THEOR. ISSUES IN ERGON. SCI., 2002, VOL. 3, NO. 2, 124–142

Toward theory-driven, quantitative performance measurement in ergonomics science: the abstraction hierarchy as a framework for data analysis

XINYAO YU, ELFREDA LAU, KIM J. VICENTE* and MICHAEL W. CARTER

Cognitive Engineering Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada

Keywords: Performance measurement; Abstraction hierarchy; Coupling.

Measurement in ergonomics science has not kept pace with theorizing. As a result, it is rare to find measures of human performance that are simultaneously objective, quantitative, sensitive, and theoretically grounded. This article proposes a new set of measures, based on the abstraction hierarchy (AH) framework, that satisfies all of these criteria. Each level of the AH can be used to define a quantitative state space that can serve as a frame of reference for objective measurement. These state spaces are complementary because they provide different views of the same human–environment behaviour. Collectively, this set of measures can be used to determine if a participant is strongly or weakly coupled to functional or physical distal properties of the work domain. Data from a longitudinal study are used as a case study to test the value of these novel measures. The empirical results show that these AH-based measures provide unique insight into participants' behaviour that was not revealed by many, more traditional measures of performance. Because it is theoretically grounded, the set of measures proposed here has the potential to be generalized to diverse work domains for which it is possible to develop an AH representation.

1. Introduction

The lack of sophisticated measurement in complex experimental settings poses a significant obstacle to ergonomics science (e.g. Moray *et al*. 1986, Moray and Rotenberg 1989, Sanderson *et al*. 1989, Howie and Vicente 1998). Traditional measures (e.g. task completion time) are objective, but frequently do not have a compelling theoretical basis and are not sensitive enough to reveal differences between experimental groups. Other measures (e.g. eye movements, verbal protocols) can provide greater scientific insight, but frequently suffer from being extremely time-consuming or subjective to analyse. It would be useful to develop novel empirical measures that are: (a) *objective*, meaning that they can be derived solely from log files of participant actions and system state; (b) *quantitative*, meaning that they can be derived computationally; (c) *theoretical*, meaning that they have a close connection to meaningful constructs; and (d) *sensitive*, meaning that they provide novel empirical insights that cannot be observed using traditional measures. The novel class of measures proposed in this article is based on the abstraction hierarchy (AH; Rasmussen 1985) and satisfies all four of these criteria, each of which is routinely used to evaluate measurement in many sciences.

* Author for correspondence. e-mail: vicente@mie.utoronto.ca

Theoretical Issues in Ergonomics Science ISSN 1463–922X print/ISSN 1464–536X online
http://www.tandf.co.uk/journals
DOI 10.1080/14639220210124102

The remainder of the article is organized as follows. First, the experimental test-bed used for this research will be introduced. Secondly, the theoretical motivation behind the AH framework will be described. Thirdly, the proposed empirical measures and their connection to the AH will be explained. Finally, the empirical sensitivity of these measures will be illustrated by using data from a 6-month longitudinal study of the impact of interface design on human performance.

2. Experimental test-bed

This research was conducted in the context of the thermal-hydraulic process control microworld illustrated in figure 1 (Pawlak and Vicente 1996). This micro-world is a real-time, interactive simulation that was designed to be representative of industrial process control systems, thereby promoting the generalizability of results from the laboratory to the field (Vicente 1991).

The micro-world consists of two redundant feedwater streams (FWSs) that can be configured to supply water to either, both, or neither of the two reservoirs. Each reservoir has associated with it an externally determined demand for water that can change over time. The work domain purposes are twofold: to keep each of the reservoirs at a prescribed temperature (40 and 20°C), and to satisfy the current mass (water) output demand rates. To achieve these purposes, participants have control over eight valves (VA, VA1, VA2, VO1, VB, VB1, VB2 and VO2), two pumps (PA and PB), and two heaters (HTR1 and HTR2). All of these components are governed by first order lag dynamics, with a time constant of 15 seconds for the heaters and 5 seconds for the remaining components.

A number of other variables, not displayed in figure 1, can be used to describe the operation of the simulation. Definitions of these variables are provided in table 1.

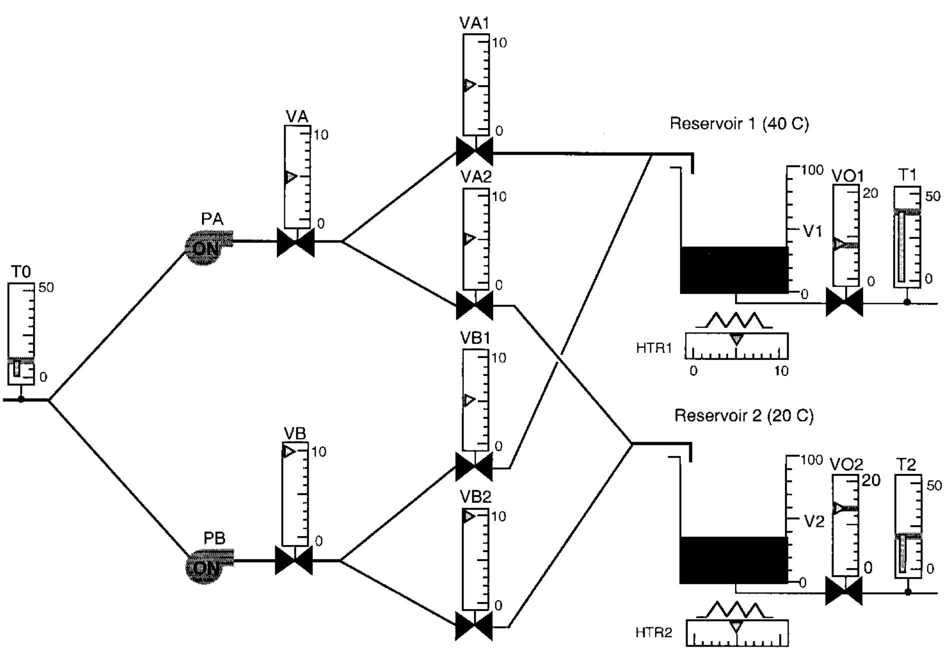

Figure 1. Thermal-hydraulic process control microworld (adapted from Pawlak and Vicente 1996).

Table 1. Definition of microworld process variables.

Variable	Description
Temperature	
T1	Temperature of res 1
T2	Temperature of res 2
Output demand	
D1	Output flowrate for res 1
D2	Output flowrate for res 2
Mass	
MO1	Mass output flowrate for res 1
MO2	Mass output flowrate for res 2
MI1	Mass input flowrate for res 1
MI2	Mass input flowrate for res 2
M1	Mass inventory of res 1
M2	Mass inventory of res 2
Energy	
E1	Total energy stored in res 1
E2	Total energy stored in res 2
EI1	Energy input flowrate for res 1
EI2	Energy input flowrate for res 2
EO1	Energy output flowrate for res 1
EO2	Energy output flowrate for res 2
Heat transfer	
FH1	Flow from heater HTR1
FH2	Flow from heater HTR2
Flowrates	
FA1	Flowrate from valve VA1
FB1	Flowrate from valve VB1
FA2	Flowrate from valve VA2
FB2	Flowrate from valve VB2
FPA	Flowrate from pump PA
FPB	Flowrate from pump PB
FVA	Flowrate from valve VA
FVB	Flowrate from valve VB
Heaters	
HTR1	Setting for heater of res 1
HTR2	Setting for heater of res 2
Pumps	
PA	Setting of pump in fws A
PB	Setting of pump in fws B
Valves	
VA	Setting of initial valve in fws A
VB	Setting of initial valve in fws B
VA1	Setting of valve 1 in fws A
VB1	Setting of valve 1 in fws B
VA2	Setting of valve 2 in fws A
VB2	Setting of valve 2 in fws B
VO1	Setting output valve in res 1
VO2	Setting of output valve 2 in res 2

fws = feedwater stream, res = reservoir.

The role that these quantitative variables play in the novel measures proposed in this article is discussed later.

3. Theoretical motivation

Explanations of the AH are usually theoretically motivated by psychological theories of human problem solving (e.g. Rasmussen 1985, Vicente 1999). However, it is also possible to explain the scientific value of the AH from the theoretical perspective of systems engineering or control theory (Vicente 1991).

3.1. *The inverse dynamics problem*

Figure 2 provides one of the simplest possible representations of a human–machine system—a negative feedback control loop. The box labelled *work domain* represents the (forward) dynamics of the controlled system. It is a model of how the actions of the worker are translated into outputs that are relevant to the goal(s) of interest. From the perspective of the worker, however, the important question is 'given where I should be, what should I do to get there'? This is frequently referred to as the *inverse dynamics problem* in systems engineering. As shown in figure 3(*a*), it involves going from the error signal (i.e. the difference between where you are and where you want to be) to what you should act on (i.e. work domain components). Unfortunately, this mapping is exceedingly complex because there are so many components that can be acted on and many interactions that must be taken into account. It is very difficult for resource-limited actors, such as workers, to solve this inverse dynamics problem unaided.

In the case of the micro-world in figure 1, the inverse dynamics problem requires participants to develop a mapping between two sets of variables, the first describing the four work domain purposes (T1, T2, D1, D2) and the second describing the 12 components that can be acted on (PA, PB, HTR1, HTR2, VA, VA1, VA2, VO1, VB, VB1, VB2, VO2). The decision as to how to act is non-trivial because there seem to be so many degrees of freedom. Furthermore, the relationship between the work domain purposes and the components that can be acted upon is far from obvious. The mapping between the two is essentially a black box from the viewpoint of workers (see figure 3(*a*)), unless designers provide some support to help workers solve the inverse dynamics problem. Thus, it is hard for participants to decide what to act on, given knowledge of where they want to be.

There are at least three different ways for workers to solve this problem (cf. Christoffersen *et al.* 1997). The first approach, *trial and error*, involves acting on components in a haphazard way, examining the result, and then iteratively making another change. Trial and error is cognitively economic because it does not require much thought, and although it can succeed in the long-run, it is clearly not very efficient. The second approach, *heuristics*, involves developing rules of thumb that

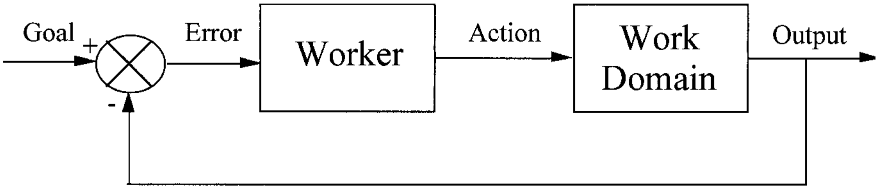

Figure 2. A negative feedback control loop.

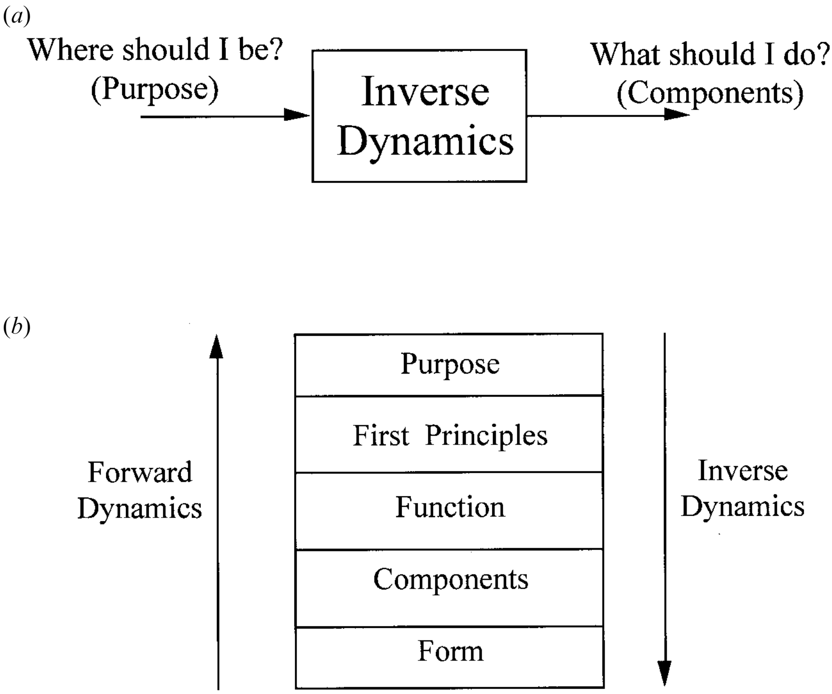

Figure 3. (*a*) The inverse dynamics problem. (*b*) The AH provides stratified feedback for solving the inverse dynamics problem.

map particular states onto particular actions (e.g. if T1 is too high, then lower HTR1). Heuristics are also cognitively economic and can be efficient. However, it can take a great deal of practice to acquire heuristics. Furthermore, because they are rules of thumb, heuristics are—by definition—fallible (i.e. they will not always work). The third approach, *model-based derivation*, involves computing the actions to be performed by using a mental model of the work domain dynamics. In the absence of faults, this approach is reliable because it is based on an understanding of the how the work domain functions. However, it is cognitively uneconomic because it requires a great deal of knowledge, memory load and computational power. None of these three approaches to solving the inverse dynamics problem is particularly attractive, because each has important limitations.

In operational settings, workers usually avoid trial and error because the consequences of an error can be quite severe. Model-based derivation is usually avoided because it is not possible to perform, given human information processing capabilities. Thus, workers frequently rely on formal or informal procedures (a form of heuristics) to cope with the inverse dynamics problem in a cognitively manageable, if fallible, way. For example, table 3 provides an example of the type of heuristics reported by a highly experienced participant in a longitudinal micro-world study described later (Christoffersen *et al.* 1998).

3.2. *Abstraction hierarchy*

Designers can help workers solve the inverse dynamics problem by providing feedback so as to open up the 'black box' in figure 3(*a*). The AH can be viewed as a modelling framework to do precisely that. Rather than requiring workers to solve the inverse dynamics problem in one complex step—a cognitively daunting task—the AH provides a hierarchically-organized set of work domain models that allows workers to transform knowledge of where they should be (i.e. the current state of the purposes) to decisions about what they should do (i.e. how to act on components) in several stratified steps. As shown in figure 3(*b*), the top level of the AH provides workers with information about the state of the work domain purposes, whereas the bottom levels provide workers with information about the form of the components on which they can act. The levels in between show how these two entities are structurally related. By providing feedback at each of these levels of abstraction, the black box is now made transparent because the relationship between purpose and form is shown as a stratified hierarchy. Furthermore, the links between levels can guide a workers' search process by showing which lower-level objects are relevant to the current higher-order function of interest. As a result, the inverse dynamics problem is made easier to solve, because it no longer has to be solved all in one complex step. Workers can focus on the high-level objectives to be achieved rather than the myriad specific actions that might be taken in any particular context.

A familiar example is keeping one's car between lane markers while driving. Because the state of the work domain (i.e. the position of the car) is clearly visible with respect to the goal (i.e. the lane markers), experienced drivers do not have to memorize a procedure for how to control the car. Instead, they can just rely on the feedback from the environment to guide their actions directly in a goal-directed fashion. Table 4 provides an example of this type of high-level, functional control reported by a highly experienced participant in a longitudinal micro-world study described later (Christoffersen *et al.* 1998). The contrast with the low-level, action-based heuristic in table 3 is notable.

In short, the AH framework is intended to provide a mechanism for coping with complexity. The AH is usually used in conjunction with a decomposition (or part-whole) hierarchy that describes the work domain at various layers of resolution. Higher levels describe the work domain at a coarse level, whereas lower levels describe the work domain at a more fine-grained level.

Again, the micro-world in figure 1 can be used to make these ideas more concrete. Figure 4 shows an AH analysis for this micro-world (Bisantz and Vicente 1994). As shown along the top, there are three levels of decomposition for this particular example, each connected by part–whole relations (System, Sub-system, and Component). As shown along the left, there are five levels of abstraction for this example, each connected by structural means-ends links (Functional Purpose, Abstract Function, Generalized Function, Physical Function and Physical Form). Four cells in figure 4 have been identified as being useful (the variables at each level are shown in parentheses):

- Functional Purpose/System—outputs to the environment (T1, D1; T2, D2);
- Abstract Function/Sub-system—mass and energy topologies (MI1, M1, MO1, EI1, E1, EO1; MI2, M2, MO2, EI2, E2, EO2);
- Generalized Function/Component—liquid flow and heat transfer rates (FPA, FVA, FA1, FA2; FPB, FVB, FB1, FB2; FH1, FH2); and
- Physical Function/Component—component settings (PA, VA, VA1, VA2; PB, VB, VB1, VB2; VO1, VO2; HTR1, HTR2).

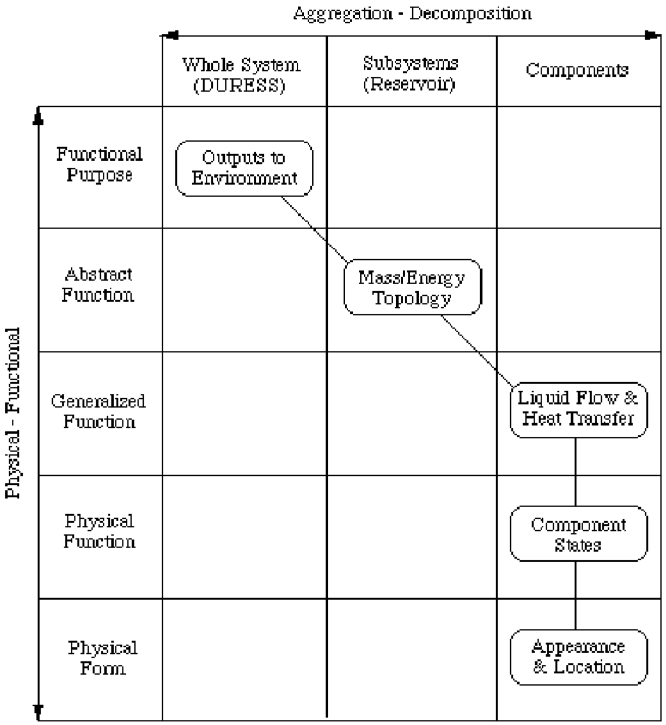

Figure 4. Representation of thermal-hydraulic process control micro-world in abstraction/decomposition space (adapted from Bisantz and Vicente 1994).

Note that this fourth cell can be considered an *action space*, because it represents the components on which participants can act. The bottom level of Physical Form was not used because the location and appearance of components are not meaningful in a micro-world simulation.

4. Quantitative measurement

The AH has been used for a wide number of purposes in ergonomics science, including protocol analysis, interface evaluation, interface design, database design, training and worker role allocation (Vicente 1999). However, before this research, it does not appear to have been used to identify objective, quantitative measures that can illuminate human performance.

4.1. *A stratified hierarchy of state spaces*

Each of the four cells in figure 4 contains a different representation of the very same work domain, each being a different frame of reference, or state space, for measuring performance. Thus, the participant-environment behaviour during any one trial in an experiment can be plotted as a trajectory over time. However, because each level of the AH defines a different state space, each trial will be revealed as a different trajectory, depending on the level of abstraction chosen for measurement.

Consequently, each frame of reference can be used to conduct a different data analysis. For example, at the Functional Purpose/System level, the micro-world in figure 1 can be described in a four-dimensional state space (not including the time dimension) defined by the four outputs: T1, D1, T2, D2 (see table 1). The behaviour of the simulation during one trial for one participant can be plotted as a trajectory in this state space. For a successfully completed startup trial, this trajectory would start at the origin of the space (because the process is initially shutdown) and would end at the area defined by the particular set-point values (and tolerances) for that trial.

If we take a block of trials for one participant, we get a series of trajectories in the same state space, one for each trial. It is then possible to calculate the multi-dimensional variance of these trajectories, using the formula shown in the appendix. This variance is a quantitative measure of consistency at this level of the AH. For example, if the path that a participant takes is exactly the same for each trial when plotted in this state space, then the variance would be zero. However, if the path that a participant takes in this state space is wildly different from trial to trial, then the variance would be quite large. Therefore, in the language of dynamical systems theory (Port and van Gelder 1995), this measure of variance in trajectories across a block of trials can be considered to be a relative measure of *coupling* to the distal properties of the work domain. If participants exhibit low variance, then they are strongly coupled to this level of the AH. If participants exhibit high variance, then they are weakly coupled to this level of the AH. While the construct of coupling is frequently used in the ergonomics science literature, it is rarely defined computationally and measured quantitatively as it is here.

The very same block of trials can be plotted as trajectories at any of the levels of the AH. Moreover, it is possible to calculate the variance of that block of trajectories at any level of the AH in a manner that is analogous to that just described for the Functional Purpose level (see the appendix). And, because each level of the AH represents a different state space for the same work domain, the same block of trials will be represented as a different set of trajectories at each level of the AH. Thus, it is possible for the same participant to exhibit high variance at one level and low variance at another. Such a pattern of results allows us to make inferences about the relative degree of coupling for a particular participant as a function of level of the AH. For example, one participant may be more strongly coupled to a higher level of the AH, suggesting that they are focusing on the functions to be satisfied. Another participant may be more strongly coupled to a lower level of the AH, suggesting that they are focusing on a particular sequence of quantitative component settings (e.g. like a detailed procedure consistently followed by rote).

In the remainder of this article, we will show that this quantitative set of measures may reveal important differences between participants, even after extensive experience.

5. Case study

5.1. *Longitudinal experiment on interface design*

The measures just defined will be illustrated with data from a longitudinal experiment investigating the impact of interface design on human performance (Christoffersen *et al.* 1996, 1997, 1998). Two interfaces were tested using a between-participants design (see Pawlak and Vicente (1996) for a detailed description of the interfaces). The P group used an interface that only presented *physical* information (i.e. the state of the components and the overall purposes to be achieved), much like the situation depicted in figure 3(*a*). Based on the earlier discussion, one might expect that participants would have to engage in lower-level control based on heuristics to do well with this interface. The P + F group used an interface that presented both *physical* and *functional* information (i.e. all levels of the AH), much like the situation depicted in figure 3(*b*). Based on the earlier discussion, one might expect that participants would have to engage in higher-level control based on feedback to do well with this interface. Unfortunately, it is difficult to test these hypotheses using traditional measures of performance.

5.2. *Previous analyses*

Only normal (i.e. non-fault) trials will be analysed here. The primary performance measure that had been used to investigate performance under normal trials was total trial completion time. Previous analyses using this measure had shown that there was no significant mean difference between interface groups, but that the P group was significantly more consistent than the P + F; P participants occasionally took twice as long as usual to complete the required tasks, even after 5.5 months of quasi-daily practice (Christoffersen *et al.* 1996). Table 2 provides a summary of the mean completion times for each participant for the first and last block of normal trials over the course of the 6-month long experiment.

These data show that participants AV and TL were the most proficient in their respective interface groups. They were clearly better than the other participants and not unlike each other, except for the difference in completion time variability. Analyses using other measures showed a similar pattern (Yu *et al.* 1997). On many measures, AV and TL seemed very alike and better than everyone else.

There was one strong exception to this pattern. When participants were asked to write down how they controlled the micro-world, AV and TL reported using qualitatively different procedures (Christoffersen *et al.* 1998). As shown in table 3, TL reported following a rote set of precise actions on components (e.g. 'set HTR2 to 3 1/3'; 'it might not make sense but it works'). However, as shown in table 4, AV reported focusing on the functions to be achieved (e.g. 'necessary input') and did not list many precise actions. If we believe these subjective data, then AV and TL were controlling the process in qualitatively different ways. TL's knowledge about the process seems to be action-based, while AV's seems to be function-based. Yet, the mean trial completion time analysis and many of the other measures that we investigated did not uncover this difference. Instead, they suggested that TL and AV were comparable and quite proficient in their control performance. Is this another case of dissociation between subjective reports and behavioural performance, or are the traditional measures failing to pick up a difference that really exists? We

Table 2. Average startup task completion time for first 22 and last 20 normal trials (from Christoffersen *et al.* 1998).

		Trials 1–22		Trials 196–217	
Group	Participant	*M*	SD	*M*	SD
P + F	AS	860.2	441.3	437.2	31.3
	AV	517.3	149.0	353.5	16.1
	IS	644.4	125.8	399.0	17.5
P	ML	660.3	261.8	390.2	48.0
	TL	493.6	80.2	357.5	20.5
	WL	624.3	170.4	437.2	97.9

Table 3. Last control recipe for TL on P interface (Christoffersen *et al.* 1998).

(1) Open valves VB2 and VB to their maximum (10) and start PB. Set HTR2 to 4.5.
(2) Set VA1 to 7, VA to maximum (10), and start PA. Set HTR1 to 10.
(3) As the temperature in Res 2 reaches to lower end of the desired range, set HTR2 to 3 1/3; then set VO2 to desired level.
(4) Set VO1 to desired level. As the temperature in Res 1 reaches the lower end of the desired temperature range, set VA1 to level 10.
(5) As the resevoirs reaches level 60, adjust VA1 to equal VO1 and VB2 to equal VO2. Modify heater settings as necessary.
Hint: HTR1 should be at same numerical setting as VA1, and HTR2 at 1/3 of VB2 (it might not make sense, but it works).

If the demand on res 2 is greater than 10:
Follow steps (1), (2) and (3) above, then:
(4) Set VA2 such that the total water input into res 2 is one level above VO2. Set VA1 such that $VA1 + VA2 = 10$. Set heater levels as defined in the hint section above.
(5) As the resevoir reaches level 60, adjust values so that $VA1 = VO1$ and $VA2 + VB2 = VO2$. Modify heater settings as required.

Table 4. Last control recipe for AV on P + F interface (experimenter notes are in brace brackets).

Turn PA & PB on
Turn VA & VA1 on (max)
Turn HTR1 on (max)
Turn VB & VB1 & VA2 on (if necessary)
{arrow underneath text, from parenthetical phrase to VA2}
Turn HTR2 on (max)
Adjust VO1 & VO2 to output
Adjust VA & VA1, VB & VB1 & VA2 to necessary input.
Adjust HTR1 & HTR2 For the inputs
Do some fine tuning. END

hypothesized that the quantitative measures based on the AH could shed some objective, quantitative light on the relative difference between TL and AV.

5.3. *Abstraction hierarchy measures*

The variability in the trajectories for each participant was calculated at each level of the AH described above, by block (see the appendix for the mathematical formulae). We will begin by discussing the results from the Functional Purpose/System level, illustrated in figure 5. These trajectories were normalized with respect to the set-point values for each trial, thereby allowing us to meaningfully compare trajectories across trials. The graphs in figure 5 show the variability in trajectories for each participant over the course of the entire experiment, as a function of 11 blocks of ~20 trials each. After the initial part of the experiment, the variances for TL and AV are about the same, and they exhibit very consistent trajectories at this level of the AH. This is not surprising, given that both participants were the most proficient in their respective groups. Thus, according to this measure, TL and AV behaved in the same fashion.

As already mentioned, the Physical Function/Component cell in figure 1 represents an action space because it presents the low-level component states. The variability of trajectories in this level was computed in the same manner except that the trajectories were not normalized with respect to the set-point values for

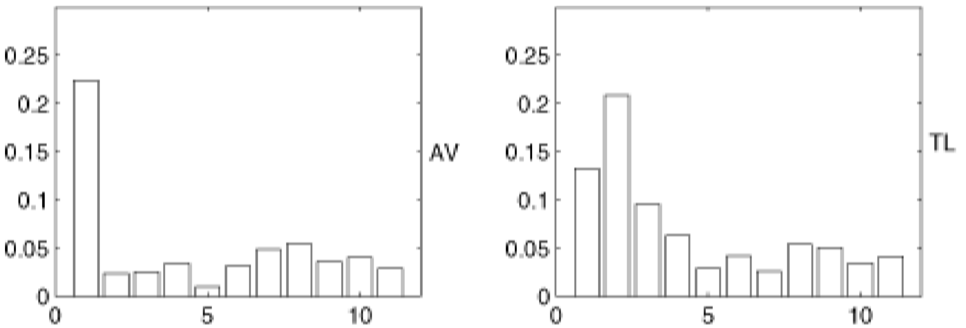

Figure 5. Variance at functional purpose/system.

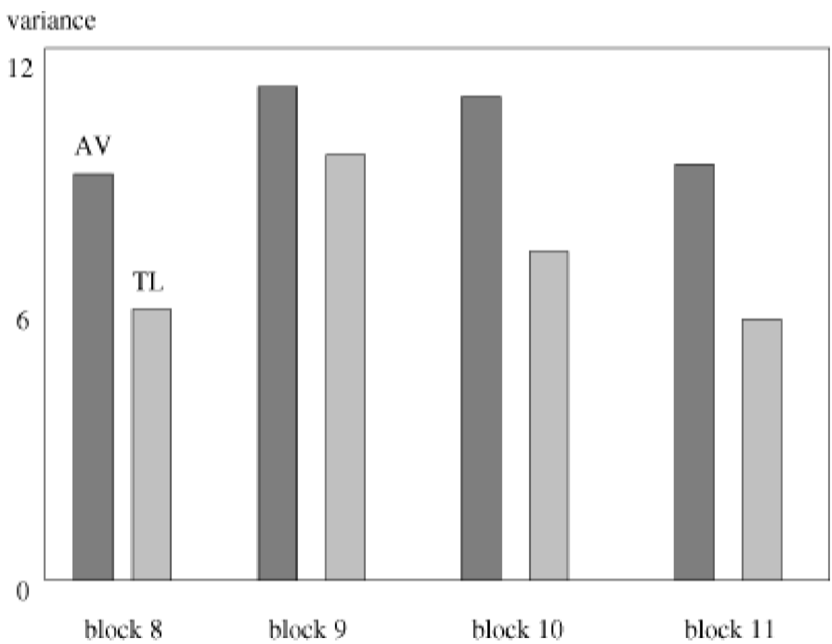

Figure 6. Comparison of variance at physical function/component at the last four blocks.

each trial. Such a normalization is not possible because there is no direct relationship between set-point values and component settings. Thus, the variability analysis at this level is based on absolute setting values (with a compensation for the fact that different components have different scale values; see the appendix). Figure 6 compares the action variability for AV and TL during the last four blocks of the experiment. These data show that TL's behaviour is consistently less variable than AV's at this level of the AH. This result is consistent with the observation that TL's behaviour is driven more by a fixed set of specific actions than AV's (compare tables 3 and 4). Thus, this finding provides support for the expectations generated by the subjective data reported by these two participants.

Figure 7 shows the results of the variability analysis at the Generalized Function/ Component level. The trajectories in this frame of reference were also not normalized with respect to the set-point values for each trial for the reasons stated above. The

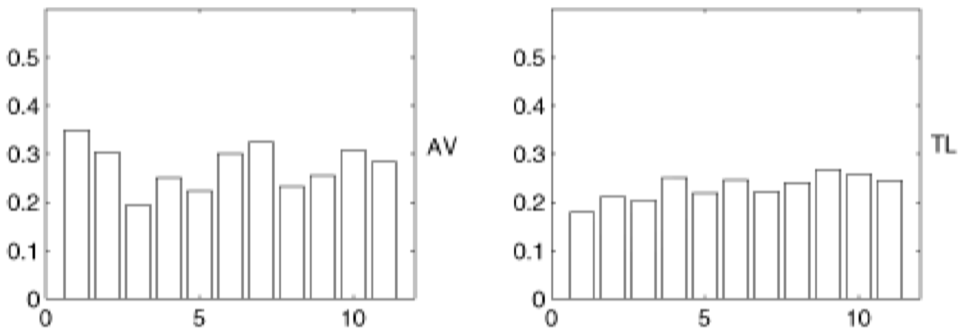

Figure 7. Variance at generalized function/component.

results are very similar to those of the previous analysis because there is a strong correlation between these two levels of the AH. By inspecting the equations describing the process dynamics, we can see that there is a direct correspondence between these two sets of variables after the transient produced by a change in component setting. In other words, if we are given the component settings we can usually uniquely derive the liquid flow rates and heat transfer rates (for normal trials). The only times during which this relationship is weakened is during the transient period after a control action. Thus, this analysis does not provide any new insights.

The final set of AH variance analyses was conducted at the Abstract Function/Sub-system level. There are two important differences between this frame of reference and the last two just described. First, the measurement is taking place at an aggregate level. We are now examining variables at the Sub-system level, which are aggregates of the variables that were examined at the level of Components (see figure 4). Secondly, measurement at this level is in terms of variables that describe the system in terms of first principles (i.e. mass and energy conservation laws). In this sense, this frame of reference is a privileged level of description. The first analysis conducted at this level was based on trajectories that were not normalized for the particular set-point values for different trials. In this case, the calculations are based on absolute data values (except for a compensation for the fact that different components have different scale values; see the appendix). The results from this analysis are presented in figure 8. It is difficult to discern any patterns in the data.

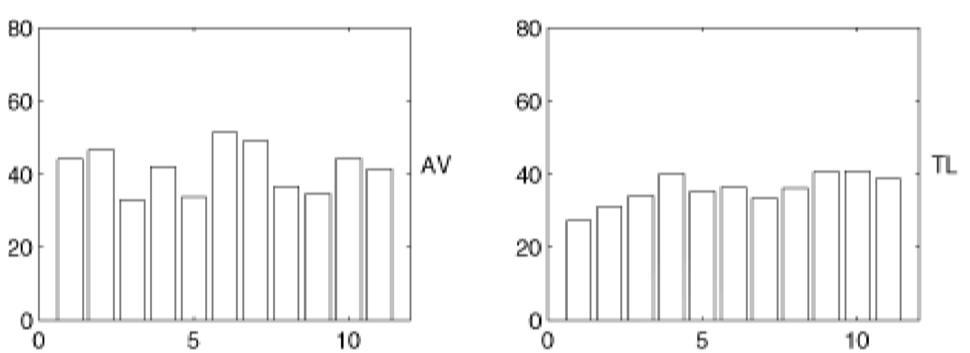

Figure 8. Variance at abstract function/sub-system (normalized by scale only).

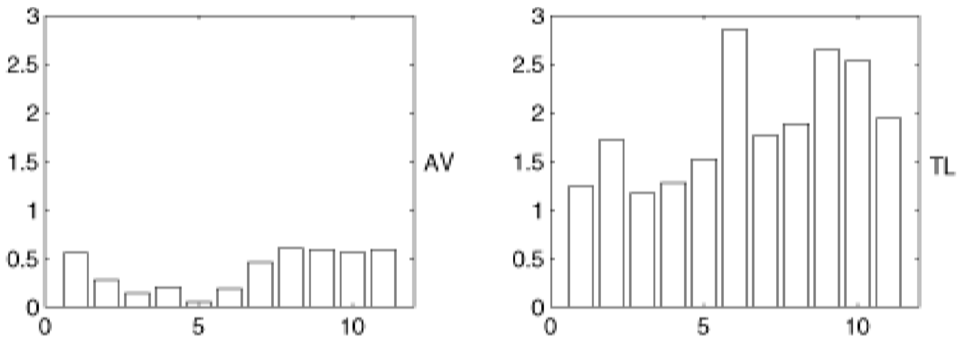

Figure 9. Variance at abstract function/sub-system (normalized by both scale and goals).

There is another way to look at these data, however. Because each trial has a different set of setpoints, we would expect there to be variance in the trajectories for this reason alone. Although the trajectory for each trial begins at the origin, the end point for each trajectory will be different for each trial as a function of the setpoint for that trial. If we assume that participants try to stabilize both volume and temperature for each reservoir, then it is possible to correct the trajectories for differences in setpoint values across trials. This is accomplished by dividing the mass input and output flowrates by the demand setpoints, and dividing the energy input and output flowrates by the product of the demand setpoints and the temperature setpoints (see the appendix). Normalizing the trajectories in this fashion eliminates any variability caused solely by differences in setpoints across trials.

The results from this second analysis are presented in figure 9. Several interesting findings emerge from this alternative way of looking at the data. The most important of all is the large difference between the variances for TL and AV. From the beginning of the experiment, but especially in the second half, the trajectory variance for AV is much lower than that for TL. This result provides objective validation of the subjective report data in tables 3 and 4. AV is thinking about and controlling the process at a high level of abstraction, focusing on the mass and energy level. Moreover, he contextualizes his control at this level based on the setpoint values for each trial. This can be observed by the noticeable difference in the data in figures 8 and 9 for AV. It is only when we compensate for differences in setpoint values that we see that, at a high level of abstraction, AV is acting in a consistent fashion across trials. In contrast, the regularities in TL's behaviour are more at the action level (figure 6), where he exhibited a lower variance than AV. Because TL's actions are relatively similar for trials with different setpoints, his behaviour is not as contextualized (or situated) as AV's. Thus, when we examine TL's data at a contextualized, functional level of abstraction, he exhibits less structure than does AV.

5.4. *Discussion*

Subjective report data had suggested that AV and TL controlled the process in qualitatively different ways (see tables 3 and 4). However, these differences had not been confirmed by many other objective measures of performance that had been used to analyse the data from this experiment (Yu *et al.* 1997). The AH-based measures proposed in this article offered unique insight by providing objective, quantitative evidence about the important differences between these two participants. Furthermore, these insights were consistent with the subjective report data. At the lowest level of the AH (i.e. the action space), TL exhibited less variability in his trajectories than AV. Because TL thought about the process in terms of specific actions on components, it makes sense that the regularities in his behaviour appear at this low level of abstraction. Conversely, at a high level of abstraction, AV exhibited less variability in his trajectories than TL. Because AV thought about the process in terms of functions, it makes sense that the regularities in his behaviour should appear at a high level of abstraction.

Perhaps most importantly of all, these findings are consistent with the theoretical rationale behind the AH, given the different interfaces used by TL and AV. AV used the P + F interface which presented him with both physical and functional information (see figure 3(*b*)), thereby providing some help in solving the inverse dynamics problem. Because he could see the state and structure of the system, he did not have to memorize a set of procedures. Instead, he could use the information in the P + F interface as an error signal to generate actions that were appropriate to the current context. Thus, there was a stronger coupling between AV's actions and the micro-world, as shown by the AH analysis at the level of first principles in figure 9. This stronger coupling also led to a larger degree of context-conditioned variability (Turvey *et al.* 1982). Because different trials had different goal setpoint values, AV's actions were more variable across trials (see figure 6).

TL used the P interface, which only displayed physical information (see figure 3(*a*)), making it more difficult to solve the inverse dynamics problem. Although the P interface provided TL with enough feedback to control the system efficiently, it does not reveal all of the interactions that govern the micro-world. As a result, TL could not rely primarily on the feedback in the interface to generate his actions. Instead, he had to acquire a rote set of detailed actions that he used as a script for each trial. Thus, TL's actions were less variable across trials because they seemed to be governed more by the steps in his procedure than by what was presently going on in the process. Consequently, TL exhibited a weaker coupling to the first principles of the micro-world (see figure 9). The regularities in his control were at the action level (see figure 6).

This theoretical interpretation of the differences between AV and TL is only possible because there is a very strong connection between the objective, empirical measures described in the appendix and the theoretical constructs of the AH framework.

6. Conclusions

This article has made a novel contribution to performance measurement in ergonomics science. A novel set of measures have been proposed based on the AH framework. As far as we know, this is the first time that the AH has been used for this purpose. In addition to being theoretically driven, these measures benefit from being objective and quantitative, thereby improving the rigour of ergonomics science. The measures were also sensitive enough to identify behavioural differences between participants in a longitudinal study of interface design. These differences had not been objectively identified by many other analyses using more traditional measures, such as task completion time.

The primary limitation of this work is that these novel measures were only applied to two participants in one experimental setting. Analogous measures can be derived for other work domains for which it is possible to develop an AH representation. Although the content of the levels of the AH will differ for various work domains, the relationship between levels will be the same. The key empirical question for future research is whether such measures will lead to important and unique insights in other contexts, as they have here. Accordingly, it is important that the AH be used as a measurement tool in diverse application domains to assess the generalizability of the approach proposed here.

Acknowledgements

This research was sponsored by a contract from the Japan Atomic Energy Research Institute. We would like to thank Dr Fumiya Tanabe (contract monitor), Greg Jamieson, Klaus Christoffersen, and Chris Hunter for their contributions.

References

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

CHRISTOFFERSEN, K., HUNTER, C. N. and VICENTE, K. J. 1996, A longitudinal study of the effects of ecological interface design on skill acquisition, *Human Factors*, **38**, 523–541.

CHRISTOFFERSEN, K., HUNTER, C. N. and VICENTE, K. J. 1997, A longitudinal study of the effects of ecological interface design on fault management performance, *International Journal of Cognitive Ergonomics*, **1**, 1–24.

CHRISTOFFERSEN, K., HUNTER, C. N. and VICENTE, K. J. 1998, A longitudinal study of the impact of ecological interface design on deep knowledge. *International Journal of Human–Computer Studies*, **48**, 729–762.

HOWIE, D. E. and VICENTE, K. J. 1998, Measures of operator performance in complex, dynamic microworlds: advancing the state of the art, *Ergonomics*, **41**, 485–500.

MORAY, N. and ROTENBERG, I. 1989, Fault management in process control: eye movements and action, *Ergonomics*, **11**, 1319–1342.

MORAY, N., LOOTSTEEN, P. and PAJAK, J. 1986, Acquisition of process control skills, *IEEE Transactions on Systems, Man, and Cybernetics*, **16**, 497–504.

PAWLAK, W. S. and VICENTE, K. J. 1996, Inducting effective operator control through ecological interface design, *International Journal of Human–Computer Studies*, **44**, 653–688.

PORT, R. F. and VAN GELDER, T. 1995, *Mind as motion: explorations in the dynamics of cognition* (Cambridge, MA: MIT Press).

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

SANDERSON, P. M., VERHAGE, A. G. and FULD, R. B. 1989, State-space and verbal protocol methods for studying the human operator in process control, *Ergonomics*, **32**, 1343–1372.

TURVEY, M. T., FITCH, H. L. and TULLER, B. 1982, The Bernstein perspective: I. The problems of degrees of freedom and context-conditioned variability, in J. A. S. Kelso (ed.), *Human motor behavior: An introduction* (Hillsdale, NJ: Erlbaum), 239–252.

VICENTE, K. J. 1991, Supporting knowledge-based behaviour through ecological interface design. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL.

VICENTE, K. J. 1999, *Cognitive Work Analysis: Toward safe, productive, and healthy computer-based work* (Mahwah, NJ: Erlbaum).

YU, X., CHOW, R., JAMIESON, G. A., KHAYAT, R., LAU, E., TORENVLIET, G. L., VICENTE, K. J. and CARTER, M. W. 1997, Research on the characteristics of long-term adaptation, JAERI Final Report, CEL 97-04, Cognitive Engineering Laboratory, University of Toronto.

Appendix

Functional purpose/system

This measure shows the consistency in subjects' performance across trials, at the level of goal variables (outputs). The state of the four goal variables, ($MO_1(t)$, $MO_2(t)$, $T_1(t)$, $T_2(t)$) can be plotted against time, creating one trajectory in five-dimensional space for each trial. The variance in these trajectories within a block of trials is then calculated.

The variance of goal variables over a block of trials is defined as follows:

(1) *Time shift*: Generally, there is a delay between the beginning of a trial and the time of the first action by the participant. The magnitude of this delay varies across participants, and within-participants across trials. For the purposes of this analysis, this idiosyncratic response delay is noise, so it should be removed. If there is a time delay, τ for the trial, then the four goal variables T_1, T_2, MO_1, MO_2 should be shifted by τ so that they all start at time 0. This leads to functions $T_1(t-\tau)$, $T_2(t-\tau)$, $MO_1(t-\tau)$, and $MO_2(t-\tau)$, respectively.

(2) *Normalization*: For each trial, the four goal variables: $MO_1(t-\tau)$, $MO_2(t-\tau)$, $T_1(t-\tau)$, $T_2(t-\tau)$ are normalized with respect to their set points, which leads to $\overline{MO}_1(t-\tau)$, $\overline{MO}_2(t-\tau)$, $\overline{T}_1(t-\tau)$, and $\overline{T}_2(t-\tau)$, respectively. Normalization allows us to compare all of the variables across trials from a common reference scale.

(3) *Linear interpolation*: The micro-world simulation only logs its state at the time of a participant action rather than at a constant sampling interval. Thus, if there is a long time between actions, then the state of the process during this time will be unknown and must be derived. To recover these data, $\overline{MO}_1(t-\tau)$, $\overline{MO}_2(t-\tau)$, $\overline{T}_1(t-\tau)$, $\overline{T}_2(t-\tau)$ are linearly interpolated every 3 seconds over the first 300 seconds of a trial (the minimum duration of a trial). This interpolation interval was chosen based on knowledge of the bandwidth of the micro-world dynamics. Thus, we get $\overline{MO}_1(t_i)$, $\overline{MO}_2(t_i)$, $\overline{T}_1(t_i)$, $\overline{T}_2(t_i)$ with $t_1 = 0$, $t_2 = 3$, $t_3 = 6, \ldots, t_{101} = 300$.

(4) The multi-dimensional, time-wise variance at each t_i is calculated by:

$$\mathrm{var}\,(t_i) = \frac{\begin{array}{r}\sum_{j=1}^{n}(\overline{MO}_1^j(t_i) - ave_{\overline{MO}_1(t_i)})^2 + (\overline{MO}_2^j(t_i) - ave_{\overline{MO}_2(t_i)})^2 \\ + (\overline{T}_1^j(t_i) - ave_{\overline{T}_1(t_i)})^2 + (\overline{T}_2^j(t_i) - ave_{\overline{T}_2(t_i)})^2\end{array}}{n-1},$$

where $\overline{MO}_1^j(t_i)$, $\overline{MO}_2^j(t_i)$, $\overline{T}_1^j(t_i)$, $\overline{T}_2^j(t_i)$ are normalized outflow rates and temperatures at time t_i of trial j within the block of sampled trials. $ave_{\overline{MO}_1(t_i)}$, $ave_{\overline{MO}_2(t_i)}$, $ave_{\overline{T}_1(t_i)}$, $ave_{\overline{T}_2(t_i)}$ are the average values of $\overline{MO}_1^j(t_i)$, $\overline{MO}_2^j(t_i)$, $\overline{T}_1^j(t_i)$, $\overline{T}_2^j(t_i)$ respectively, over the same block of trials. For example,

$$ave_{\overline{MO}_1(t_i)} = \sum_{j=1}^{n} \overline{MO}_i^j(t_i)/n.$$

(5) The multi-dimensional variance over the entire 300 s span can then be calculated as follows:

$$variance = \frac{\int_0^{300} \operatorname{var}(t)dt}{300} \approx \frac{\sum_{i=0}^{300} \operatorname{var}(3i) \times 3}{300}.$$

Abstract function/sub-system

At this level of the AH, there are 12 variables that describe the state of the microworld: MO1, EI1, EO1, M1, E1, MI1, MO2, EI2, EO2, M2, E2 and MI2 (see Bisantz and Vicente 1994). With the addition of time, they form a 13-dimensional space. Multi-variance is defined in the same way as variance at the goal level, except that the variables are normalized with respect to their maximum possible scale values. This normalization process removes any artificial, differential-weighting effects caused by heterogeneous numerical scales across variables.

There are actually two ways to calculate variance at this level of the AH, one that is context-sensitive and another that is not:

(1) The first way is by normalization with respect to the setpoint variables (D1, T1, D2 and T2), as well as scale (shown in table A1).
(2) The second method is by normalization with respect to scale only (shown in table A2).

Generalized function/component

At this level of the AH, there are 10 variables that describe the state of the process: FA, FA_1, FA_2, FB, FB_1, FB_2, FO1, FO2, HTR_1 and HTR_2. Including time, they form an 11-dimensional space. As before, the variables are normalized with respect to their maximum values before calculating the variance in trajectories.

Table A1

For reservoir 1:	For reservoir 2:
$\overline{MO}1 = MO1/D1$	$\overline{MO}2 = MO2/D2$
$\underline{EI1} = EI1/D1 \times T1 \times 2,090,000$	$\underline{EI2} = EI2/D2 \times T2 \times 2,090,000$
$\underline{EO}1 = EO1/D1 \times T1 \times 2,090,000$	$\underline{EO}2 = EO2/D2 \times T2 \times 2,090,000$
$\underline{M}1 = M1$	$\underline{M}2 = M2$
$\underline{E1} = E1/168,000,000$	$\underline{E2} = E2/168,000,000$
$MI1 = MI1/D1$	$MI2 = MI2/D2$

Table A2

For reservoir 1:	For reservoir 2:
$\overline{MO}1 = MO1$	$\overline{MO}2 = MO2$
$\underline{EI1} = EI1/2,090,000$	$\underline{EI2} = EI2/2,090,000$
$\underline{EO}1 = EO1/2,090,000$	$\underline{EO}2 = EO2/2,090,000$
$\underline{M}1 = M1$	$\underline{M}2 = M2$
$\underline{E2} = E1/168,000,000$	$\underline{E2} = E2/168,000,000$
$MI1 = MI1$	$MI2 = MI2$

Physical function/component

In the fourth level of the AH for this micro-world, there are 12 different components that participants can act on: PA, PB, VA, VA1, VA2, VB, VB1, VB2, VO1, VO2, HTR1 and HTR2. With time, these variables form a 13-dimensional action state space. Multi-variance is defined in the same way as variance at the functional purpose level, except that: (a) the variables are normalized with respect to their maximum settings; and (b) time is represented on an ordinal scale (e.g. time of first action, time of second action, etc.) rather than on an interval scale. The latter decision was made because the order of actions seemed to be more important and more meaningful than their precise timing.

About the authors

Xinyao Yu received a BSc (1990) from Harbin Institute of Technology, and PhD (1995) from Beijing University of Aeronautics and Astronautics in China, both majoring in control engineering. He was a research fellow at the International Institute for Software Technology, United Nations University from 1993–1994, and a postdoctoral research fellow with the System Control Group and the Cognitive Engineering Lab at the University of Toronto from 1996–1998. His research interests include advanced control theory, safety critical embedded system specification and development methods, and human factors theory. Xinyao joined Ontario Power Generation Inc. in 1998, where he is an engineer working on the design of safety critical control software, instrumentation and control, and human machine interface for nuclear power plants. He is author of more than 20 scientific papers. Xinyao has been invited to present his work in several universities including Oxford University, Technical University of Denmark, City Polytech of Hong Kong, and University of Toronto. He is still active in academe while working in the power generation industry.

Elfreda Lau is a recent graduate of the University of Toronto, earning a Bachelor of Applied Science degree in Industrial Engineering. During her studies, she spent 16 months on a professional experience year internship working as a software engineering consultant. Elfreda is currently working as an Associate Consultant for Bain & Company, Inc., a global management consultancy firm.

Kim J. Vicente received his BASc (1985) in Industrial Engineering from the University of Toronto, his MS (1987) in Industrial Engineering and Operations Research from the Virginia Polytechnic Institute and State University, and his PhD (1991) in Mechanical Engineering from the University of Illinois at Urbana-Champaign. During 1987–1988, he spent 1 year as a visiting scientist in the Section for Informatics and Cognitive Science of the Risø National Laboratory in Roskilde, Denmark. During 1991–1992, he was on the faculty of the School of Industrial and Systems Engineering at the Georgia Institute of Technology. Since 1998, he has been professor of Mechanical & Industrial Engineering and of Biomaterials & Biomedical Engineering at the University of Toronto, and director of the Cognitive Engineering Laboratory there. He is also an adjunct professor of psychology at Miami University, Ohio, and a registered Professional Engineer in the province of Ontario. Currently, Kim serves on the editorial boards of *Human Factors* and *Theoretical Issues in Ergonomics Science*, and on the Committee for Human Factors of the US National Research Council/National Academy of Sciences. He is also a Senior Fellow of Massey College. Kim is the recipient of several research awards, including the Premier's Research Excellence Award, valued at \$100 000. He has authored or co-authored over 70 journal articles, and over 75 refereed conference papers. He is the author of *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-based Work* (Lawrence Erlbaum Associates).

Michael Carter is a Professor in the Department of Mechanical and Industrial Engineering at the University of Toronto. He is cross-appointed to the Department of Computer Science, the Institute of Biomedical Engineering and the Faculty of Nursing at the University of Toronto. He received his PhD in Combinatorics and Optimization from the University of Waterloo. He is a member of the editorial board for the *Journal of Scheduling* and the *Journal of Health Care Management Science*, and an Associate Editor for the journal INFOR of CORS. He is co-author of the book *Operations Research: a Practical Introduction*, July 2000, CRC Press. He was the winner of the Annual Practice Prize for Operational Research from the Canadian Operational Research Society three times (1988, 1992 and 1996). In 2000, he received the CORS Award of Merit for lifetime contributions to Canadian Operational Research. He also received an 'Excellence in Teaching' Award from the University of Toronto Student Administrative Council. His research has been primarily in the area of practical scheduling problems. Michael has graduated eight PhD students and 27 Masters. His student timetabling algorithms are currently being used at Carleton University (Ottawa), the London School of Economics, the University of Otago (New Zealand), the University of Limerick (Ireland), l'École des Hautes Études Commerciales de Montréal and the University of Western Ontario (London, Ontario). His current research focus is in the area of Healthcare productivity and effectiveness with a variety of projects at several hospitals, home care and mental health institutions.