Advancing Performance Measurement in Cognitive Engineering: The Abstraction Hierarchy as a Framework for Dynamical Systems Analysis

Xinyao Yu, Elfreda Lau, Kim J. Vicente and Michael W. Carter
Proceedings of the Human Factors and Ergonomics Society Annual Meeting 1998 42: 359
DOI: 10.1177/154193129804200338

Advancing Performance Measurement in Cognitive Engineering: The Abstraction Hierarchy as a Framework for Dynamical Systems Analysis

Xinyao Yu, Elfreda Lau, Kim J. Vicente, and Michael W. Carter
Cognitive Engineering Laboratory
Department of Mechanical and Industrial Engineering
University of Toronto
Toronto, Canada

This paper describes a novel approach to performance measurement that is based on the abstraction hierarchy (AH) and dynamical systems theory (DST). Each level in the AH provides a systematic way of identifying a state space that can be used to conduct complementary DST analyses. This approach was applied to data from a longitudinal experiment that measured subjects' performance in interacting with a process control microworld simulation on a quasi-daily basis over a period of six months. The variability in trajectories at each level of the AH was examined over successive blocks of trials. The analyses at different levels revealed complementary insights into subjects' behavior. Collectively, the results provided objective, quantitative evidence that highly experienced and very proficient subjects were actually performing the task using very different strategies. Thus, integration of the AH and DST provides a novel measurement approach that can reveal unique and important insights into performance.

INTRODUCTION

Cognitive engineers have proposed advanced interface design principles to facilitate the job of operators of complex sociotechnical systems (e.g., Bennett & Flach, 1992). But as ideas for new interfaces have developed, our ability to evaluate them empirically has not kept pace. Thus, there is a pressing need for novel measures that distinguish, not just between more and less proficient subjects, but also between equally proficient subjects who are performing tasks in qualitatively different ways (Sanderson, Verhage, & Fuld, 1989). In this paper, we describe how the abstraction hierarchy (AH; Rasmussen, 1985) and dynamical systems theory (DST; e.g., Port & van Gelder, 1995) can be combined for this purpose.

Dynamical Systems Theory

DST has been used for decades by scientists and engineers from diverse disciplines to model dynamic systems. In DST, the behavior of a system over time is described using a *state variable*. Complex systems are usually modeled using several state variables, each representing a different attribute of the system. Collectively, the state variables of a system define a multidimensional *state space*, with each state variable defining one dimension of that space. Any sequence of system behavior can be mapped onto the state space as a sequence of states. This line in the state space is referred to as a *trajectory*. Although DST has its roots in the physical sciences, it is being increasingly used in cognitive science (Port & van Gelder, 1995). In the remainder of this paper, we will show that DST can be very relevant to cognitive engineering as well.

Abstraction Hierarchy

The AH is a multileveled representation format for describing complex work domains (Rasmussen, 1985). Each level of the AH is a model of the same work domain, with each level relying on a different set of attributes or `language'. Higher levels of abstraction represent the work domain in terms of purpose and function, whereas lower levels represent the work domain in terms of physical implementation. 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.

Figure 1 provides an example of how the abstraction and decomposition hierarchies can be used to develop multiple representations of a work domain (Bisantz & Vicente, 1994). The system being represented is DURESS (DUal REservoir System Simulation) II, a process control microworld that was used as a testbed for this and previous research. As shown along the top of Figure 1, there are three levels of decomposition for this particular example, each connected by part-whole relations (System, Subsystem, and Component). As shown along the left side of Figure 1, there are five levels of abstraction for this example, each connected by means-ends links (Functional Purpose, Abstract Function, Generalized Function, Physical Function, and Physical Form).

The abstraction and decomposition hierarchies, while conceptually orthogonal, are coupled in practice (Rasmussen, 1985). At higher levels of abstraction, operators tend to think of the work domain at a coarse level of resolution, whereas at lower levels of abstraction, they tend to adopt a detailed level of resolution. Therefore, certain cells in the space are not very meaningful (e.g., Functional Purpose / Component).

Aggregation - Decomposition

		Whole System (DURESS)	Subsystem (Reservoir)	Component
Functional	Functional Purpose	Outputs to Environment		
	Abstract Function		Mass/Energy Topology	
	Generalized Function			Liquid Flow & Heat Transfer
Physical	Physical Function			Component States
	Physical Form			Appearance & Location

Figure 1: Representation of DURESS II in abstraction / decomposition space (from Bisantz & Vicente, 1994).

In the case of DURESS II, four cells in Figure 1 have been identified as being useful for the present purposes (Bisantz & Vicente, 1994):

- Functional Purpose / System - outputs to the environment
- Abstract Function / Subsystem - mass and energy topologies
- Generalized Function / Component - liquid flow and heat transfer rates
- Physical Function / Component - component settings

Note that this fourth cell can be considered an *action space* in that subjects can only act directly on the components in DURESS II. The bottom level of Physical Form will not be used here because the location and appearance of components are not meaningful in a microworld simulation.

Integration

Each of the four cells just described contains a different representation of the very same work domain, each being a different abstraction of process behaviors. This insight allows us to use the AH to conduct a DST analysis. Each level of the AH provides a different frame of reference, or state space, for measuring performance. Thus, the behavior during any one trial can be plotted as a trajectory over time. But because each level defines a different state space, each trial will be revealed as a different trajectory, depending on the level of abstraction that we chose for measurement.

Consequently, each frame of reference can be used to conduct a different DST analysis. For example, at the Functional Purpose/System level, DURESS II can be described in a four dimensional state space (in addition to the time dimension) defined by the four outputs: two temperature goals and two output flowrates. The behavior of the work domain during one trial is a function of time and can be geometrically plotted as a trajectory in the five dimensional space. For a successfully completed startup trial, this trajectory would start at the origin of the space (because the system is initially shutdown) and would end at the small area defined by the particular goal values (and tolerances) for that trial. In the remainder of this paper, we will show how the various levels of the AH can be used to define multiple, complementary frames of reference that reveal important differences between subjects, even after extensive experience.

METHOD

The research was conducted with DURESS II, a thermal-hydraulic process simulation. DURESS II was designed to be representative of industrial process control systems, thereby increasing the generalizability of research results to operational settings. Two interfaces for DURESS II were designed. The first only presented physical (P) information about the process (i.e., the settings of all of the components and the goal variables), which represents the state-of-art in traditional interface design methods. In contrast, the second interface presented both physical and higher-order functional (P+F) information about the process (i.e., flows, heat transfer rates, mass and energy balances). The P+F interface was designed based on the analyses of AH of DURESS II (for a detailed description, see Pawlak & Vicente, 1996).

Three subjects were assigned to each interface group. They were asked to control DURESS II until the goals (two temperature goals and two water outflow rates) were achieved. Subjects controlled the process for approximately one hour a day on a quasi-daily basis for six months. The control tasks included startup, tuning, and shutdown. Different fault modes were introduced in some of the trials (unknown to the subjects), requiring the subjects to detect, diagnose, and compensate for the abnormality. The total number of trials for each subject was 244.

A time-stamped, log file was created during each trial, recording the subjects' actions, and the values of all of the process variables. At several points during the experiment, subjects were asked to write down a "control recipe" describing how they controlled the system during a startup task (see Christoffersen et al., 1994).

RESULTS

Due to space limitations, we will present here only the results for the startup tasks of normal trials for the best subject in each interface group (AV in the P+F group, and TL in the P group). Previous analysis of the control recipes written by the subjects suggested that the best subject in the P group, TL, had followed a rote set of precise actions on components, while the best subject in the P+F group, AV, had focused on the functions to be achieved. This qualitative result seemed to show that TL's knowledge about the process was action-based, while AV's was function-based.

Based on these self-reports, we made the following predictions. At the lowest level of abstraction (i.e., the action space), TL should exhibit a lower level of variability in his trajectories than AV. Because TL thinks about the process in terms of specific actions on components, the regularities in his behavior should appear at this low level of abstraction. Conversely, at a high level of abstraction, AV should exhibit a lower level of variability in his trajectories than TL. Because AV thinks about the process in terms of functions, the regularities in his behavior should appear at a high level of abstraction. There is not enough information available to predict which of the three higher-level frames of reference AV is focusing on. Nevertheless, we would expect an inversion of results when we adopt a different frame of reference.

The variability in the trajectories for each subject was calculated at each level of the AH described above, by block (see Yu et al., 1997 for the mathematical formulae). We will begin by discussing the results from the Functional Purpose/System level, illustrated in Figure 2. These trajectories were normalized with respect to the setpoint values for each trial, thereby allowing us to meaningfully compare trajectories across trials. The graphs in Figure 2 show the variability in trajectories for each subject over the course of the entire experiment, as a function of 11 blocks of approximately 20 trials each. After the initial part of the experiment, the variance for TL and AV are at the same level, and they exhibit very consistent trajectories at the goal space level. This is not surprising, given that both subjects were the most proficient in their respective groups. Thus, according to this measure, TL and AV behaved in the same fashion.

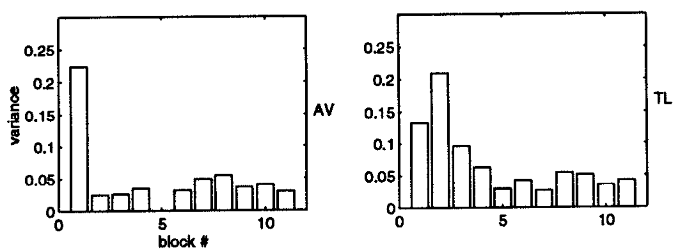

Figure 2: Variance at Functional Purpose/ System.

As already mentioned, the Physical Function/Component cell in Fig. 1 represents an action space because subjects could only act on the process by directly changing the state of individual components. The variability of trajectories in the action space level was computed in the same manner except that the trajectories were not normalized with respect to the goal values for each trial. Such a normalization is not possible because there is no direct relationship between goal 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 Yu et al., 1997). Figure 3 compares the action variability for AV and TL during the last four blocks of the experiment. These data clearly show that TL's behavior is consistently less variable than AV's. This result is consistent with the observation that TL's behavior is driven more by a fixed set of specific actions than AV's. Thus, this finding provides support for the hypotheses generated by the control recipe data for these two subjects.

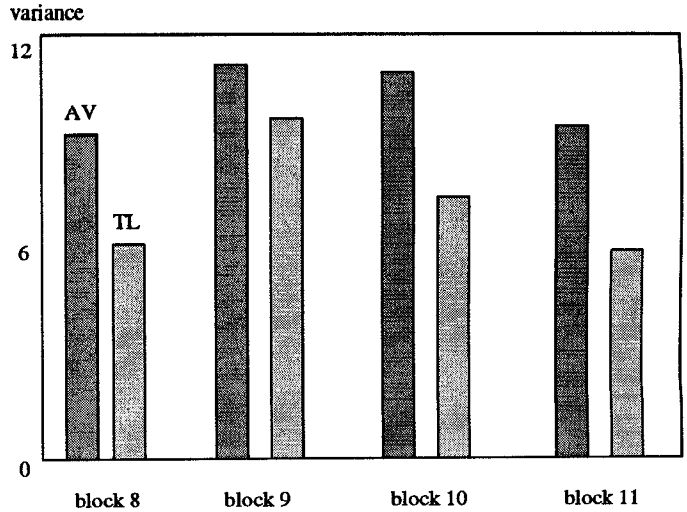

Figure 3: Comparison of variance at Physical Function / Component at the last four blocks.

Figure 4 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 goal values for each trial for the reasons stated above. The results at this level are very similar to those from the action space level because there is a strong correlation between the two levels. 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 short period after an action. Thus, this analysis does not provide any new insights.

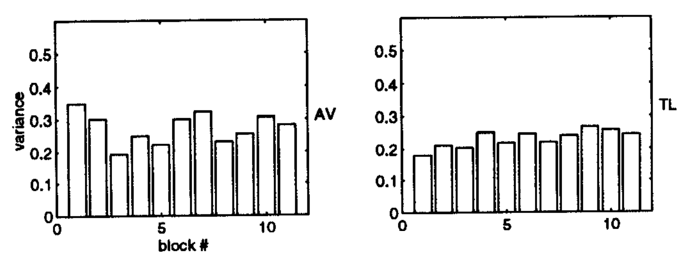

Figure 4: Variance at Generalized Function / Component.

The final set of AH variance analyses was conducted at the Abstract Function/Subsystem 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

Subsystem level, which are aggregates of the variables that we examined at the level of Components (see Figure 1). Second, 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 goal 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 Yu et al., 1997). The results from this analysis are presented in Figure 5. It is difficult to discern any patterns in the data.

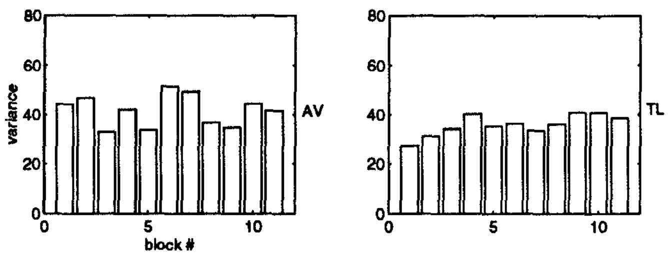

Figure 5: Variance at Abstract Function / Subsystem (normalized by scale only).

There is another way to look at these data, however. Because each trial has a different set of goal 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 different trials as a function of the setpoints for those trials. If we assume that subjects 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. Normalizing the trajectories in this fashion eliminates any variability caused solely by differences in goal setpoints across trials.

The results from this second, more informative analysis are presented in Figure 6. 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 strong result is consistent with the predictions we made from the control recipe data. 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 5 and 6 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 behavior are more at the action level (Figure 3), where he exhibited a lower variance than AV. Because TL's actions are relatively similar for trials with different setpoints, his behavior 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.

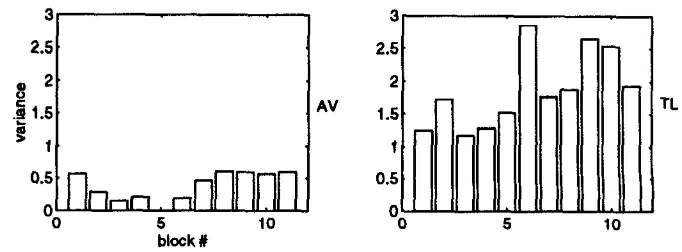

Figure 6: Variance at Abstract Function / Subsystem (normalized by both scale and goals).

DISCUSSION

This paper has shown that it is possible to integrate the AH and DST to develop novel and informative measures of performance. DST analyses of variability in trajectories were conducted at the state spaces defined by different levels of the AH. Two important insights emerged from the results, one of theoretical interest and the other of methodological interest.

The most important theoretical insight stems from the behavioral evidence showing that the best subjects in each interface group adopted very different ways of controlling DURESS II. Previous analyses had shown that both TL and AV attained a very proficient level of performance that was superior to that obtained by the other subjects in their respective groups. However, the DST analyses presented here show that TL and AV were using qualitatively different strategies to control the process. TL adopted a rote set of procedures that allowed him to control the system efficiently, at least under normal conditions, but without any deep understanding of the system. Thus, the regularities in his behavior across trials can be found primarily at the level of control actions (see Figure 3). In contrast, AV focused on the functions that had to be satisfied rather than the specific actions that were required to achieve those goals. Thus, the regularities in AV's behavior could be best revealed at the level of first principles (see Figure 6). The AH analyses show that AV controlled the system by focusing on higher-order functional variables and used whatever actions were required to satisfy the current goals. Because he was not following a script, AV exhibited a greater level of variability in his actions than did TL. However, this variability was not chaotic. Instead, it was situated to the current context.

These differences can be interpreted in terms of the different interfaces that TL and AV used to control DURESS II. AV used the P+F interface which presented him with both physical and functional information. 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 DURESS II, as shown by the AH analysis at the level of first principles in Figure 6. 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 3).

TL used the P interface which only displayed physical information. Although it provided him with enough feedback to control the system efficiently, the P interface does not reveal all of the interactions that govern DURESS II. 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 DURESS II (see Figure 6). The regularities in his control were at the action level (see Figure 3).

The methodological novelty generated by these analyses is that the AH and DST (Port & van Gelder, 1995) can be profitably combined to develop novel and informative measurement tools. The different levels in the AH can be used to define frames of references for the measurement of operator actions, or process behavior, over time. As far as we know, this is the first time that the AH has been used for this purpose. The AH provides a relatively principled basis for identifying frames of reference for dynamical systems analysis. This suggests that it may have applicability in different domains beyond DURESS II. Although the content of the levels of the AH will differ for various work domains, the relationship between levels will be the same. Thus, the combination of the dynamic systems approach and the AH as a measurement tool is a potentially widely generalizable technique. If the results obtained here are any indication, then this tool should allow researchers to obtain important and unique insights into their data.

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

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. Human Factors, 34, 513-533.

Bisantz, A. M., & Vicente, K. J. (1994). Making the abstraction hierarchy concrete. International Journal of Human-Computer Studies}, 40, 83-117.

Christoffersen, K., Hunter, C. N., & Vicente, K. J. (1994). Research on factors influencing human cognitive behavior (I) (CEL 94-05). Toronto: University of Toronto, Cognitive Engineering Laboratory.

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

Port, R. F., & 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 Trans. on Systems, Man, and Cybernetics, 15, 234-243.

Sanderson, P. M., Verhage, A. G., & 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., & 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 (pp. 239-252). Hillsdale, NJ: Erlbaum.

Yu, X., Khan, F. S., Lau, E., Vicente, K. J., & Carter, M. W. (1997). ADAPT user's manual: A data analysis tool for human performance evaluation in dynamic systems (CEL 97-03). Toronto: University of Toronto, Cognitive Engineering Laboratory.