (VA, VA1, VA2, VB, VB1, VB2, VO1, VO2), and two heaters (H1, H2).

2.2. *The ecological interface design framework*

Vicente and Rasmussen (1992) proposed the ecological interface design framework. This framework suggests that an interface should contain five levels of information: (1) the *functional purpose* (the purposes for which the system was designed); (2) the *abstract function* (the causal structure of the process in terms of mass, energy, information, or value flows); (3) the *generalized function* (the basic functions that the plant is designed to achieve); (4) the *physical function* (the characteristics of the components and connections between them); and (5) the *physical form* (the appearance and spatial location of those components). The DURESS system was prototyped (Vicente et al., 1995) as shown in Fig. 2. Five levels of abstraction hierarchy are captured and displayed on the screen as follows. The functional purpose is captured and displayed by the state variables, i.e., the desired flow rate DV1, DV2 and temperature DT1, DT2 setpoints. The generalized function is captured by the flow rate and the heat transfer rate, which are displayed beside the valves and the heaters, respectively. The abstract function is captured as the representation of the mass balance and energy conservation principles. The

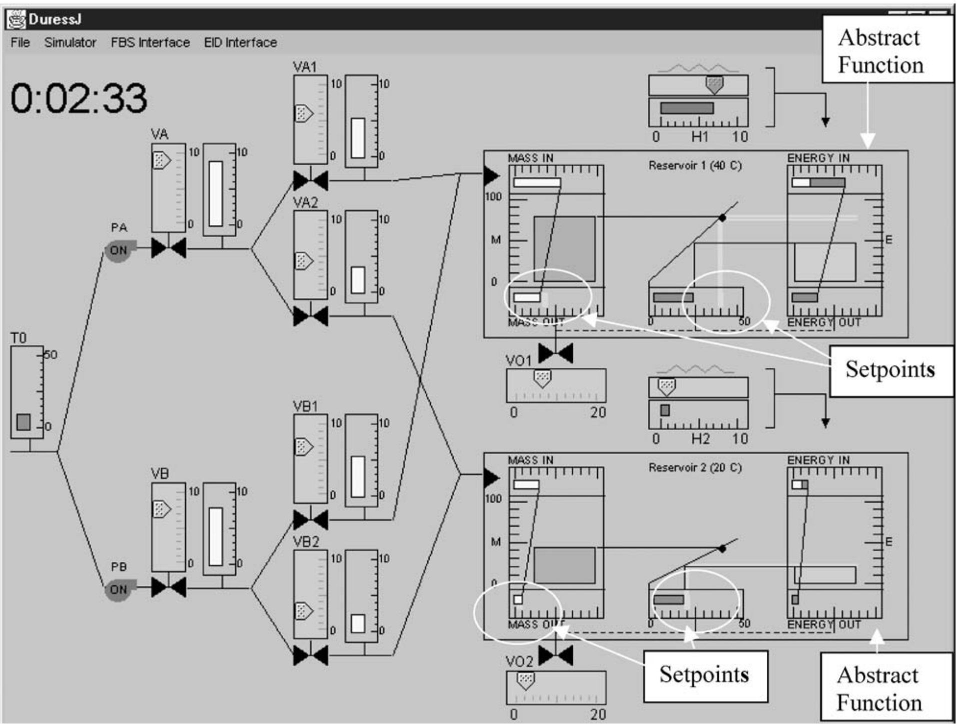

Fig. 2. EID–DURESS interface.

abstraction function for this particular plant is shown on the right in Fig. 2. According to Vicente et al. (1995), inputs are shown at the top (“mass in” and “energy in”) and outputs are shown at the bottom (“mass out” and “energy out”). The graph between the mass balance and the energy balance illustrates the relationship between volume, energy, and temperature. The physical function is captured as the state of the pumps, valves, and heaters. These states are displayed by the small triangular pointers on the respective scales. The physical form is represented by the topographic layout of the plant system (i.e., the location of the components and the connections between them). It is noted that the displays of the physical function and the physical form overlap.

2.3. The function–behaviour–state framework

2.3.1. Theory

The function–behaviour–state framework for interface design was proposed by Lin (2000). The most important assertion with this framework is that a generic design model of a plant should be compatible with a mental model of an operator, and an interface should display information to enforce such a compatibility between a designer and an operator. There are various proposals of models for the design of a general system (including a plant system), such as the axiomatic design theory (Suh, 1990), the systematic design methodology (Pahl and Beitz, 1995), the function–behaviour–state model (Umenda et al., 1990), and the integrated bond-graph and function-means methodology (Bracewell and Sharp, 1994). Based on an analysis of generic features of these design models and methods, a set of concepts, which founded the function–behaviour–state framework, was concluded and is described in the following. It is also noted that from the point of view of interface design, information about these concepts of a plant system should be displayed in an appropriate way.

System, *structure*, *system decomposition*: The *structure* of a plant is a set of entities (physical or conceptual) connected in a semantically sensible manner. It was assumed that a structure (a system) has a boundary which separates the structure from the rest of world which is called the *environment*, so the set of entities for a particular system is fixed. A system can usually be further decomposed into subsystems and components. Components are not further decomposed depending on the purpose of manipulating the system. Systems/subsystems have two patterns in terms of their topology: the tree pattern and the network pattern.

State and state variable: Entities are perceived by a set of properties, and these properties are called the *states*. The properties and the states are given a name. The name of the state is the state variable. The state variable can be further divided into the active state variable and the passive state variable. The *active* state variable corresponds to the actuation entity in a physical system, and to those that can be directly manipulated by the operator in the context of an interface. The *passive* state variable corresponds to those variables that depend on the active state variable. This dependency is governed by the 'principle' (see later discussions). It is clear that the passive state variable is not directly manipulated by the operator.

Behaviour: The *behaviour* is the causal relationship or structure among a set of related state variables/states. Such kinds of relationships have two parts: the intensional and the extensional. The intensional part is represented by algebraic or differential equation. The extensional part is the exhibition of the states of the related state variables. The extensional part of the behaviour is either *actual* in the sense that the passive state is based on the measurement via the sensor in the structure, or *expected* in the sense that the passive state is calculated based on the constraint equation. There are two situations in which the states are related: (a) if State 1 changes from 0.3 to 0.4, State 2 will change from 5 to 6; (b) at time t, State 1 is 0.3, and after Δt, State 1 will become 0.5.

Principle: The *principle* governs or accounts for the behaviour in a way that the causal relationship or structure is derived from the principle; for example, Newton's laws are used to derive the dynamic model of a robotic system.

Function: The *function* is defined as a purpose in the mind of humans and can be realized by the system (structure) through the provision of certain behaviours by the structure. The semantics of the function are given by an assertion which has the following syntactic form: Function: = verb | noun | [proposition] | [value 1] | [proposition] | [value 2], where the notation '[ ]' means optional, and the notation ': = ' means the assertion. For example, there are two functions of the DURESS plant system, i.e., function 1: generate the flow rate of water to 8 and 2 kg/s, respectively; function 2: generate the temperature of water to 40°C and 20°C, respectively.

It is further noted that the behaviour, the principle, and the function follow system decomposition. This means that it makes sense to speak of the behaviour (function, principle) of a component (subsystem, system).

2.3.2. Display

The concepts discussed above regarding the function–behaviour–state framework are displayed in the following way (see Fig. 3, which is a prototype interface of DURESS implemented based on the function–behaviour–state framework).

Display of the state: The active state variables are displayed in the lower part of Fig. 3. The component corresponding to an active state is displayed by a widget representing its active state or states (if a component has more than one state variable).

Display of the behaviour: The (actual) behaviour is displayed by simultaneously highlighting a set of related active states and passive states. Take the behaviour of valve VA1 as an example (see Fig. 3). The active state of VA1 is the valve opening, and the passive state of VA1 is the flow rate; sliding the widget of the active state

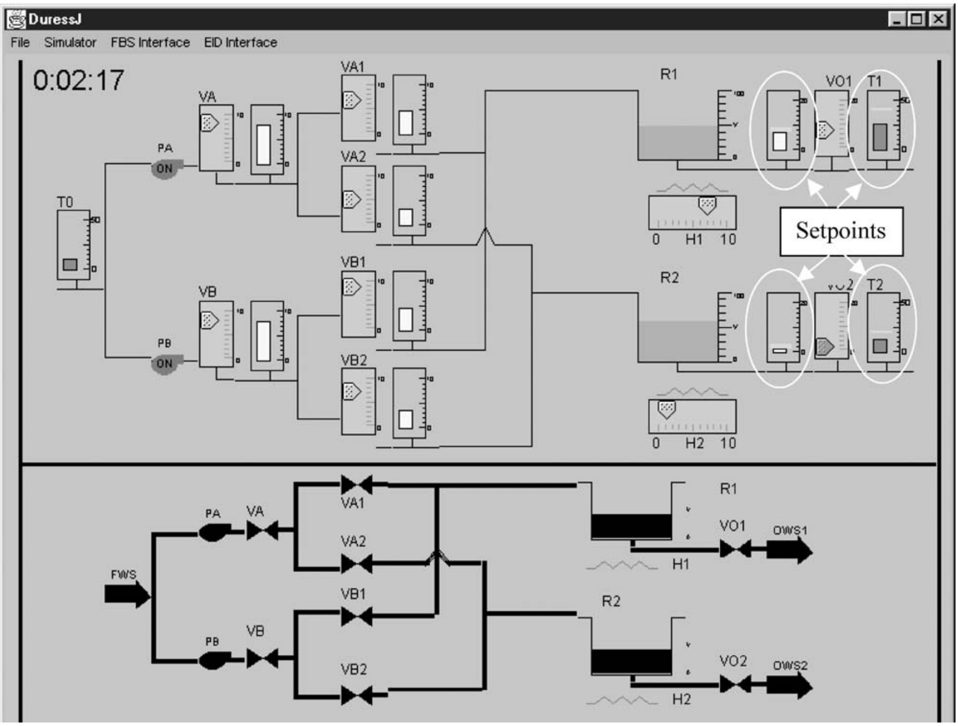

Fig. 3. FBS–DURESS interface.

(i.e., the valve opening) will immediately cause a display of the passive state (i.e., the flow rate). It is noted that the display of the behaviour of some structures (e.g., the pump) is omitted in order to create a similar reference or ground upon which comparison of the two interfaces is more unbiased.

Display of the system decomposition: Another important feature in the display strategy with the function–behaviour–state framework is the enforcement of compatibility between the operator and the designer on system decomposition. This enforcement is offered by the provision of a widget on the state network diagram (Fig. 3); triggering of this widget leads to the display of a system decomposition diagram for the plant system in question. Further triggering of the tag of a particular subsystem/component in the state network diagram causes the display of the states of the state variables of that particular subsystem/component. For example, if one triggers the tag of the subsystem (the tank R1 and the valve VO1) in the DURESS plant system, the following states should be displayed or highlighted: the level of the water in tank Rl, the flow rate at the outlet of VA1, the flow rate at the outlet of VB1, the flow rate at the outlet of valve VO1, and the temperature of the water at the outlet of valve VO1.

Display of the function: The function follows the behaviour. The function can be displayed in various ways. The most compact way is to have a widget on the passive state variable with which functions are associated. For example, for the entire DURESS plant, the four setpoints of the flow rate and the temperature of the water at the exit of R1 and R2, as illustrated in Fig. 3, correspond to the function of the plant.

Display of the principle: In the function–behaviour–state framework, the principle is not displayed explicitly because it is thought that the principle is something used for the analyst or modeler to derive the constraint equation for the state variable, which is not a part of the front-end plant model. The so-called ‘correct’ passive states are displayed as a means of exposing the constraint. It is believed or assumed that an interface should be capable of computing the ‘correct’ passive states. As such, the active state and the ‘correct’ passive state come to form the “correct” behaviour. Clearly, the ‘correct’ behaviour can be interpreted as the expected behaviour, and further, the expected behaviour is the goal to be achieved (thus, the function). The display of the expected behaviour should be in proximity to the actual behaviour (measured by the sensor). Furthermore, the expected behaviour is displayed under the control of the operator by triggering a control widget on an interface. A typical situation for applying the display of the expected behaviour may be as follows: in monitoring a plant operation, the operator may switch on/off the display widget for the expected behaviour. If the actual behaviour of a plant system is not the “same” as the expected behaviour, the root of a fault is thus implied (i.e., fault diagnosis).

2.4. Discussion

There must be some correspondence between the two interface frameworks, as they have the same purpose. The correspondence between them is shown in Table 1. The following remarks are noticed:

Table 1
The ecological interface design vs. the function–behaviour–state interface

The ecological interface design	The function–behaviour–state framework
Physical form	Structure
Physical function	State
Generalized function	Extensional actual behaviour (the component level)
Abstract function	Extensional expected behaviour
Functional purpose	Function (the system or subsystem level)

Remark 1. In the EID literature (e.g., Vicente et al., 1995), it appears that the generalized function is only associated with the component of a system. Janzen and Vicente (1998) did mention that the generalized function may also apply to the subsystem; unfortunately, this point is not exemplified. While in the function–behaviour–state framework, the behaviour can be applied to the component, the subsystem, and the system, as discussed previously.

Remark 2. The abstract function in the ecological interface was stated in several different ways: Vicente (1999) stated that the abstract function is applied to the storage components; Janzen and Vicente (1998) drew a table on which the abstract function is associated with both the system and the subsystem; Burns (2000a) considered that abstract function is applicable to the component, subsystem and the system. The principle in the function–behaviour–state framework may correspond to the abstract function, but the principle is applied to the system, subsystem, and the component.

Remark 3. The functional purpose in the EID framework appears only applicable to an entire system. While in the function–behaviour–state framework, the function follows the behaviour and is applicable to the system, subsystem, and component.

Some other differences between the two interfaces will also be discussed later in this paper; a more complete and comprehensive discussion refers to Lin et al. (2002).

2.5. Hypotheses

As mentioned in Section 1, based on a comprehensive literature study of the ecological interface design framework and many discussions with the industrial partner, the following hypotheses were proposed (note that in the following, words, such as ‘more readily’, ‘more difficult’, and so on are in the context of the comparison of one framework with another).

First of all, the two levels of information displayed on the EID display screen (Fig. 2), (the abstract function and the generalized function), and the procedure for using these levels of information with the other three levels of information do not comply well with the operator's mental model; it is difficult to educate or train operators to accept these levels of information as their mental models of a complex plant system. Here, the mental model means the representation of a plant in the operator's mental world. Operators more readily accept the connections between state, behaviour, and function (namely, the function–behaviour–state framework). Because of such an incompatibility between the operator and the designer in the question of what the plant is about, and because of the added visual complexity in displaying the abstract function with the EID, operators can experience a higher mental workload with the ecological interface design framework and its interface than with the function–behaviour–state framework and its interface. *Secondly*, the level of information called the abstract function in the EID may not easily be applied by operators in performing their tasks. As such, access to this level of information may not render an appreciable performance improvement or may not imply rational application of this level of information into the cognitive process associated with the tasks being performed. It should be noted that the said tasks are of the normal control operation and fault detection kinds.

3. Method

3.1. Experimental design

A statistically oriented, randomized complete block design (RCB) (Montgomery, 1997), with a factorial methodology, was employed for the present experimental design. In this experiment, there are two within-participants factors: interface and scenario. As the main objective of the study was to compare two interfaces, naturally, the interface factor was designed to have two levels: function–behaviour–state interface and EID. The scenario factor has four levels: L01, L02, F01 and F02. Therefore, the total treatments are 8 (2×4). The replicates for each trial are 3. Therefore, total trials for each participant are 24 (8×3). Among the four levels of scenario, L01 and L02 are normal scenarios (which means that a plant system works well, and the operation goal could be reached if the operator runs the plant correctly), and F01 and F02 are abnormal scenarios (which means that a plant is out of order, and a fault detection procedure is required from operators). Both the normal scenarios (L01 and L02) and the abnormal scenarios (F01 and F02) were designed with task difficulty in mind. In particular, in the normal scenario, task L02 is more difficult than L01, while in the abnormal scenario, task F01 (additional water into Reservoir 1) is more difficult than F02 (Heater 2 failure). The level of task difficulty of the normal control operation is basically determined by the initial setting of the active state variable and the properties of the fluid. These different settings can lead to different states of the passive state variables describing the flow rate and the fluid temperature at the outlet of the output valve VO1, VO2, as well as different dynamics of the states (the change rate of the states). The gaps between the states at the initial setting and the goal states and how fast the states change, become the measures of the task difficulty level, namely, a bigger gap and faster change rate imply a more difficult task. This is the rationale for making the task L02 more difficult than the task L01 in the normal control operation situation.

In the present study, the participant factor was out of the range of interest and was considered as a nuisance factor. Through a blocking technique, the potential effects of operators on the statistical comparisons among treatments could be systematically eliminated. Based on the preliminary experimental result, 20 blocks (operators) were considered necessary for this experiment. As a result, in total there were $24 \times 20 = 480$ runs for the entire experiment.

The other effects, e.g., the order of the operation of interface, etc., were also eliminated through the complete randomization of treatments within blocks. In addition, in order to eliminate possible "contamination" due to the order of interface, participants were further divided into two groups (half of the participants in each group). One group proceeded with the EID interface first and the other with the FBS interface first.

3.2. *Participants*

The criteria used for the selection of participants were constrained by considerations of experimental control and representativeness. With regard to control, the reliability and responsibility of participants were two main concerns. Applicants who showed an interest in participating in the experiment were interviewed and in some cases, references giving evaluations of applicants' reliability and responsibility were requested. With regard to representativeness, all participants had to pass the required credits in thermal fluids courses in a four-year engineering program and were associated with engineering disciplines.

The gender issue was considered as well. In many applications where manual control still prevails (e.g., vehicle driving), the gender factor is highly significant. In applications of computer user interfaces, the significance of the gender factor remains an open question. That said, in the future a nearly even number of female and male participants should be selected to address the gender issue.

As a result of the selection process, there were 20 students (6 female and 14 males, from the University of Manitoba) participating in the experiment. The average age of the participants was 31. Of the participants, 3 were undergraduate students and 17 graduate students. All had an engineering background. The participants were paid $12.00 per hour for the experiment.

3.3. *Apparatus*

3.3.1. *Eyegaze system*

The eye movement information was collected using Eyegaze from LC Technologies of Fairfax, Virginia, USA. This Eyegaze system is designed for measuring, recording, playing back, and analysing an operator's eye behaviour (the movement and change in pupil diameter) when the operator works on a computer interface. The configuration of the system is shown in Fig. 4. A video camera is located below the computer screen, and it continually observes the participant's eye.

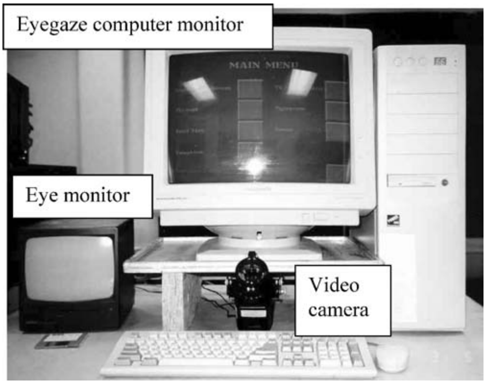

Fig. 4. Configuration of eyegaze system.

A small, low power, infrared light emitting diode (LED) located at the centre of the camera lens illustrates the eyes. This Eyegaze system uses the Pupil-Centre/Corneal Reflection method to determine the gaze direction of the eyes. It tracks the participants' eyes and generates data at the rate with which the video camera captures images, i.e., at the camera field rate of 60 Hz, which means that the gaze points and fixations are recorded every 17 ms. The accuracy required by the calibration program is 6.35 mm. In order for the Eyegaze system to work well, the experimental environment was illuminated with fluorescent lights, which emitted low levels of infrared light upon the participant's eye.

The Eyegaze system shakedown and testing included the following procedures: (1) the positioning of the participants (or the positioning of the monitor tray such that the participants were between 559 and 711 mm distance away from the screen, and their eye level was aligned with the top edge of the monitor screen); (2) the adjusting of the camera (it is recommended by the manufacturer that the focus ring of the lens be set to the maximum value of 1.3); and (3) calibration for each participant. This included two steps.

Step 1: The participant gazed at a sequence of dots, and the participant's gaze point on the screen was predicted.

Step 2: The participants were adjusted to reduce the discrepancy between the actual position of the dots on the screen and the eyegaze point on the screen until sufficient accuracy (less than 6.35 mm) was obtained.

The field of view of the Eyegaze system camera is approximately 38 mm wide, 30 mm high, and 38 mm back and forth (or as long as the participant is looking within about 40° of the camera *Z*-axis). The Eyegaze tracker is not head-mounted; therefore, a change in the position of the participant's head will affect a correct tracking of the participant's eyegaze (e.g., if the participant moves his/her eye outside the camera's field of view, the system does not detect a pupil and thus the associated corneal reflection). One of the solutions is to put a sensor on the participant's head so that the position of the head can be detected; the discrepancy between the position of the head at the calibration stage and the position at the operation stage can be used to obtain the correct eyegaze position on the screen. Another solution that was adopted in the present study was to have an accessory (a wooden plate in this case) "fix" the participant's head. When a participant operated the computer, he or she was required to sit back against the wall. Consequently, the position of the participant's eye with respect to the computer monitor screen where the application program is running could be fixed. It was noted that with such a helmet, participants could easily become fatigued (Lin, 2000; Burns et al., 2002). This was actually one of the reasons that the trial duration was short (also see the discussion in Section 3.4).

3.3.2. DURESS-Eyegaze integration system and data acquisition

Two programs, the application program (i.e., DURESS in this case) and the eyegaze system program, needed to run simultaneously. In the present study, one computer was used to run these two programs at the same time in the Microsoft Windows 3.2 system.

The DURESS program system was extended to the capability of recording. This apparatus system directly gave high level information, such as (i) the eyegaze fixation and the pupil diameter; (ii) the real-time participants' mouse movement information (i.e., the mouse icon movement and clicking on the screen); (iii) the real-time states regarding the process (valves, heaters and reservoirs); (iv) the time used by operators to control the system in order to reach desired system behaviours (in the normal control operation scenario) and to detect faults (in the fault detection situation). These real-time recordings were implemented by a log file or mechanism which is a part of the DURESS program. In addition, the non-real-time verbal protocols of the participants were manually written by the experimenter and were tape recorded as well.

3.4. Procedure

At the beginning of the experiment, all of the participants were asked to sign a consent form. Then they were given an introduction (30 min) outlining the purpose of the experiment, the DURESS, the function–behaviour–state framework, and the ecological interface design framework. After that, the participants had some limited time (5 min for each interface) to get familiar with both interfaces as well as the DURESS plant simulation system. The calibration procedure then followed. After calibration, the formal trials began. In the formal trial, the participants were not told the scenarios (i.e., four scenarios: L01, L02, F01, F02) with which they were working. If the participants felt that there was nothing wrong with the system, they just controlled the plant system (e.g., adjusting the openings of the valves and the heaters) until the demanded temperature and flow rate of the water out of the reservoirs were reached within a predefined duration (1 min in this study). Otherwise, the participants had to report any fault and then stop the trial. The participants could fail to fulfil the task either because they could not control the systems reaching the goal when there was no fault in the system, or because they could not detect a fault which was present in the system. After each trial was done, the participants were asked to proceed with a subjective evaluation. The subjective evaluation included the filling out of a subjective rating form called "rating scale mental effort (RSME)" and a narrative comment form (details of the subjective evaluation forms will be described in later discussions).

It is noted that the maximum trial duration was 8 min. If the participant could not fulfil the task during this maximum trial duration, the experimenter would stop the trial. The legitimacy of the maximum trial duration was roughly examined. Seven and 8 min as the maximum trial durations were tried. The experimental results were then collected to compare their differences. Statistically, the result showed that no significant difference existed between these two situations. The relatively short trial duration was also due to operator fatigue problems associated with the use of Eyegaze equipment (in both the eye and the posture that the participant had to maintain throughout the trial). Nevertheless, analysis of the literature revealed that duration length appeared not to be as sensitive as expected. For example, the duration length in the study performed by Ham and Yoon (2001) is about 8 min; yet they produced the same result (e.g., the performance variation in the presence of the generalized function and the abstract function with the EID is smaller) as Christoffersen et al. (1996) did with a duration of 30 minutes over half a year period.

3.5. *Measures*

One of the major features of the current study that is different from all the related studies in the literature regarding the EID is that a larger suite of measures was used. These measures can be classified into two categories in terms of their purposes: the performance measure and the mental workload measure. From the point of view of the measuring technique, these measures can be classified into three categories: the physiological measure, the performance measure and the subjective measure. In the following, each measure used in the present study is described.

3.5.1. *Physiological measures: the eye fixation and the pupil diameter*

There are a number of parameters associated with the eye behaviours, and they have been studied for many purposes other than the one in this paper; see the study presented by Goldberg and Kotval (1998). The present study was focused on eye fixation and pupil diameter parameters, as they were generally found sensitive to the measurement of human mental workload. Furthermore, eye fixation and duration is an excellent tool to track the operator's cognitive process.

Fixations are pauses over informative regions of interest (Salvucci and Goldberg, 2000). In this study, one *eye fixation* is defined as a duration of the eye gazepoint within a circle (with radius being 6.35 mm) longer than 100 ms (LC Technologies, 2000). Measures of fixation include two attributes: the fixation duration (or dwell time) and the number of fixations. Longer *fixation duration* implies more time spent on interpreting, processing or associating a target with its internalized representation. The fixation duration is also called *dwell time* (Backs and Walrath, 1992). A larger *number of fixations* implies a larger magnitude of required information processing. It is clear that the eye fixation measure can generally be used to correlate the efficiency of task execution. Given a definite amount of attention resources, eye fixation can reflect performance; while under a certain restriction on performance (say, an operator has to complete a task within 5 min), eye fixation reflects the operator's mental workload.

In the present study, the eye gazepoints and fixations were traced, and the tracking results were dumped into a file which was separated into two parts. Part I was comprised of the gazepoint data, including the sampling time, *X* coordinate, *Y* coordinate, and validate/invalidate (if the participant looked off the screen, the data is invalidate). Part II included the fixation time data, including the position of the eye fixation, the fixation duration, and the saccadic duration.

In order to analyse the eye fixation data, the function–behaviour–state interface and EID were intentionally divided into several regions (see Figs. 5 and 6, respectively). These regions were defined to be targeting or bounding variables in question or in interest. For example, regions Nos. 11 and 14 (see Fig. 6) target the abstract function of the EID.

Pupil diameter usually increases with an increase in mental workload (May et al., 1990). A change in the pupil diameter reflects difficulty levels of tasks; greater complexity either in the task itself or the interface form leads to greater pupil dilation (Stern, 1997). Also, there is suggestive evidence that the pupil constricts as a physiological sign of 'fatigue', or a function of time on task (TOT) (Stern, 1997).

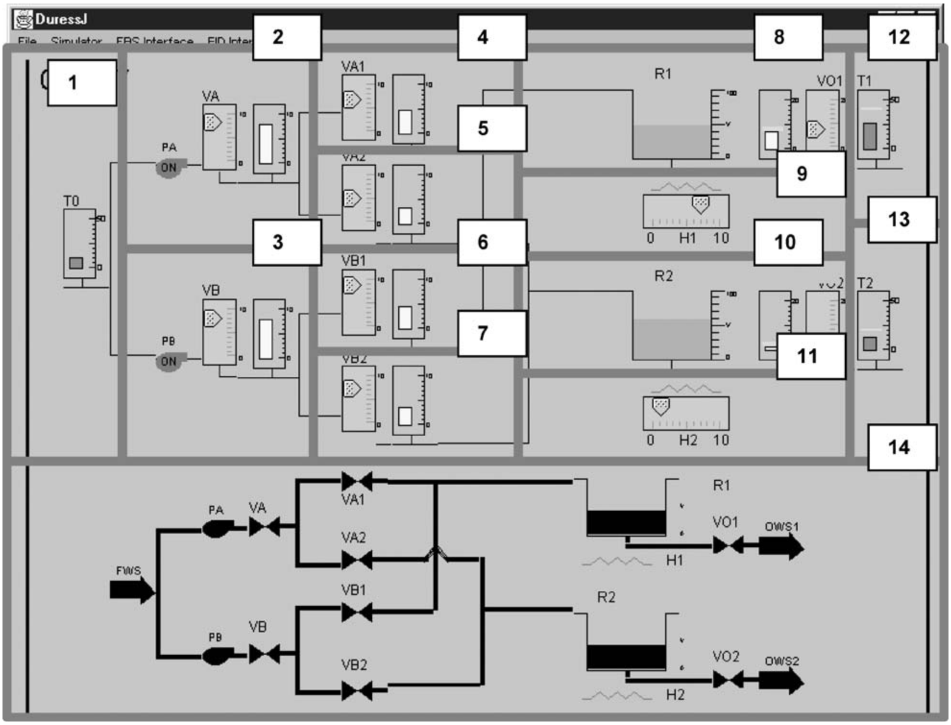

Fig. 5. Region decomposition of FBS–DURESS.

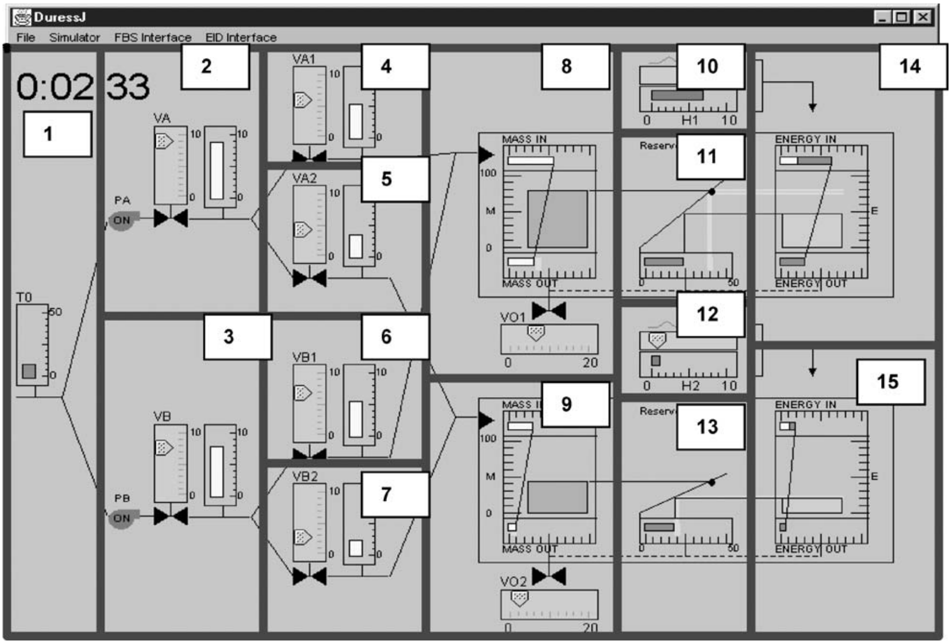

Fig. 6. Region decomposition of EID–DURESS.

Interestingly, it was reported that the pupils were significantly more dilated (by 8%) in the more difficult trials than in the easier trial initiation conditions, which may suggest an increase in a participant's level of effort or arousal and the achievement of improved subsequent task performance (Washburn and Putney, 1995). Jorna (1997) reported that pupil size increased as a function of the controller data link interface and traffic density in a pilot simulation study. It was reported that human cognitive processes, such as problem solving and language comprehension, are accompanied with pupillary dilation (Beatty, 1982). Known as a cognitively highly demanding task, simultaneous interpretation was studied using the pupillary response as an independent on-line measure of cognitive load, and clear differences did appear among experimental tasks; for example, words more difficult to translate induced higher levels of pupil dilation than did easily translatable words (Hyona et al., 1995). However, the relative sensitivity of the change in the pupil diameter as a measure of mental workload is highly debatable. For example, one of the major controversies in pupillary research was found in Hess (1964) regarding pupil response to arousing visual stimuli. Also, of special interest is that pupil dilation reaches a maximum as maximum effective storage is reached, and pupils may even decrease in diameter when memory is overloaded (Pealver, 1974). In other words, pupils may be sensitive to both the extent of processing capacity as well as the breakdown of capacity (Granholm et al., 1996). This kind of finding is not very surprising, as any experiment makes its result legitimate only under certain conditions. That said, pupil diameter is a measure which is susceptible to experimental conditions and noise.

3.5.2. *Performance measures*

Performance measures are those variables whose values give indicators of how operators behave in achieving their operation goal. In particular, the following variables were defined and used in the present study.

Normal control operation time: This measure refers to the time for an operator to complete a goal which is predefined and known to the operator. This measure is only applied to the normal control operation situation.

Fault detection time: This measure refers to the time required for an operator to detect a fault. It is only applied to the fault detection situation.

Fault detection rate: This measure refers to the successful rate of a fault being correctly identified within a predefined period of time (up to 8 min in the present study). This measure is only applied to the fault detection situation.

Operation failure rate: This measure refers to any error in operation which results in an early breakdown of the whole process system. The symptoms of these errors are related to (a) pump blow up, (b) reservoir emptying due to over-heating, (c) reservoir overflow, and (d) reservoir reaching the boiling point of fluids due to over-heating.

3.5.3. *Subjective measures*

The subjective measure is an indicator related to the participants' internal experience and is widely considered as an effective indicator about human mental workload in human factors engineering. Typically, the subjective measure is assessed with a rating scale which gives quantitative information about mental workload. Several subjective workload scales were developed and used in the recent years, such as the modified Cooper-Harper (MCH), NASA task load index (TLX), overall workload (OW), subjective workload assessment technique (SWAT) and rating scale mental effort (RSME). Among them, RSME is a unidimensional scale which was developed by the Delft University of Technology (Zijlstra, 1985). In RSME, ratings of invested effort were indicated by a cross on a continuous line. The line ran from 0 to 150 mm, and every 10 mm was indicated. Along this line, at several anchor points, statements related to invested effort were given, e.g., 'almost no effort' or 'extreme effort'. The scale was scored by the measurement of the distance from the origin to the mark in mm. A higher scale implies a higher mental workload. It was shown that the RSME is a relatively good indicator for self-report workload assessment (de Waard, 1996). Therefore, the RSME was adopted in the present study. In addition, the participants' narrative comments were also elicited as an additional subjective measure.

3.6. *A general note*

In general, the experimental design in the present study was fairly close to those experimental designs in the ecological interface design literature (e.g., Vicente et al., 1995; Janzen and Vicente, 1998; Burns, 2000a). Other aspects of the experimental settings, such as the number of participants, the participants themselves, the pre-training of the participants, the task difficulty level, and the trial duration, are also fairly compatible to those published studies on ecological interface design. For example, the number of participants in Burns (2000a) was 20 (6 female and 14 male), the trial duration in Burns (2000a) was 15 and 8 min in Ham and Yoon (2001).

4. Results and discussions

Data analysis aims to determine whether any factor significantly affects response (the operator mental workload and the performance in this case). For instance, data analysis on whether the interface factor imposes a significant effect is frequently investigated, as the main purpose of the present experimental study is to examine the differences in interfaces developed by two frameworks. The experimental design conforms to the standard conditions of an analysis of variance ANOVA (e.g., within-participants design). Therefore, ANOVA was used to examine whether the effects of the factors led to significant differences in the results. If a significant difference was found, Duncan's multiple range test (Montgomery, 1997) could be further employed to give more detailed causes for the difference identified. In the following, results are presented and discussed in terms of the three categories of measures, followed by a discussion.

4.1. Subjective measures

4.1.1. RSME

The RSME was calculated for the designed experiment, and the results are shown in Table 2. First of all, from Table 2 it can be seen that the increase of the task difficulty level (e.g., L02 is more difficult than L01) led the RSME to increase. Further, from Table 2, it can be seen that the mean value of the RSME (RSME = 57, S.D. = 23) for the FBS interface was significantly ($p<0.05$) less than that for the EID interface (RSME = 62, S.D. = 24) (Duncan's multiple range test, $\alpha = 0.05$), which implies that the participants experienced a higher mental workload with the EID

Table 2
Rating scale mental effort (RSME) means

Scenario/task	Interface Pr > F*	RSME mean	RSME mean	
			FBS	EID
Total	<0.0001		57.188^{B}	62.229^{A}
L02	0.0164	65.458^{A**}	62.083^{B}	68.833^{A}
L01	0.2032	60.317^{B}	58.717^{A}	61.917^{A}
F01	<0.0001	58.850^{B}	53.367^{B}	64.333^{A}
F02	0.7464	54.208^{C}	54.583^{A}	53.833^{A}

* Significance level $p<0.05$.

** Means with different letters (comparison within each column) are significantly different (Duncan's multiple range test, $\alpha = 0.05$).

interface than with the FBS interface. A closer examination of the results further shows that in the case of relatively easy tasks, i.e., L01 for the normal control operation and F02 for the fault detection situation, the differences between the two interfaces in mental workload were not significant; while in the case of relatively difficult tasks, e.g., L02 for the normal control operation and F01 for the fault detection situation, the differences between the two interfaces in mental workload were significant (in particular, the mental workload with the EID interface is higher than that with the FBS interface). These results positively test the hypotheses that operators can experience a higher mental workload with the EID interface than with the FBS interface.

4.1.2. Narrative comment

A summary of narrative comments showed that 55% of participants preferred the FBS interface, 10% of participants preferred the EID interface, and the remaining 35% of participants had no special preference to either of the two interfaces. Further collection of participants' comments showed the following: *For the FBS interface*, the positive impressions included the following (more than 90% of the participants): (i) participants felt it was easier to get accustomed to this interface and to pick up on what was occurring, (ii) participants felt to make less errors with this interface, (iii) participants felt less effort was required with this interface, and (iv) participants felt more comfortable with this interface. The negative impressions of the FBS interface included the following (less than 15% of the participants): (i) this interface did not provide enough information, and (ii) it did not align related information closely enough.

For the EID interface, the positive impressions included the following (less than 15% of the participants): (i) its abstract function seemed to be helpful, and (ii) participants felt able to detect faults more quickly with this interface. The negative impressions of the EID interface included the following (more than 90% of the participants): (i) this interface had a complicated layout and was difficult to comprehend and memorize, (ii) it made concentration difficult due to the large amount of information, especially the abstract function, (iii) participants felt more overwhelmed by cluttered information and had difficulty picking up the correct information per se, (iv) participants were very likely to overlook some important information, and (v) participants felt more nervous and used more eye resources and attention.

The results of participants' narrative comments can be further analysed. The function–behaviour–state framework matches better than the ecological interface design framework to the participants' mental model of the plant and the design model of the plant (considering only the normal control operation and the fault detection situations). It may be argued that the ecological interface design framework is more useful in problem-solving (specifically, the fault diagnosis and the fault compensation). However, the root cause of this advantage is not necessarily the presence of the abstract function and the way it is defined (only applicable to the storage component in a system (e.g., Vicente, 1999)). In the function–behaviour–state framework, the extensional expected behaviour (which is readily applicable to any level of a system decomposition) has more potential for operators to apply a fundamental problem-solving strategy/technique called analytical redundancy (Frank, 1990).

Comments about the difficulty in memorizing have an important implication on the effectiveness of an interface. Vicente et al. (1995) used memory recall techniques to evaluate the interface. The proposition for this technique is: the more variables recalled, the better the interface. According to this proposition, the prediction that the FBS interface should be more effective than the EID interface can be made, as participants had more difficulty recalling the semantics of the levels of information presented on the EID interface. One might argue that this is a matter of training. This argument needs to be cautiously applied, however, as it actually disputes a general fact; that is, there is an issue concerning how easy or how difficult it is to learn a given theory or method.

It is interesting to note that the general impression obtained from the participants, i.e., visual complexity, busyness and clutter information, is in strong agreement with the result obtained by Itoh et al. (1995).

4.2. *Performance measures*

4.2.1. *Normal control operation time*

It can be seen from Table 3 that for the relatively easier task L01, the normal control operation time with the two interfaces is not significantly different ($p<0.05$). However, for the more difficult scenario L02, the normal control operation time with the two interfaces is significantly different, and in particular the normal control operation time with the FBS interface is much shorter than that with the EID interface (FBS: mean = 215 s, s.d. = 89 s; EID: mean = 253 s, s.d. = 114 s) ($p<0.05$). Specifically, the operator with the FBS interface would take 15% less time than with the EID interface. Note that the term normal control operation time corresponds to the term normal trial completion time in Janzen and Vicente (1998). It is recalled from the previous discussion that the FBS interface (not the theory behind it) can be viewed as an equivalent interface to the EID interface with the absence of the abstract function; therefore, the result shown here implies that the abstract function information does not help in normal control operation in terms of the normal

Table 3
Comparison for normal control operation time (NCOT)

Scenario	Interface Pr > F^*	NCOT Mean	NCOT mean	
			FBS	EID
L02	0.050	233.858^{A**}	214.82^{B}	252.90^{A}
L01	0.3220	171.158^{B}	178.28^{A}	164.03^{A}

*Significance level $p<0.05$.

**Means with different letters (comparison within each column) are significantly different (Duncan's multiple range test, $\alpha = 0.05$).

Table 4
Comparison for fault detection time (FDT)

Scenario	Interface Pr > F^*	FDT mean	FDT mean	
			FBS	EID
F01	0.8309	111.083^{A**}	112.05^{A}	110.117^{A}
F02	0.096	112.883^{A}	122.35^{A}	103.42^{A}

* Significance level $p < 0.05$.

** Means with different letters (comparison within each column) are significantly different (Duncan's multiple range test, $\alpha = 0.05$).

control operation time. The result appears to be contradictory to that obtained by Janzen and Vicente (1998) (where they showed a positive indicator of an interface with the presence of the abstract function for normal trials).

4.2.2. Fault detection time

It can be seen from Table 4 that the fault detection time with the two interfaces is not significantly different for the F01 and F02 task situations. The result in the present study is in agreement with the result obtained by Janzen and Vicente (1998), who said: "…No variables were entered into the model, indicating an absence of a significant linear relationship between attention allocation strategies and fault detection time" (p. 537). The result of the present study would likely also be in agreement with the result obtained by Burns (2000a) (who found that the EID interface with spatial integrated and temporal not integrated display of information led to the fastest fault detection time, assuming that the access of the operator to the abstract function is significantly less).

4.2.3. Fault detection successful rate

It is shown from Table 5 that (1) by averaging both F01 and F02, the fault detection with the EID interface is 72% (correct counts = 86 out of 120) and 86% (correct count = 103 out of 120) with the FBS interface. It is further noted that for the relatively more difficult task F01, the fault detection with the EID interface is very low, 48% (29 out of 60), while with the FBS interface it is 77% (46 out of 60). It is noted that the term *fault detection successful rate* may also be called *fault detection accuracy* (e.g. Burns (2000a)). The result of the present study is contradictory to the result obtained by Burns (2000a) (who found that the EID interface with spatial and temporal integrated display of information led to the highest fault detection accuracy, meaning that the presence of the abstract function is positively helpful to a better performance in terms of fault detection accuracy). One could argue that Burns' study was not about what information should be put on the interface display screen, but instead about how (spatial integration) and when (temporal integration) to display information. Therefore, such a comparison is not legitimate. However, in certain circumstances, when to display can imply what to display. For example, in Burns (2000a), given an extreme case, if an operator never triggers his or her control

Table 5
Fault detection successful rate

No. of correct detection	F01	F02	Total
FBS	46 (out of 60)	57 (out of 60)	103 (out of 120)
EID	29 (out of 60)	57 (out of 60)	86 (out of 120)
Total	75 (out of 120)	114 (out of 120)	189 (out of 240)

Table 6
Operation failure rate

Interface	No. of error operation
FBS	30 (out of 480)
EID	57 (out of 480)

action to the abstract function information content in the case of the temporally controlled display of information, then the abstract function information will never actually be exposed to that operator, meaning that the piece of information called abstract function is absent. In the result reported by Burns (2000a), there was considerable reduction in the display of the "abstract function/component" (in proportion to the other information) in the case of the temporally controlled display of the EID interface, compared to the display of the "abstract function/component" (in proportion to the other information) in the temporally controlled display of information. So, it appears sensible to relate the EID interface with the temporally controlled display of information to the FBS interface without the display of the principle level information.

4.2.4. *Operation failure rate*

Table 6 shows the result of the operation failure measurement. The operation failure rate is 6% with FBS and 12% with EID. This means that the participants had twice as many failures with EID as with FBS.

4.3. *Physiological measures*

4.3.1. *Pupil diameter (PD)*

The results of the change in pupil diameter are documented in Table 7. It can be seen from Table 7 that (i) the pupil diameter measurements of the two interfaces are not significantly different ($p<0.05$), and (ii) Duncan's grouping result further shows that with an increase in the task difficulty level (from L01 to L02, and from F02 to F01), the pupil diameter mean significantly decreased from L01 to L02, whereas it significantly increased from F02 to F01. This result implies that the pupil diameter is not a sensitive measure in this application. One possible cause may be that the difference in terms of task difficulty level is too small. It is noted that in the present

Table 7
Pupil diameter (PD)

Scenario	Interface Pr > F^*	PD mean	PD mean	
			FBS	EID
Total	0.4988		3.189^{A}	3.193^{A}
L02	0.9189	3.182^{B**}	3.181^{A}	3.182^{A}
L01	0.1062	3.188^{A-B}	3.177^{A}	3.200^{A}
F01	0.8348	3.204^{A}	3.206^{A}	3.203^{A}
F02	0.7976	3.189^{A-B}	3.192^{A}	3.188^{A}

* Significance level $p < 0.05$.

** Means with different letters (comparison within each column) are significantly different (Duncan's multiple range test, $\alpha = 0.05$).

experiment the fault detection tasks may not necessarily be more difficult and more demanding than the normal control operation tasks. When the symptom of a fault is more obvious, the detection of the fault may be less challenging.

The pupil diameter measurements on individual participants were investigated in the present case. It was found that the pupil diameter with each participant (12 participants in total) was not significantly different with the two interfaces (respectively), the pupil diameter with each participant (6 participants in total) was significantly smaller with FBS than with EID, and the pupil diameter with each participant (2 participants in total) was significantly larger with FBS than with EID. This result may imply that an individual operator's level of expertise and visual perception capability are significant factors and need to be further examined.

4.3.2. Eye fixation

The results with the eye fixation measure are shown in Figs. 7 (FBS) and 8 (EID), respectively. When the eye fixation on individual components and subsystems is examined, significant differences in the two interfaces are found (see Table 8).

Regarding the system target zone and other zones, it can be found that most portions of participants' eye fixations were placed on the target zones of both interfaces (Figs. 9 and 10). The average percentage of the eye fixation on these zones was up to 61.9%. The average percentage of eye fixation on the abstract function zone on the EID interface (i.e., Zones 11 and 14 in Fig. 6) is only 3.4%, and the state network zone with the FBS interface (i.e., Zone 14 in Fig. 5) is 0.6%. This may imply that during the entire period of trial (157 ± 75 s), the participants were most interested in the target zone and those generalized functions connected with the target zones.

A further statistical analysis revealed: (1) the eye fixation on target zone 1 (i.e., VO1 + T1 + H1) with the FBS interface is significantly different from that with the EID interface, in particular FBS (37%) and EID (42.6%); (2) the eye fixation on target zone 2 (i.e., VO2 + T2 + H2) with the FBS interface was not significantly different from that with the EID interface, in particular FBS (21.9%) and EID

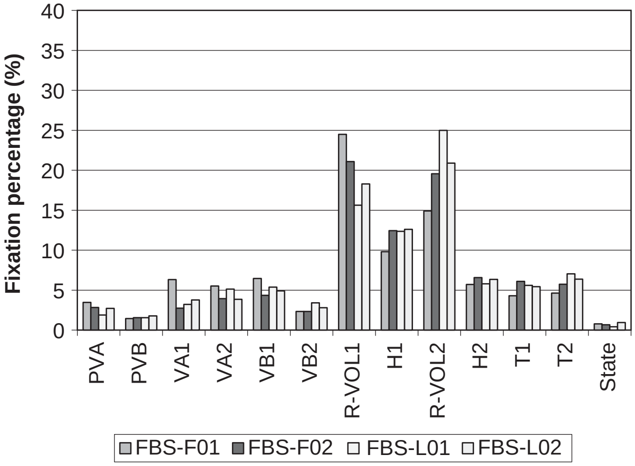

Fig 7. Eye fixation distributions over different regions in FBS interface.

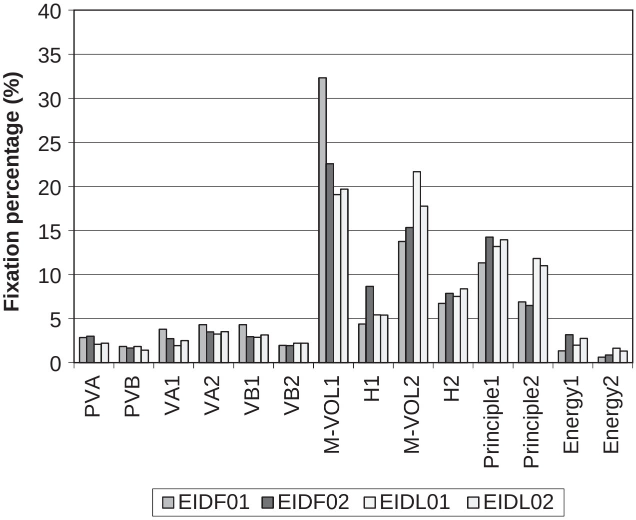

Fig. 8. Eye fixation distributions over different regions in EID interface.

(22%); and (3) the total percentage of eye fixations on the EID interface (95.8%) was significantly larger than those with the FBS interface (90.8%).

The above results on eye fixation are in agreement with the subjective measure, as the view from the participants was that the EID interface was more complex and more attention had to be paid to the interface in order to achieve similar comprehension of a problem. That is to say, the eye fixation on the EID interface should be of a higher percentage.

Table 8
Eye fixation for the components and subsystems

	VO1	VO2	T1	T2	H1	H2	State or Principle	VO1 + T1 + H1	VO2 + T2 + H2	Total
FBS (%)	19.9	20.1	5.3	5.9	11.8	6.1	0.6	37.0	21.9	90.8
EID (%)	23.4	17.1	13.2	9.1	6.0	7.6	3.4	42.6	22.2	95.8
Interface Pr >F*	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.792	<0.0001
Scenario	<0.0001	<0.0001	4E-04	<0.0001	<0.0001	0.01	0.112	<0.0001	0.0085	0.078

*Significance level $p<0.05$.

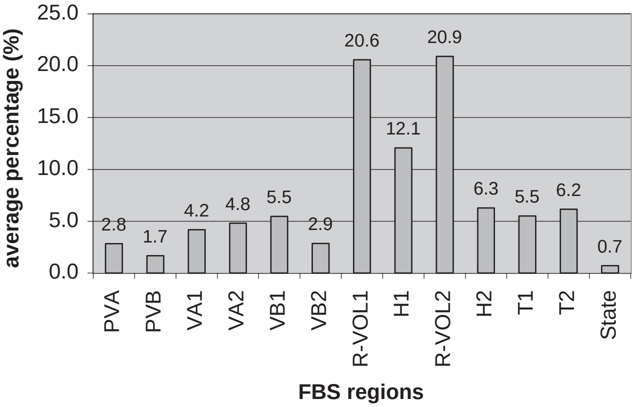

Fig. 9. Average percentage of eye fixation on the regions of FBS interface.

One general observation was made about using eye fixation as a tool to correlate the use of a particular piece of information (e.g., the abstract function). The common procedure seems to assert that the larger duration or higher number of eye fixation on a piece of information, the more attention on it and the more effective use of it; if in this case the corresponding trial also shows performance improvement, then a positive indicator or correction between the usefulness of that piece of information and the performance improvement could be established. This procedure is problematic in the sense that higher attention does not necessarily imply more effective use of that piece of information. Nevertheless, it should be possible to derive the following conclusion: if attention by any measure (e.g., eye fixation) to a particular piece of information is very low, that particular piece of information is unnoticeable (thus not useful). Therefore, in this connection, it can be concluded that the abstract function on the EID interface is not useful to the normal control operation and the fault detection task situations. It can also be concluded that the state network diagram in the function–behaviour–state framework is not useful; this is because the state network diagram is a subset of diagram in the behaviour/function window. This diagram may be useful in fault diagnosis and compensation.

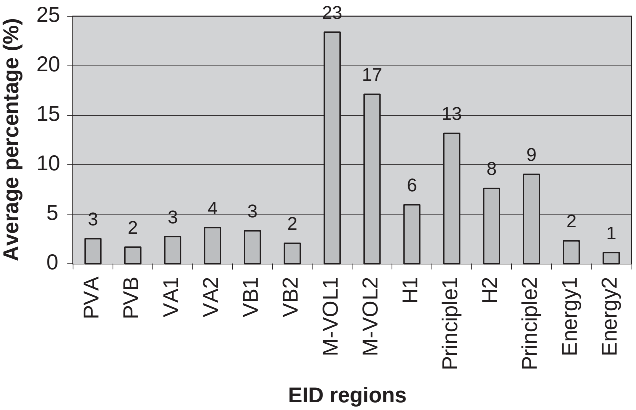

Fig. 10. Average percentage of eye fixation on the regions of EID interface.

For example, in a potential fault diagnosis task, the problem is assumed to be pump blow-up, and the root cause is the wrong operation procedure (e.g., the pump is set on before the valve is set on). The diagnosis procedure for this task may get to a certain stage where the states of the pump and the valve need to be checked. The states from the state network diagram can be viewed by clicking on the relevant components, the pump and the valve in this case.

5. Further discussion and related studies

5.1. The major controversial issue

Most studies in the literature on the experimental evaluation of interface design frameworks or interfaces appear relevant to the ecological interface design framework (Vicente, 2002). These studies applied experimental methods, such as verbal protocol, performance measure, memory recall, and subjective narrative description. None of these studies seem to have explicit accounts of mental workload. It is widely agreed that a display of the physical information and the physical function information only is not enough, regardless of any visual forms. There has been consistent verification of the need to display the generalized function. This study suggests that the most controversial issue is a need to display the abstract function and the questions of how and when to display. In general, the generalized goal of the mentioned studies was to verify the usefulness of abstract function in operation tasks, including monitoring (or fault detection), control (adjusting a plant system to new setpoints, and fault management (fault diagnosis and fault compensation). A summary of these studies is found in a recent paper by Vicente (2002). An exceptionally relevant study is the work that led to an alternative framework, i.e., the function–behaviour–state framework (Lin et al., 2001). If one notes the correspondence of these two frameworks and the scope of this paper, one can view this study as examining operator performance and operator mental workload in normal control operation and fault detection situations in the absence of the abstract function (in the context of the EID). The experimental result generated from the present study does not agree with some of the results in the existing literature. In the following, a critical analysis of selected studies among these studies is given for the purpose of elaborating on possible causes for the results obtained from this experimental study. There is also a theoretical analysis of the usefulness of the abstract function under normal control operation and fault detection task situations. This analysis also serves as a rationale for deriving the limitations of our study and for future work (these two issues will be discussed in the last section).

5.2. The operation task versus the information need

Variability in many experimental results is closely related to the categories of tasks and also to the difficulty levels of tasks in each of these categories. The experimental study result in Ham and Yoon (2001) implies that the display of all five levels of information may not positively improve performance (not to mention mental workload) in any operation. In particular, Ham and Yoon (2001) showed that the best performance of operators was achieved by displaying the generalized function (note that the physical function and functional purpose are always displayed) for a certain kind of fault diagnosis task, which is related to component failure. Further, they showed how the increase of task difficulty levels affects the usefulness of the abstract function; i.e., they considered that the difficulty level of tasks in their Experiment 2 is much higher than that of tasks in their Experiment 1. However, careful examination of the pools of tasks for Experiment 1 and for Experiment 2, respectively, hardly provided a sufficient clue about the quantitative measure of the task difficulty level for the tasks of two pools. Ham and Yoon (2001) did mention or imply that in Experiment 2, the task scenario was a mix of normal control operation and fault diagnosis, which is more “demanding”, in the words of the authors; this demanding effort was responsible for the increased level of usefulness of the abstract function. However, it could be argued that task demanding effort and difficulty level are different categories of concepts; they are not necessarily additive to one another. Experience from the present experimental study seems to suggest that when a demanding task is combined with a relatively short trial duration (e.g., 8 min), operators could have insufficient time to process the abstract function; the processing of the abstract function generally requires more attention resources because it reflects the deepest level of knowledge.

It is believed that in a task such as monitoring faults (or fault detection), the symptom was more easily captured, and this seems to imply that the abstract function in such a task situation may be less useful. Burns (2000b) showed, experimentally, that with the same amount of information displayed in different ways, the operator performance on fault detection (the detection time was considered) differed from that on fault diagnosis (the diagnosis time was considered). The comment made by Ham and Yoon (2001) in explaining why the abstract function does not work for certain tasks is a generalized finding regarding task versus information content: "…In the experiment, the PG showed a notably better performance than the PA display…Another, and simpler, explanation is also possible. The diagnosis may have required less use of AF-level information than GF-level information in experiment…" (p. 208). PG means the physical function plus the generalized function, PA means the physical function plus the abstract function, GF means the generalized function, and AF means the abstract function.

Janzen and Vicente (1998) experimentally showed that the P + F information led to a more accurate fault diagnosis. It is interesting to note that in the conclusion section of their paper they stated that fault diagnosis requires identifying broken constraints and was best accomplished by consulting the flow levels defined by the AH. The participants adopting this strategy exhibited the best performance on fault trials. Janzen and Vicente's statement implies that the flow levels information (i.e., the generalized function) is more useful than the abstract function, which agrees very well with the result obtained from Experiment 1 by Ham and Yoon (2001). Janzen and Vicente (1998) did not provide sufficient evidence that the abstract function is useful to a reduction of fault detection time.

Based on the above discussion, it appears clear that the result generated from the present experiment, i.e., the function–behaviour–state interface without the display of the extensional expected behaviour (Fig. 3) (which corresponds to the absence of abstract function in Fig. 2) achieves a better fault detection performance (in terms of detection successful rate) with yet a lower mental workload than the EID.

5.3. *Information display: what, how and when*

By definition, the framework is responsible for what information needs to be displayed. There are two other issues in interface design, that is, how to display and when to display. How to display includes the information content layout and the form of display. Interface design theory and methodology refer to all three issues. It was found in the literature that this connection is not clear. In Ham and Yoon (2001), the framework appears to be more concerned with answering the what-question, while in many other studies (e.g., Burns, 2000a) ecological interface design seems to cover all three issues. This differentiation is, however, very important because in many situations, the experimental result was intended to address the what-issue (e.g., the usefulness of the abstract function) although the experiment was dependent on the how- and when-issues. In other words, when the experimental result shows, for example, the usefulness of the abstract function, one can hardly perceive whether this is solely because of the presence of the abstract function or perhaps it is because of an effect from an aggregated combination of both the presence of the abstract function and the way and/or the time in which it is displayed. Janzen and Vicente (1998) explained their contradictory result regarding diagnosis accuracy and compensation time as the cause of the "graphic emergent features of the Principles display" (p. 539). Ham and Yoon (2001) also appeared conscious of this differentiation; however, their result was obtained with a mix of the what-factor and a part of the how-factor (the information layout).

There are two serious concerns in the context of this paper: the reliability and the practicalness of experimental results. Confusion arising from the scope of interface design framework is one of the causes for variability in many experimental results, including the one presented in this paper. Reliability is obviously related to the applicability of a particular experiment. There is a dilemma presented in the reliability and the practicalness. Ham and Yoon's work (2001) may produce more reliable results as they eliminated the effect of forms of information contents, but their results and conclusions may not be generalized to the a more practical situation—the display forms (which are more sophisticated) of the two interfaces (Figs. 2 and 3).

5.4. Brief theoretical analysis of the role of the abstract function

The nature of the abstract function is the constraint equation among a set of relevant state variables (the active state variable and the passive state variable in the terminology of this paper). For example, in the case of the DURESS plant system (e.g., tank 1), this set of variables includes: the flow rate entering the tank (further corresponding to the two valves, VA1 and VB1), the flow rate out of the tank (further corresponding to the valve VO1, the energy entering the tanks (further corresponding to the heater), and the energy out of the tank. The constraint equation is derived from physical laws, chemistry laws, and other axioms from various science and technology domains (e.g., the design axioms for design of general systems; see Suh, 1990). In the EID framework (see Vicente, 1999), the constraint equation is represented by a differential equation or some general algebraic expression as a solution to the differential equation (the latter point might just be speculation) (p. 170). From the point of view of the mathematical representation of physical and/or chemical behaviours of an underlying system, it is clear that neither the differential equation nor the general algebraic solution accurately represents the behaviour of the system per se because the system boundary and initial conditions have not been considered. Instead, the very idea in the FBS framework (see the previous discussion in Section 2) relevant to the abstract function in the EID framework is that a specific algebraic solution should be considered as a level of information stored in an interface and displayed appropriately. This study further elaborated on the behaviour having two aspects: intensional and extensional. The intensional is the specific solution to the general differential equation, which should not be displayed to the plant operator. The extensional is a series of related active states and passive states that satisfy the intensional part of behaviour, which should be displayed. Furthermore, there are two kinds of extensional behaviours: the expected extensional and the actual extensional. The former is subject to the intensional part of behaviour (i.e., the constraint equation). The latter is based on the information of sensors (e.g., sensors for the level of water in a tank). Although it is clear that the abstract function in the EID interface, as it is defined, corresponds to the intensional behaviour, it is not very clear whether the display of the abstract function as shown in Fig. 2 corresponds to the expected extensional or the actual extensional behaviour.

Vicente (1999) indicated that the very general technique for fault detection and fault management is analytical redundancy (Frank, 1990). In the context of the process control of a plant, the model (the constraint equation in general) is evaluated to derive the plant outputs, and then the expected outputs are compared with the actual outputs (which are measured from the sensors) to determine if there is any discrepancy. However, the discrepancies may not necessarily mean any fault; they may mean that a plant has not reached its targets (prescribed by a set of setpoints) under the steady state of the plant. Therefore, when the task scenario is a mix of control and fault detection or fault diagnosis, the abstract function does not help in a theoretical sense. This analysis implies that the result generated by Ham and Yoon (2001), which shows a significant performance improvement with the presence of the abstract function (plus the generalized function) to the mix of control operation and fault diagnosis (including fault detection), needs a further careful analysis of causes.

Assume that there is the presence of the generalized function and the functional purpose in the terminology of the EID framework. It can theoretically be explained why the abstract function either does not help or provides a redundant help under the normal operation and fault detection task situations. Fault means that there is at least one broken constraint event. The constraint can be in the form of the generalized function or the abstract function. In the following, three typical cases are explained.

Case 1: When a fault occurs on a component, which is not directly related to the storage component (the tank in Fig. 2), e.g., valve VA, its generalized function will be violated and the sensor should get an abnormal state. In this case, the fault is identified by the abnormal symptoms, not by the mass and energy imbalance which may be shown in the sophisticated graphic representation of the abstract function (Fig. 2). This point is confirmed by Ham and Yoon (2001) based on their experimental result and their analysis, positively stated by Janzen and Vicente (1998), and verified in the present study as well.

Case 2: A fault exists directly related to the tank (Fig. 2), e.g., on valve VA1. First of all, the symptom of this case is the same as that of Case 1. It is true that in this case, the abstract function (mass conservation part) will cause the rate of volume change in the tank to be greater or smaller than "normal" depending on the nature of the problem in valve VA1. The graphical representation of the mass conservation in Fig. 2 may show that Mass-in is greater or smaller than Mass-out. It is clear, however, if an operator detects this fault not by looking at the abnormal symptom (or the generalized function or flow function in the terminology of the EID) but instead by reasoning about the graphical representation of the abstract function per se, not only extra attention (which increases mental workload) needs to be paid, but also a wrong fault detection result could be produced as the cause of the imbalance of the mass conservation may also be the problem of valve VO1, for example.

Case 3: When a fault exists on any situation other than those state variables that are called the active state variables, say the pump on/off, the valve opening/closing, etc. (e.g., unknown extra water pours into the tank), the symptom can still be captured by the generalized function (i.e., the rate of volume change in the tank will be either greater or smaller than normal). An argument may be made that the symptom in this case is derived from the broken generalized function; an interface may not provide the rate of volume change in the tank but instead may provide information of the abstract function (mass-in versus mass-out). In this situation, it is true that there is no clue to the fault because of the absence of the related generalized function, yet there is a clue to the fault based on the information possibly displayed, such as mass-in > mass-out. However, as analysed in Case 2 above, a broken mass conservation law does not sufficiently exclude other possible cases (instead of the extra water being poured into the tank). On a general note, a situation that fails to give the operator the information about the rate of volume change in the tank is a mistake in interface design; this mistake can be called the incomplete provision of the generalized function.

In the impact of the abstract function to the normal control operation, there are basically two cognitive and manipulative issues: (1) a goal of action and (2) a plan of action. The functional purpose in the EID interface gives the information for the first issue. With regard to the second issue, one needs to examine whether the abstract function, as it is defined and displayed, could provide a clue for operators to develop an effective plan in their mental world with which they can effectively manipulate the widgets of these active state variables to make the plant system reach its goal.

As discussed earlier, the abstract function in the EID interface is expressed by a kind of differential equation (Vicente, 1999). The constraint created by the differential equation is valid to more situations than a specific situation that aligns with a specific boundary condition of a plant. This point can be best explained by the example of the analysis of a robot mechanism. Fig. 11 shows a five-bar mechanism. The two actuators are represented as the active state variables, and they are manipulated by operators. The control goal is to make the end-effect point Q reach a new setpoint $\tilde{Q}$. There are two branches with this mechanism (see Fig. 11a, Q and $\tilde{Q}$). The kinematic constraint equation for this mechanism is $\vec{l}_1 + \vec{l}_3 = \vec{l}_5 + \vec{l}_2 + \vec{l}_4$ (Shigley and Uicker, 1995). From kinematics, it is known that both branches (Q, Q') satisfy this loop constraint equation.

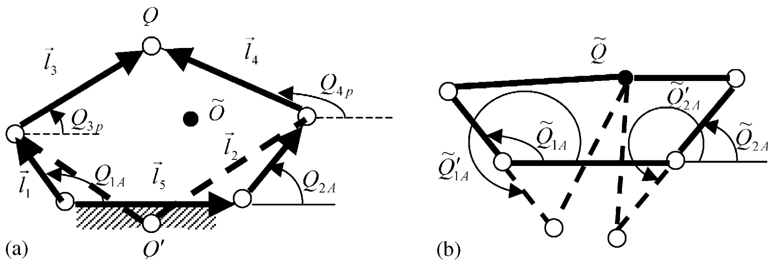

Fig. 11. A five-bar robot mechanism. (a) The schematic kinematic diagram (the dashed line represents a different branch other than that represented by the solid line. Q_{1A}, Q_{2A} are active state variables and Q_{3A}, Q_{4A} are passive state variables. (b) Two branches (the solid line: branch A; the dashed line: branch B) that satisfy the loop constraint equation. Furthermore, $\tilde{Q}_{1A}, \tilde{Q}_{2A}$: the active state variables when the target is reached from branch A; $\tilde{Q}'_{1A}, \tilde{Q}'_{2A}$: the active state variables when the target is reached from branch B).

Clearly the question now is, is there any unique role with the abstract function in the aspect of producing an effective action plan in the operators' mental world? The abstract function may play a role in creating information to operators about how "far" the current state of the plant is away from the target state which is defined as a new setpoint ($\tilde{Q}$ in Fig. 11) and as DT, DV in the DURESS plant (in Fig. 2). This role is fulfilled by showing, for example, mass-in greater or smaller than mass-out in the case of the DURESS plant, or by showing that the loop equation is violated in the case of the robot mechanism. However, a direct examination of the difference between the flow rate of the water at the outlet of valve VO1 and the desired flow rate DV1 should be far more effective and efficient in fulfilling that role. For the robot mechanism (Fig. 11), a direct examination of the difference between the coordinate of Q and the coordinate of $\tilde{Q}$ is more effective and efficient in fulfilling that role.

As to the question of whether the abstract function could give operators a sort of effective path to reach the target goal (for example, in the case of the robot mechanism driven by two actuators (Q_{1A}, Q_{2A}) whether Q_{1A}, Q_{2A} should be increased or decreased, for Q to be adjusted to $\tilde{Q}$), the abstract function may wrongly guide operators in such a way that the mechanism does reach the target $\tilde{Q}$ but through a branch different from its starting branch. In Fig. 11(b), branch A is the starting branch (which corresponds to the specific solution to the loop constraint equation) and branch B is the other one also satisfying the constraint equation; both branches make the mechanism possibly reach the target $\tilde{Q}$. This wrong guidance is possible because the abstract function, the loop constraint equation in this case, does not rule out branch B.

In conclusion, the abstract function, as it is defined and displayed in the literature, plays a redundant role with respect to the functional purpose in the first basic issue in the normal control task, and plays no role, being possibly insensitive to the incorrect action plan which may be executed by operators, in the second basic issue. It is noted that the function–behaviour–state interface design approach can offer assistance to operators in addressing the second basic issue; a separate report is underway.

6. Conclusion

6.1. The main conclusions

The present study aimed at evaluating two human–computer interface development frameworks, namely the ecological interface design framework and the function–behaviour–state interface design framework. The former is named EID, and the latter is named FBS for short in the present discussion. A comprehensive experiment with three categories of measures (i.e., the performance measures, eye movement measures, subjective rating measures) was conducted, together with a theoretical analysis of the experimental results. In the experiment, the tasks include the normal control operation and fault detection situations. The participants were those whose knowledge is compatible to that behind the process of the simulated plant used in the experiment. The following conclusions are derived from this experiment.

(1) The mental model based on FBS is the one rooted in a generic model of any system (including the plant system) in the computer-aided design of the system (e.g., Zhang, 1994; Zhang et al., 1997), while the mental model based on EID is the one rooted in unanticipated problem solving. The former is more easily acceptable to the plant operator than the latter.

(2) The abstract function in the EID interface is not positively correlated with performance improvement for the normal control operation and fault detection tasks, yet its visual complexity, compared with the interface without the abstract function, can cause higher mental workload. This experimental result is further in agreement with our theoretical analysis of the role of the abstract function for these two task situations. This result is also in agreement with the result presented by Experiment 1 in Ham and Yoon (2001).

(3) The eye fixation measurement further shows that the operator pays little attention to the abstract function for the normal control operation and the fault detection tasks, which provides another confirmation of the second conclusion.

(4) The eye fixation measurement confirms that the operator pays little attention to the state network diagram in the FBS interface, which means that the state network diagram is of no use to the normal control operation and fault detection tasks.

(5) The eye fixation measurement, the performance measurement, and the subjective measurement are of good consistency in the experiment.

(6) The change in the pupil diameter is not sensitive to the characterization of mental workload in this experimental condition.

6.2. *Limitations and future work*

First, the present experiment has not considered fault diagnosis and fault compensation. Consequently, the conclusion cannot be generalized to these two task situations. Further studies are considered to compare FBS and EID in fault diagnosis and fault compensation task situations. In that intended work, a full FBS interface which includes the mechanism to display the expected extensional behaviour at the system level, the subsystem level, and the component level will be shown. It is believed that the display of information should be made adaptive to different task categories; in particular, expected extensional behaviour for the fault diagnosis and compensation tasks will be displayed.

The *second* limitation is that the characterization of tasks needs to be enhanced. The goal of the characterization is to develop a quantitative measure about the task demand level and the task difficulty level. It should be very interesting to move forward this work to fault diagnosis and fault compensation with the comprehensive task characterization. The conclusion, made by Ham and Yoon (2001), is further speculated, that the positive indicator of the abstract function to the performance improvement is due to the increased level of task difficulty in their second experiment.

The *third* limitation of the present study is that the expertise of operators affects their performance and mental workload characterization has not been considered. Further work needs to address this issue. To move towards this end, expertise level of operators needs to be quantified. An ANOVA based analysis could help to determine the level of expertise versus the performance and mental workload.

Finally, there are two general approaches in determining whether or not a particular level of information is useful. The first general approach is to design an interface with or without a particular level of information in question and then analyse the experimental result. Analysis of the result can lead to the conclusion of whether or not that particular level of information has a significant impact on the performance. The present study falls into this general approach. The second general approach is to examine "how" operators "use" a particular level of information while in the meantime recording the performance. After that, it is possible to develop a correlation between the performance and the use of that particular information. The work given by Janzen and Vicente (1998) falls into this general approach. Theoretically speaking, the second general approach should lead to a more accurate picture of what is going on in the process. However, contemporary studies on EID usually rely on the frequency of visits and the dwell time to a particular level of information (e.g., the abstract function), e.g., Janzen and Vicente (1998), Burns (2000a). This technique needs to be carefully examined, as a visit does not mean a correct use of particular information. Future work includes plans to develop a more reliable experimental procedure not only to examine visits but also to examine how the operator uses a particular level of information in reasoning.

6.3. Contributions

Despite these limitations, the study presented in this paper has provided a new angle for evaluating the EID framework and its interface described in the studies. The experimental methods used in this study, especially including the eye movement parameters (the eye fixation and the change in the pupil diameter), considerably differ from the existing studies on this subject which are essentially based on the participants' verbal exposure and the experimenters' manual record with the help of a video system. From the point of view of experimental design in statistics, the method used in this paper does not differ from existing studies, i.e., factorial design, ANOVA, and post hoc analysis (e.g., Duncan); however, the inclusion of three categories of measures in this experiment gives relatively higher reliability to the results generated from it. Some of the results are in agreement with those produced in the existing studies, but some others are not, which should initiate more rigorous experimental studies on the EID in a more comprehensive way.

More specifically, this study has given a more comprehensive explanation on the function–behaviour–state framework and its interface, though the level of information in this framework called the expected extensional behaviour is not shown. This study suggests that the function–behaviour–state framework should be the mental model of operators in the case of process plants, as this framework is the same as the most generic model for the design of plants. The study also implies that the operator mental model of a plant should be the same as the design model of the plant. More generally, the following speculations are provided by the study: the whole mental model of operators should have three elements: the function–behaviour–state framework, the system decomposition, and the analytical redundancy for fault detection, diagnosis, and compensation. Furthermore, the display of these levels of information should be such that in the normal control operation and fault detection task situations, only the state and the extensional actual behaviour (corresponding to the generalized function in the EID) should be displayed; while in the fault diagnosis and fault compensation task situations, the expected extensional behaviour (the component level, the subsystem level, or the system level) should be displayed.

The theoretical proof of either no use or redundant use of the level of information called the abstract function in the EID to the normal operation control and the fault detection task situations is first given in this paper. This proof is made through a more generic concept in the problem solving called analytical redundancy (e.g., Frank, 1990; Vicente, 1999).

It has also been shown that the change in the pupil diameter is generally not sensitive to the mental workload characterization to the application of evaluating an interface under situation of normal control operation and fault detection task situations in the process plant.

Acknowledgements

The research reported here is supported by the Natural Sciences and Engineering Research Council (NSERC) and the Atomic Energy of Canada Limited (AECL) through a CRD grant awarded to the second author. The authors would like to thank A. B. Thornton-Trump from the University of Manitoba, Canada for his support to set up the Eyegaze experiment apparatus and fulfil of the experiment described in this paper. The authors would also like to thank Kim J. Vicente for his kind supply of the DURESS software system upon which a further implementation of the new features is facilitated.

References

Backs, R.W., Walrath, L.C., 1992. Eye movement and papillary response indices of mental workload during visual search of symbolic displays. Applied Ergonomics 23 (4), 243–254.

Beatty, J., 1982. Task-evoked papillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91, 276–292.

Bracewell, R., Sharp, J., 1994. A computed aided methodology for the development of conceptual schemes for mixed energy-transforming and real-time information systems. In: Sharp, J., Ho, V. (Eds.), Computer Aided Conceptual Design. LooseLeaf Co., UK, pp. 79–94.

Burns, C.M., 2000a. Navigation strategies with ecological displays. International Journal of Human–Computer Studies 52, 111–129.

Burns, C.M., 2000b. Putting it all together: improving display integration in ecological displays. Human Factors 42 (2), 226–241.

Burns, C.M., Thompson, L.K., Rodriguez, A., 2002. Mental workload and the display of abstraction hierarchy information. 45th Human Factors and Ergonomics Annual Meeting, Baltimore, MD, pp. 235–239.

Christoffersen, K., Hunter, C.N., Vicente, K.J., 1996. A longitudinal study of the effects of ecological interface design on skill acquisition. Human Factors 38 (3), 523–541.

De Waard, D., 1996. The measurement of drivers' mental workload. Ph.D. Thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Center. [http://home.zonnet.nl/waard2]

Frank, M.P., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. Automatica 26, 459–474.

Goldberg, J.H., Kotval, X.P., 1998. Eye movement-based evaluation of the computer interface. In: Kumar, S.K. (Ed.), Advances in Occupational Ergonomics and Safety. IOS Press, Amsterdam, pp. 529–532.

Granholm, E., Asarnow, R.F., Sarkin, A.J., Dykes, K.L., 1996. Pupillary responses index cognitive resource limitations. Psychophysiology 33, 457–461.

Ham, D.H., Yoon, W.C., 2001. Design of information content and layout for process control based on goal-means domain analysis. Cognition, Technology and Work 3, 205–223.

Hess, E.H., Polt, J.M., 1964. Pupil size in relation to mental activity during simple problem-solving'. Science 143, 1190–1192.

Hutchins, E.L., Hollan, J.D., Norman, D.A., 1986. Direct manipulating interfaces. In: Norman, D.A., Draper, S.W. (Eds.), User Centered System Design: New Perspectives on Human–Computer Interaction. Erlbaum, Hillside, NJ, pp. 87–124.

Hyona, J., Tommola, J., Alaja, A-M., 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. The Quarterly Journal of Experimental Psychology 48A (3), 598–612.

Itoh, J., Sakuma, A., Monta, K., 1995. An ecological interface for supervisory control of BWR nuclear power plants. Control Engineering Practice 3 (2), 231–239.

Janzen, M.E., Vicente, K.J., 1998. Attention allocation within the abstract function. International Journal of Human–Computer Studies 48, 521–545.

Jorna, P.G.A.M., 1997. Context simulation: an interactive methodology for user centered system design and future operator behaviour validation. Technical Report NLR-TP-97032, National Aerospace Laboratory NLR, Amsterdam, The Netherlands. Also available on website: http://atm-seminar-97.eurocontrol.fr/jorna.htm.

LC technologies, 2000. The Eyegaze Development System for Windows NT, User's Guide (March). LC Technologies, Inc., Fairfax, Virginia, USA.

Lin, Y., 2000. Experimental study based on eye gaze measurement for computer interface. Technical Report (AEDL-2000-L7Z01), Advanced Engineering Design Laboratory, Department of Mechanical Engineering, University of Saskatchewan, Canada, 105 pages.

Lin, Y., Zhang, W.J., Watson, L.G., 2001. On proposal of function–behaviour–state framework as refinement of EID framework of human–machine interface design. In: Eiji Arai, Tatsuo Arai, Masaharu Takano (Eds.), Human Friendly Mechatronics. Elsevier, Osaka, Japan, pp. 61–66.

Lin, Y., Zhang, W.J., Watson, L.G., 2002. A study of ergonomic measurement indices in evaluating human–machine interfaces. 45th Human Factors and Ergonomics Annual Meeting, Baltimore, MD, pp. 61–66.

May, J.G., Kennedy, R.S., Williams, M.C., Dunlap, W.P., Brannan, J.R., 1990. Eye movement indices of mental workload. Acta Psychologica 75, 75–89.

Montgomery, D.C., 1997. Design and Analysis of Experiments, 4th Edition. Wiley, New York.

Moray, N., Rottenberg, I., 1989. Fault management in process control: eye movements and action. Ergonomics 32 (11), 1319–1342.

Norman, K.L., Weldon, L.J., Shneiderman, B., 1986. Cognitive layouts of windows and multiple screens for user interfaces. International Journal of Man–Machine Studies 25, 229–248.

Pahl, G., Beitz, W., 1995. Engineering Design: A Systematic Approach, 2nd Edition. Springer, London.

Peavler, W.S., 1974. Pupil size, information overload, and performance differences. Psychophysiology 11, 559–566.

Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols. Eye Tracking Research and Applications Symposium 2000, Palm Beach Gardens, FL, USA, pp. 71–78.

Shigley, J.E., Uicker, J.J., 1995. Theory of Machine and Mechanisms, 2nd Edition. McGraw-Hill, Inc., New York.

Stern, J.A., 1997. The pupil of the eye: what can it tell us about 'mental processes'? Human Engineering for Quality Life (HQL) Quarterly 8, (website: http://www.hql.or.jp/gpd/eng/www/nwl/n08/ergo.html), 1–2.

Suh, N.P., 1990. The Principles of Design. Oxford University Press, Oxford.

Umenda, Y., Takeda, H., Tomiyama, T., Yoshikawa, H., 1990. Function, behaviour and structure Applications of Artificial Intelligence in Engineering. In: Gero (Ed.), Proceedings of the Fifth International Conference, Boston, MA.

Vicente, K.J., 1991. Supporting knowledge-based behaviour through ecological interface design. Unpublished Doctoral Dissertation, University of Illinois, Urbana-Champaign.

Vicente, K.J., 1999. Cognitive Work Analysis, Toward Safe, Productive, and Healthy Computer-Based Work. Lawrence Erlbaum Associates, Publishers.

Vicente, K.J., 2002. Ecological interface design: progress and challenges. Human factors 44 (1), 62–78.

Vicente, K.J., Rasmussen, J., 1990. The ecology of human–machine systems II: mediating 'direct perception' in complex work domains. Ecological Psychology 2, 207–250.

Vicente, K.J., Rasmussen, J., 1992. Ecological interface design: theoretical foundations. IEEE Transactions on Systems, Man and Cybernetics 22 (4), 589–606.

Vicente, K.J., Christoffersen, K., Pereklita, A., 1995. Supporting operator problem through ecological interface design. IEEE Transactions on Systems, Man and Cybernetics 25 (4), 529–545.

Washburn, D.A., Putney, R.T., 1995. Pupillometric indices of dimensions of difficulty in visual-task performance. Proceedings of the 39th Annual Meeting of Human Factors and Ergonomics Society, Santo Monica, CA, pp. 949–954.

Wickens, C.D., 1992. Engineering Psychology and Human Performance, 2nd Edition. HarperCollins, New York.

Zhang, W.J., 1994. An Integrated Environment for CAD/CAM of Mechanical Systems. Unpublished Doctoral Dissertation, Delft University of Technology.

Zhang, W.J., Zhang, D., Van Der Werff, K., 1997. An integrated data model of function, behaviour, and structure of computer-aided conceptual design of mechanisms. In: Usher, J.M., Roy, U., Parsaei, H.R. (Eds.), Integrated Product and Process Development—Methods, Tools, and Technologies. Wiley, New York, pp. 85–124.

Zijlstra, F.R.H., Doorn, L. van, 1985. The construction of a scale to measure subjective effort, Technical Report, Department of Philosophy and Social Sciences, Delft University of Technology.