

Mental Workload Dynamics in Adaptive Interface Design

PETER A. HANCOCK, ASSOCIATE, IEEE, AND MARK H. CHIGNELL, PENDING MEMBER, IEEE

Abstract—In examining the role of time in mental workload, a different perspective is presented from which to view the problem of assessment. Mental workload is plotted in three dimensions, whose axes represent effective time for action, perceived distance from desired goal state, and level of effort required to achieve the time-constrained goal. This representation allows the generation of *isodynamic workload contours* that incorporate the factors of operator skill and equifinality of effort. An adaptive interface for dynamic task reallocation is described that employs this form of assessment to reconcile the joint aims of stable operator loading and acceptable primary task performance by the total system.

I. INTRODUCTION

THE GROWTH in systems complexity has served to generate an increasing interest in the nature and development of the locus of contact between operator and machine: the interface. The present necessity to reconcile progressively larger fluctuations in time-constrained task demands drives the search for innovative solutions to interface problems. Adaptivity represents one avenue for improving the efficiency of human-machine interaction [1], [2]. Adaptivity in the interface may assume a number of different forms and their parallel implementation need be neither antagonistic nor operationally exclusive [3]. The present work focuses upon one such form of adaptivity which is founded upon the evaluation of operator capability. This assessment is achieved through a dynamic approach to mental workload as realized in the model presented below.

In the following sections, a brief overview of differing forms of adaptivity precedes the introduction of a new framework for mental workload assessment. This differs from the traditional forms of assessment that have been summarized elsewhere [4]–[7]. The present approach is based on the premise that task-induced mental workload results from the combination of perceived distance to a desired goal state and the effective time horizon within which to reach such a goal. It is this information concerning operator loading status that is integrated with models

of the task and the system to derive a template for the form of adaptive interface that we propose. The adaptive interface acts primarily as an intelligent intermediary [8], [9] that dynamically allocates the reallocated tasks and task components to either system or operator while preserving the dual aims of acceptable primary task performance by the system and stable levels of workload imposed upon the system operator. The advantages of this approach, together with current technical limitations to its full implementation, compose the summary section of the paper.

II. ADAPTIVE INTERFACES

There are a number of ways in which an interface can adapt to the needs of the task, the machine, and the human. One simple way is for the interface to have modes of operation that adapt to the ability or skill level of the operator [10], [11]. Computer software applications often have different interface characteristics that cater to the needs of experts and novices. This also occurs in other operational realms. For instance, a student driver may experience a modified interface where steering or braking can be assumed by the instructor when needed. Interfaces may also adapt to the user by changing conversational style [12]. Some aspects of style will apply to wording and the degree of authority or subservience implied by machine dialogue, while others relate to issues such as the way in which the machine gives advice to the user or requests additional information from them [13].

A third sense in which interfaces adapt is through the use of a format in which the task presented in the interface and the operations available to the human are immediate and obvious. This idea of immediacy, as typified by “what you see is what you get” approaches to software design, is related to the concept of cognitive compatibility [14]. The emphasis on cognitive compatibility stems from the realization that human-machine interaction has evolved from a primary focus on manual activity to one which requires a broad range of perceptual, cognitive, and motor abilities. This necessitates the development of cognitive interfaces [15] that supplement the physical interface with systems via devices. Research in this area continues in the use of pointing devices, speech recognition systems, screen designs, character fonts, etc. The technology of the physical interface continues to evolve. For example, a fairly stable set of features is emerging based on the original Xerox Star

Manuscript received June 2, 1987; revised October 21, 1987 and December 21, 1987. This work was supported by the National Aeronautics and Space Administration under Grant NCC 2-379.

P. A. Hancock is with the Department of Safety Science and the Department of Human Factors, Institute of Safety and Systems Management, University of Southern California, Los Angeles, CA 90089.

M. H. Chignell is with the Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089. IEEE Log Number 8821874.

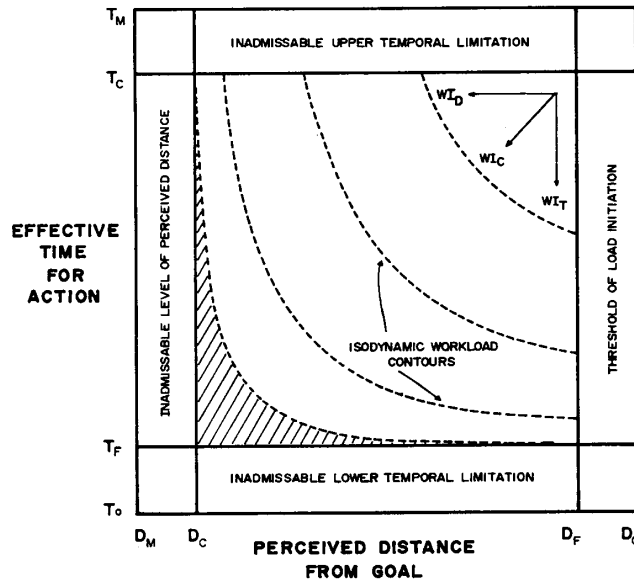


Fig. 1. Mental workload expressed as function of perceived distance from goal state and effective time for action. Salient points along distance axis D_0 , D_F , D_C , D_M , and time axis T_0 , T_F , T_C , and T_M , are described in text. Within tolerable workload space, bounded by D_F , D_C , T_F , and T_C , isodynamic workload contours (dotted lines) describe points of equal loading. Workload increases at upper right of tolerable workload space, show constant increment in load due to increase in perceived distance (W_{ID}), decrease in effective time (W_{IT}), or both in combination (W_{IC}). Hatched area at lower left indicates region of intolerable workload that results from combination of time and goal distance that makes task completion impossible.

machine [16]. In contrast, the cognitive interface remains in a state of flux. A number of desirable features of the cognitive interface have been identified, including the use of object metaphors, direct engagement [17], [18], and the selection of appropriate mental models [19]. Ideally, these interfaces remove the need for adaptivity by unifying the intentions of the user with the operations of the machine directly.

The three forms of adaptive interface described are all important and have received individual attention. This paper focuses on a fourth aspect of adaptivity, directed at the problem of task loading. This form of adaptivity supervises the match between the demands of the task and the capacity of the operator, which is reflected in level of mental workload [20].

The construction of load-mediating adaptive interfaces requires detailed knowledge of the task and its component elements. Models of cognition and human performance emphasize perceptual, cognitive, and motor processes. Adapting these models to specific tasks can be problematic. One task that has received considerable attention in this regard has been word processing. Theoretical treatments of the word processing task have been provided by various investigators [21], [22]. These task-oriented models emphasize the goals of the task and the operations that the human employs to achieve such goals. Related methods of task analysis use logic programming to develop declarative representations of tasks which then provide both a model of the task and a machine-reasoning system for augmenting the task [23].

Complex human-machine systems must perform effectively in the face of tasks that vary in structure and demand and operators who are subject to periodic oscillation in response capability. Such activity occurs against the background of perturbing and occasionally hostile environmental conditions. The human-machine interface is a critical element in the reconciliation of system performance and these operational constraints. It is at the interface that information, communication, and control functions are integrated with the intent to satisfy the diverse demands imposed upon the total system. One prime example of this integration is illustrated in the flight environment. An aircraft has to cover a number of different task phases beginning with takeoff, continuing through cruise, and terminating with final approach and landing, each of which may be performed by operators of differing skill level against a background of continual threat. Each phase of flight presents a different set of task requirements. However, each phase must also be handled by a common interface. Cockpits represent one solution to the problem of systems that are designed to adapt to a variable set of tasks and operational environments. A key principle in flightdeck display format is separation of function between instruments so that the pilot can associate particular operations with corresponding instruments. This mapping between instruments and tasks is further reinforced by training the pilot to focus on different instruments for different phases of flight. As with most existing systems, adaptivity here in the form of differing scanning strategies is initiated by the operator and the interface itself remains a static

display representation. In contrast to traditional instrument displays, new technologies have enabled the development of adaptive or context-sensitive instrumentation where the nature of certain critical displays may change between different phases of the flight; i.e., where the adaptivity is initiated by the interface, rather than by the pilot. The development of context-sensitive interfaces poses many challenges for designers as they seek to integrate new machine capabilities with the needs of those who operate complex systems [24].

III. TIME AND MENTAL WORKLOAD

Successful performance depends upon the reconciliation of the complexity and difficulty of the imposed task with the time within which the goal must be achieved. This simultaneous focus upon the characteristics of the task and the role of time in performance completion provides a different perspective from which to view the problem of mental workload assessment.

Fig. 1 illustrates a view of mental workload that includes both goal achievement and temporal constraint. The x axis of Fig. 1 represents the perceived distance between the operator's current state and the desired goal state. There are four salient points along this axis. The first of these, D_0 , represents the goal state itself. It presents no effective load with the perceived distance being zero. In the large majority of practical pursuits the operator has multileveled time-varying goals. This makes residence in the goal state a transient event, and when multiple task-related goals are involved in performance, an extremely rare occurrence. The second point along this axis, labeled D_F , represents a discontinuity. As the perceived distance from the desired goal state increases, mental workload grows accordingly. Increasing the distance from the goal brings the operator to the third salient point along the axis D_C . D_C represents the perceived maximal distance threshold specific to an individual operator. This is the level that individual operators are able to reconcile with their own maximal capacity. Transgression by an individual across this threshold of maximally acceptable perceived personal distance D_C results in a failure of effective behavioral action. Task-related mental workload returns to essentially a zero level. Therefore, one contingency of mental workload is both the *desire and belief of the operator that the task at hand is amenable to physical and/or cognitive accomplishment, within the time frame established for successful problem resolution*. Reneging on a task may have little penalty in some circumstances, while in others may have serious or even fatal consequences [25]. In some cases, individuals in the latter conditions continue to expend task-related effort, despite an irresolvable demand. This behavior is still considered goal-directed but not primarily toward the resolution of the operational task. The final point along the axis labeled D_M represents the level of perceived distance that may be reconciled by the expert or highest-level performers with respect to the specific task at hand. In the diagram, the distance between D_M and D_C is foreshortened

as it is simple for us to conceive of tasks, either singly or in combination, that even after extended practice exceed our individual best performance. Clearly, as the task-related skill of the individual increases, so the distance between D_C and D_M decreases. The distance D_C might be represented by a discrete state change analogous to a point of singularity in catastrophe theory [26]. Alternatively, failure may occur as progressive degradation in efficiency. One of the factors that apparently differentiates between these two alternative failure modes is the task-related competence of the operator [27].

On the y axis is the effective time for action. The origin of the axis T_0 is the immediate or *specious* present. This represents no effective time for action and is the temporal zero point. In a manner corresponding to the irreconcilable distance to the goal state D_M , the lower temporal limit represents a state of no load, as the operator has insufficient time available for any task-related response. The next threshold along the time axis T_F is the effective floor for operator response time. Such minimal times are well-established for most human sensory and motor processes [28] and have been used to construct models of operator performance with some degree of fidelity, e.g., the model human processor [21]. With technological refinement of current systems, this floor function or cycle time for the machine is some orders of magnitude lower than that for the human operator. In an adaptive human-machine complex, this discrepancy between human and machine cycle time creates a potential problem that requires incorporation of a more detailed understanding of the operator's temporal behavior than is currently available [29]. The effective upper limit of the temporal axis T_C , provides a horizon that is defined as coherent action. Coherent action is taken to represent *linked events designed to achieve a common goal*. An example might be a flight mission, which with its clear temporal boundaries contains primary actions directed toward achieving predetermined mission aims. Finally, there is a conceptual maximal time T_M for any activity which links a series of task-related activities. This does not connote mental workload attributed to a single common goal but rather represents the cognitive demands of the diverse sources of challenge that emerge over a longer time span. The iterations upon the temporal axis are contingent upon both endogenous (operator specific) temporal information, and temporal invariants intrinsic to the task and/or environment.

The area of admissible workload level is therefore defined by the boundaries of the thresholds D_F , D_C , T_F , and T_C , in Fig. 1. Within this space is a region of inadmissible mental workload which results from an excess of distance in combination with the brief effective time available for action. This reflects the failure of operators when a task is perceived to be within their range of capability, but some deadline restricts successful resolution. Initially, and for the sake of simplicity, an arbitrary equivalence is assumed for the ranges represented on each axis. If reduction of perceived distance per unit of effective time is considered constant, then the contours appear as in the central por-

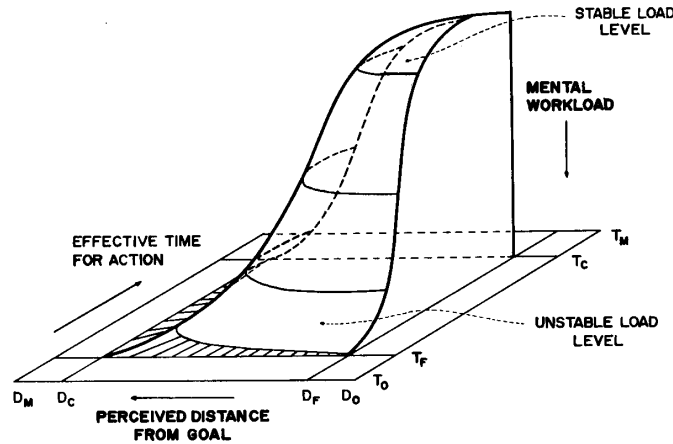


Fig. 2. Base axes are as described in Fig. 1. Vertical axis represents mental workload level that decreases as function of height. Consequently, regions of stable load are those described at upper right of three-dimensional figure. Unstable loads are those represented at lower left. Progressive increase in load beyond these latter levels results in irreconcilable demand, as given by hashed area, and thresholds T_F , D_C , T_C , and D_F . Effort is force that opposes load induced through increasing task constraints and acts to return operator to stable workload regions. Nonlinearities in vertical axis imply differential effects for equal effort at various states of loading.

tion of Fig. 1. These lines represent a family of hyperbolic curves designated as *isodynamic workload contours*. These contours are viewed as possessing equivalence in that load resolution at any location along any one contour requires a constant *effort* on behalf of the performer to retain a stable level of workload as represented by the upper right portion of the workload space. In this latter area it may be seen that equal workload increases (WI), may be generated by increasing the perceived distance from the goal (workload increase due to distance; WI_D), by decreasing the effective time for action (workload increase due to time; WI_T) or by the two attributes in combination (WI_C). The arrows shown on Fig. 1 give equal increases in load when the axes are scaled to an arbitrary equivalence as described.

In terms of the present concepts, mental workload represents the product of the task *skill* of the operator (which parameterizes the time/distance axes) and the *rate* at which reconciliation occurs. An immediate representation of this is given if the contours are developed in a third dimension giving a surface that ascends toward the top right of the diagram (see Fig. 2). The perceived demands of the task and actual constraints on performance drive the individual toward the lower left region. The *effort* of the performer opposes this tendency and acts to return the individual toward the minimal or stable load region at the upper right of the diagram. While it appears parsimonious to suggest that an operator has superior control over the perceived distance parameter, i.e., that increasing task-related effort brings the goal effectively closer, control over effective time (which is more closely linked to external temporal constraints) may be greater than generally recognized. This picture indicates that, for a given task, less mental workload is experienced by more competent operators than by their novice counterparts. Further, the range available to the experienced operator before some

threshold transition occurs is correspondingly greater than for inexperienced individuals.

There are a number of additional considerations to this view. For example, the notion of *automatic* processing [30] may explain why the perceived *distance* for a specific task is less for the experienced versus the novice performer. The experts possess the ability to extract task-related consistent *matrices of information* from the environmental display at little attentional cost. This allows available capacity to be directed to additional sources of information and extends the full range of operation. With respect to a particular task, a competent operator carries a representation not only of the difficulty but also the viable temporal duration. Indeed, the learning element of this process is suggested by a number of skilled performers who are exceptionally temporally accurate, when temporal accuracy is an important component of the task, e.g., musicians [31]. Thus the relationship between perceived (and actual) temporal constraints and necessary reconciliation effort is scaled by the competent operator to achieve an effective resolution rate. The novice, without immediate access to such information, must engage in a costly *search* among possibilities. This opens novices to misjudgment and a greater potential for threshold transgression, especially when faced with the added demands of emergency or unexpected loads. With greater task uncertainty, skill becomes of lesser use in reducing workload level.

Equifinality of effort is an adaptive strategy whereby the knowledgeable operator scales a constant rate of effort expenditure to attain the desired goal within the temporal horizon set. This adaptive action on behalf of the operator might have little distortional effect in the center of the two-dimensional space in Fig. 1 but is liable to change contour shape at extremes where threshold transition is imminent. The shape of the isodynamic workload contours shown in Fig. 1 will change depending upon the intrinsic

constraints of the task under consideration. Such shapes will be empirically driven according to particular task demands. However, the presentation in Fig. 1 is a reasonable first-pass approximation. The axes are not completely orthogonal conceptually in that, for the adaptive operator, the effective time unit is in part scaled to the content of the task at hand, and vice versa.

Having derived this generic picture of mental workload it is important to indicate the ways in which the suggested axes may be quantified so that prediction may be possible. First, the isodynamic workload contours as phenomena may be tested through performance assessment under sequential manipulations of the two axes. This requires that numerical values be provided along such continua. For the temporal axis it is possible to scale the time to the recorded output of the individual performer, such as may be found in fastest possible response latency for the lower temporal threshold T_F , and freely chosen rate of responding as representative of an individually comfortable work load range, defined *not* as the absence of mental work, but as the experience of an easily reconciled load. Also, end point estimation of extremes may be reflected in an individual operator's points of failure. These can be simply established by manipulating the effective time for response through the reduction of time to perform a task of constant demand level.

As indicated in previous formulations [32]–[36], these scaled axes are liable to be nonlinear functions, where transition from stability to instability will occur rapidly after a critical point is reached. It is progression toward, and transition across such boundaries that are the key elements in predicting workload, as it is the *outcome* of excessive workload rather than the phenomenon alone which is vital in the action of adaptive human-machine systems. To examine these nonlinearities in effort and workload, we have expanded Fig. 1 to give a third dimension as shown in Fig. 2.

The base axes in Fig. 2 are those in Fig. 1. The vertical axis represents mental workload. The function for the change in load on each axis is not arbitrary but adopts an ogival form representing a summated normal range of response [33]. The nonlinearities in the topological structure imply that equal changes in either spatial or temporal constraint do not result in equal changes of mental workload. Rather, in the region of stability, at the top right, a change in workload is easily accommodated. However, toward the identified maximal thresholds accommodating a similar work change requires considerably greater effort on behalf of the operator. Effort is the cost of resolution and represents the force that acts to oppose the load driven by increasing constraints. In response to the nonlinearities noted, the result of a constant effort expenditure toward the goal state is highly contingent upon the operator's location on the surface of the topological structure. Effort expended around the point of stable load is therefore of greater utility than a similar effort in an unstable region.

The profile illustrated in Fig. 2 does not include the operator's progress with respect to the changing loads of multifaceted goals but rather represents progress toward a

unitary goal state. In any real-world task, the operator has a number of embedded hierarchically structured goals that need simultaneous reconciliation. For example, one goal may be related to effective system control in the face of external perturbing influences, while a simultaneous need to coordinate activity with other operators, e.g., formation flight, presents a second source of workload. These embedded concerns summate to provide an overall level of workload that must be considered for any real-world implementations.

In a number of real-world operations, a question of growing importance concerns operator underload. Such conditions occur when the human is required to monitor automatic system operation for prolonged periods with little necessity for active intervention. The model of workload we have presented in Figs. 1 and 2 does not explicitly address the underload problem. In the present formulation, change in workload state is directly proportional to the effort expended by the operator. In simple terms the operator has direct control over task resolution. This affirms that the workload surface (in Fig. 2) is free to move with respect to change in task constraints. However, where the task restricts the effectiveness of operator effort, either overload or underload can be generated that is beyond the control of the operator to reconcile. In vigilance or sustained attention tasks, which typify underloaded situations, the operator's response is contingent upon the appearance of signals for action. These can occur so infrequently, or so randomly, that both operator response speed and response accuracy is depressed. As systems incorporate a greater proportion of computer-based control, the problem of operator underload will grow. As yet there is relatively little understanding concerning this question and the appropriate techniques through which to reduce such unwanted effects [7].

Our conceptual formulation of dynamic mental workload includes the notions of operator skill, effort, and the time pressure under which the task is carried out. In our view, the exploration of this relationship will be assisted by the description of a mathematical function that will capture this relation under varying task conditions. One function that has proven useful in fitting psychophysical data is the power function [37]. Although some have used the power function as a theoretical prediction of psychophysical response, the power function is more frequently used as a convenient framework for describing psychophysical data which form the basis for scaling techniques. Using a similar approach, a workload function can be formulated that will describe and parameterize experienced workload in terms of the critical parameters of skill, effort, and temporal constraints. A workload function was chosen whose mathematical properties conformed to the model of workload outlined before.

Thus we assume as a first approximation that overall workload may be captured in the following power function:

$$W = 1/et^{s-1} \quad (1)$$

where W equals the workload level, e is the effort ex-

pendent by the individual operator, t is the effective time available for action, and s is the skill level of the performer under consideration. The shape of this function for different values of t , e , and s , is shown in the panels of Fig. 3. t is normalized to vary between zero and one, as do e (strictly greater than zero) and s . One feature of this equation is that it can be linearized through logarithmic transformation, allowing parameter estimation through linear regression and related statistical techniques. We suggest that this equation may be a useful way of organizing research findings relating to the dynamic definition of workload proposed in this paper. However, the use of this workload function does not solve the problem of workload assessment, since the problem of how effort, skill, and temporal constraint should be scaled or quantified still remains. It is possible that techniques such as monotone regression, conjoint measurement, and multidimensional scaling may prove useful in this endeavour, but such topics lie outside the scope of this paper.

Returning to the qualitative characteristics of the present model, two advantages of operator skill are, first, the reduction of effort needed to reconcile the demands of a known task. Second, knowledge of the load profile that a particular task imposes allows the experienced operator to project epochs where workload would be greatest and effort most needed. This proactive perspective promotes the ability to remain in the stable region where load resolution per unit effort is most effective. Like *equifinality of effort*, this process is viewed as an adaptive strategy that maximizes the use of this energetic characteristic of limited effort and thus represents the most economic approach to task resolution.

In the extended view, instantaneous workload can be seen as the changing angle of the slope as the operator expends more or less task-related effort. The overall workload of a task is represented as the area under the curve of the effort profile that the operator engenders to achieve the goal state. Dynamic assessment of mental workload in terms of both time and task goals may provide the capability of adaptive task loading. In Fig. 4, mental workload level is plotted not as a function of effective time to reach a desired goal, as in Fig. 1, but simply as a chronological progression, although workload is recognized as a multidimensional representation (line D) in the figure. In a similar manner, task demand has also been compacted to a unidimensional representation, given by line E. In a manner analogous to tracking behavior, the workload signal follows that of demand with a small lag as is illustrated by the two initial peaks. We have also illustrated a case where the workload signal exhibits a small negative gain with respect to task demand as shown by the region of unacceptable underload. Further, it is possible that minor fluctuations in task demand and workload may exhibit an insensitivity in the region of stable loading centered around line B, as illustrated in the ascending portion following the region of unacceptable underload. The primary purpose of the adaptive interface is to maintain task response efficiency while preventing the operator workload from violating the thresholds of stable acceptable load as given by

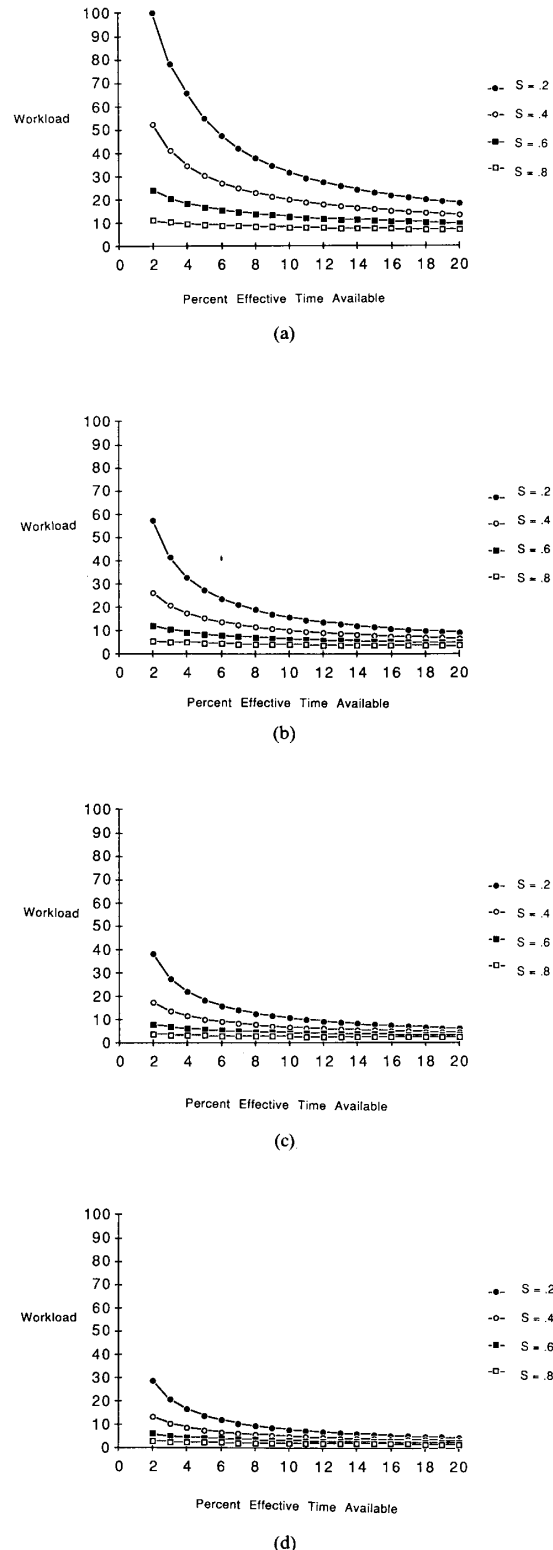


Fig. 3. Plots of mental workload function (shown in (1)) for different values of e , s , and T , with each value scaled to range between zero and one. (a) $e = 0.2$. (b) $e = 0.4$. (c) $e = 0.6$. (d) $e = 0.8$.

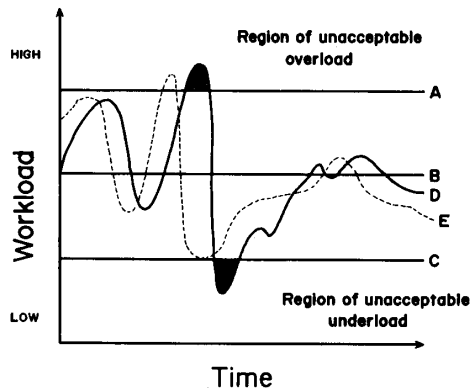


Fig. 4. Change in level of mental workload as function of chronological progression. Within illustration, level of workload (line D) follows that for task demands, represented by line E. Workload may track demand with small lag, as in fluctuations at left. Some gain may be expressed as shown for region of underload, or workload signal may be insensitive to minor changes in load in stable region as illustrated around line B. There is little evidence concerning conditions where workload may lead changes expected in task demand. Purpose of adaptive interface described in text is to eliminate unacceptably high- or low-load levels as given by hashed areas.

lines A and C in the diagram. It is possible that workload and related effort may lead task demand, when the parameters of the task are available to the operator prior to performance. As yet, experimental evaluation of the notion of load is somewhat sparse, although the retrospective impact of previous workload level upon current load has been the subject of recent investigation [38].

Fig. 4 presents a change in mental workload level as a task proceeds. In this illustration the time axis is not the perceived temporal horizon for action within the task, but the external chronological progression or reference temporal frame for the whole system.

IV. ADAPTIVITY IN TASK LOADING

Adaptive task loading assumes that a reliable assessment of both demands and resources is available. In addition to understanding and describing the task itself, it is important to distill the salient features of the task in terms of the demand imposed on the operator. There are a number of different approaches to such an assessment. One method is to assume that demands and resources are independently measurable [39]. In this case independent assessment is made of the levels of different types of resource required and the demands of the task, and these are compared explicitly with assessments of the current resources of the operator. Currently, the measurement of specific resources is problematic in spite of the promise of indicators such as pupil dilation [40] or more specific (and possibly diagnostic) measures such as evoked cortical potentials [41].

Adaptive interfaces should not only assess operator workload, they should also diagnose the features of the tasks that are contributing to the current workload to provide a basis for task reconfiguration. This implies a workload-oriented view of the task and of operator processing opera-

tions. Given the many ways that tasks and related perceptual and cognitive processing have been described, what specific approaches provide the workload-oriented view required in adaptive interfaces? A helpful initial distinction is between "horizontal" and "vertical" models of human information flowing through a series of stages, beginning with perception and ending with a motor response of some kind [28], [42]. In contrast, vertical models view task performance in terms of the types (rather than stages) of processing that occur. The difference between horizontal and vertical models of human performance is a somewhat analogous to the difference between control processes and data structures in computer programming. Horizontal models approximate a flowchart, indicating which sequence of processes transformed the input into the output. Vertical models indicate which data structure (and, by implication, which processes) are utilized for different tasks and operations.

Horizontal models focus on the flow of information processing activity engendered by the task. At present the modal horizontal model [42], [43] distinguishes between basic processes of encoding, decisionmaking, and response selection and execution. In a complex activity, several stages of the horizontal model might be active as the person engages in the task. Overlap of information processing stages is plausible [44] but difficult to identify with respect to alternative serial or parallel models of activation [45]. For purposes of simplicity most tasks are modeled as if stimuli are handled one at a time, but this is clearly a convenience rather than a veridical reflection of actual operations. Tasks of certain types can be carried out in parallel [43], but existing methods of measurement cannot identify what activities occur and in what sequence. This problem is magnified in actual complex multiple activity tasks. Reliance on identifying the processing resources required by complex tasks is problematic since there are many such resources [43], [46].

Previous attempts at linking task components to operator workload assumed that under some circumstances tasks can be broken down into separate units which consume distinct and separable processing resources using the horizontal model [40], [41]. However, it is also possible to link task models to workload using vertical models. This is particularly true in the case of skill-based and rule-based behaviors as elaborated by Rasmussen [47] where there is a fairly close mapping between conditions in the environment and resulting performance. However, the relatively abstract nature of knowledge-based behavior is more difficult to quantify and, as Rasmussen pointed out, less amenable to experimental observation.

According to vertical models such as that proposed by Rasmussen or the GOMS model [21], human operators have a repertoire of activities available to them. The way in which this repertoire is used is constrained not only by the requirements of the task but also by the context in which the task is performed. The two most pertinent aspects of this context are the environmental stressors (both physiological and psychological) that bear on the person and the temporal sequence of the task in relation to deadlines,

circadian rhythms, and other temporal factors that impact the operator.

In analyzing tasks into workload related components, we initially examined horizontal models [39]; however, vertical models appear to hold greater promise for adaptive interface design. The current focus on vertical models is reinforced by the observations that more progress appears to have been made in analyzing tasks from the structural (vertical) rather than process (horizontal) view [22] and the processes considered in current horizontal models are generally too microscopic and short-lived to be readily adapted to longer cycles of task activity.

V. A FRAMEWORK FOR AN ADAPTIVE INTERFACE

Complex systems require interfaces that adapt to the needs of the task and the operator. There are a number of ways in which a system might adapt to the operator by utilizing information about the level of mental workload experienced. The first step is the development of a basis or criterion for adaptivity. In the case of adaptivity to task loading, this signal may represent the mismatch between the current task demands and the available capacity of the human operator. In keeping with the view expressed in Fig. 3, task demands generate workload, although the relationship is time-lagged and filtered by individual skill, motivation, and arousal level [48]. The identification of this workload criterion upon which adaptation is to be based is then a two-stage process.

- 1) A global estimation of the MWL mental workload is made, i.e., is MWL too high or too low and by how much?
- 2) Diagnostic information is used to assess the load on specific capabilities.

We distinguish here between two classes of method for estimating mental workload, namely assessment methods and predictive methods. A number of assessment methods have been proposed based on techniques for physiological and psychological measurement [4]–[7]. Predictive methods are advantageous in that they allow anticipation in the adaptation process. However, prediction of workload has to the present been problematic. One of the advantages of the dynamic definition of workload given earlier in this paper is that such prediction can now become the primary source of assessment.

In consideration of (1), regression methods may be used to estimate the parameters of this function and then predict workload based on operator skill and experience, combined with the time constraints and the nature of the task components. Such methods will give an approximation, due to experimental error and related predictive inefficiencies in the regression equation. As our understanding of tasks and the relationship between task demands and mental workload increases, these simple regression techniques will be replaced by superior methods of workload prediction. One feature of the formulation of

workload given above is the interaction between operator skill and the components of the task. This interaction will tend to confound regression methods.

Given a satisfactory method of workload prediction, the next step in adaptivity is to ascertain whether or not stable workload levels (as shown in Fig. 1–3) are being exceeded. If the workload is within acceptable limits, the system continues to monitor the situation without taking adaptive steps, otherwise the source of overload is diagnosed in terms of features of the task that can be adjusted. The type of diagnostic information that is appropriate depends on the way in which the task can be adjusted. Once this diagnosis has been made, one strategy is simply to offload some or all of the task components to another operator, assuming a multioperator system. This offloading strategy can also be extended to machine components [20].

In our previous work we have focused on the use of dynamic task reallocation to maintain an optimal level of workload for the operator [20]. Dynamic task reallocation requires an adaptive mechanism that can assess the mismatch between task demands and available capacity and redefine the task so as to reduce this mismatch. More complex strategies may involve reconfiguring the task, removing some components and adding others. The ability to reconfigure tasks often occurs when a complex sequence of activities is to be performed and there is some freedom in scheduling those activities. A third type of strategy is to reduce the temporal component of workload by providing more time in which to carry out the current task components. This may be achieved by extending task deadlines or by offloading future activities to other sources, thereby mixing the first and third strategies, or by using augmenting aids such as quickened displays. In general, various combinations of these strategies may be found useful in adaptive task loading.

In this section we focus on a model of adaptivity where task components are offloaded to other agents (either human or machine). In this case the current workload should be diagnosed in terms of the task components which are contributing most to the current overload condition. This diagnosticity requires a finer analysis of workload into task-related components. Although it is possible to define these task components in terms of task-specific operators, it is preferable to develop a framework for adaptive interfaces that applies across a variety of tasks. Two approaches which provide this general perspective are multiple attentional resource theory [42], [43], and the process-oriented views of Rasmussen [47]. Diagnosing tasks in terms of the attentional resources that they require appears to be difficult [39]. A simpler analysis in terms of the relative amounts of knowledge-based, rule-based, and skill-based behavior in task performance may be used instead [47].

Given that workload is assessed, and the task-related sources of workload are identified, adaptivity is then implemented through a reasoning process that selects a task allocation (or reconfiguration or scheduling) policy that changes the loading on the human in such a way as to

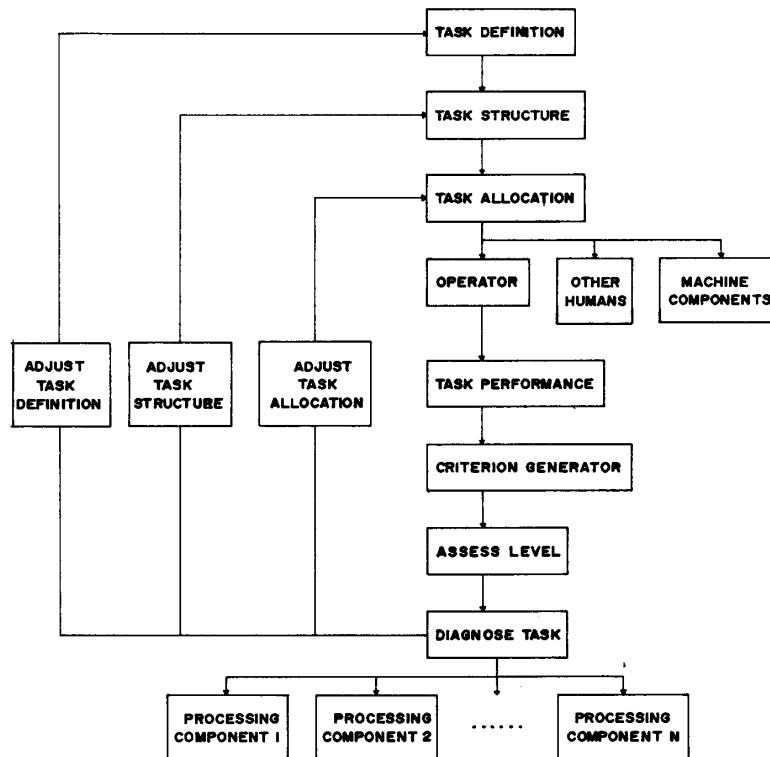


Fig. 5. Overall structure of knowledge-based adaptive mechanism that serves as adaptive interface for dynamic task allocation and reallocation.

improve overall system performance. This process must have access to both overall system goals (a model of the task) and information about what the person and machine components of the system are capable of accomplishing (person and system models). In addition, the reasoning process should decide which loading strategies and task allocation policies are available for selection, and the implications of the current task loadings.

We have illustrated this conceptual framework for adaptive task loading in Fig. 5. We begin with a view of the task as being first defined, then structured, and finally allocated among different machines and operators. The operator then performs the task as assigned and a level of performance ensures that reflects the operator's capabilities and effort compared with the nature and difficulty of the assigned task. One may then assess the operator's performance and identify a criterion for adaptivity. This criterion might, for instance, be a measure of mental workload, a measure of primary task performance, or some combination of workload and performance.

Once a criterion of operator response to the task is known, an adaptive policy may be implemented. In general, the criterion will be a dynamically changing input which has to be combined with the static information provided by the other knowledge sources available to the loading strategy reasoner. It should provide information about how well the person is performing particular sub-tasks and indicate any discrepancies between overall sys-

tem performance and system goals since it would be inappropriate to optimize the workload experienced by the operator if the overall system failed or suffered as a consequence.

In principle, models of the task, system, and person allow prediction of the effect of alternative task redefinitions and allocations, but in practice the models currently available do not provide the quantitative predictions that are needed. However, certain features of adaptive task loading appear to be self-evident, and these may be combined into a framework for the design of adaptive interfaces which utilize such information as is already available. For instance, only tasks that can be switched reasonably between the human and machine components of the system should be considered for reallocation, except under emergency conditions which demand continued performance, with the result of system failure having intolerable consequences. Similarly, the output of the adaptive interface will be a definition of the task that alters (where necessary) the loading of task components between the human and machine so as to improve future measures of the criterion. The basis for adapting the task will be a diagnostic assessment of which processing components in the current task allocation are leading to problems as observed in the criterion.

This diagnosis then implies an adjustment in the basic task definition, the structure of the task, or the way in which the task is allocated. This threefold distinction in

mechanisms for adaptation corresponds to the product design, the process plan, and scheduling of human and machine activities in manufacturing. Continuing with this analogy, one can attempt to solve production problems by changing the schedule (switching machine allocations, etc.), by changing the process plan, i.e., changing the way in which processes are carried out, or, most radically, changing the actual design. In most situations, of course, changing the allocation or scheduling policies will be appropriate. There appear to be a number of useful analogies between the problems faced in operating flexible manufacturing systems and the problem of designing adaptive interfaces. We shall not pursue these in the present paper because our focus is on the implications of adaptive interfaces for mental workload definition and assessment. We expect, however, that adaptive task loading will generally involve adaptive human-machine interface that acts as a servomechanism minimizing the difference between current demands and available capacity.

As discussed before, Fig. 5 provides an overall conceptual framework for adaptive task loading. Different adaptive interfaces are possible depending on how the three fundamental questions posed by the figure are answered. These questions are, What is the criterion generator? What is the model for task diagnosis in terms of operator processing components? and What models of the task allow appropriate redefinition, restructuring, and reallocation while a task is being carried out?

The simplest form of adaptation reallocates task components. The question of reallocation policy is contingent upon what is known of the operator's capabilities. As Rasmussen [47] has noted, one approach to this question differentiates a *set* of models rather than one global structure. As a result, this distinction between skill-, rule-, and knowledge-based behavior defines multiple levels of action. For the adaptive interface it is important to distinguish between these loading characteristics. Offloading an operator of skill-based activity would be of little use when the overload emanates from a task primarily composed of knowledge-based demands. Consequently, the permanent knowledge base concerning the operator characteristics, embedded in the interface, would need to distinguish between these types of activity. Similarly, the load-leveling mechanism would require some assessment of the relative loading of each element from the incoming mental workload signal.

Perceived distance to goal state is also contingent upon the appraisal of the capabilities needed to acquire such a target. For example, even an operator highly skilled in manual control will perceive a task to be particularly loading if task demands emphasize knowledge-based resolution. Thus, for each of Rasmussen's three levels of performance, differing types of load could be distinguished. The single axis in both Figs. 1 and 2 reflects just one of the multidimensional aspects of human capability and their differential demand by changing human-machine tasks.

Thus adaptation may occur in three major ways. First, the allocation of subtasks between human and machine components may be adjusted. Second, the structure of the

task, including the processes and methods used to carry it out may be adjusted. These adjustments will then be propagated through the ensuing task allocation. Finally, the definition of the task itself may be adjusted. For instance, evaluative criteria could be changed so that more operator errors are tolerated, or the scope of a task could be altered so that, when propagated through the task structuring and task allocation stages, the resulting load on the human operator would be reduced. Needless to say, the type of adjustments to task definitions, task structures, and task allocations will depend greatly on the application domain. Adjustments in this framework are based on a diagnosis of task performance in terms of workload. The model of workload outlined in this paper is a global one which provides a measure of overall experience workload. However, adjustment of the task and its implementation must be based at some point on diagnostic criteria which point to the adjustments that should be made. In the future, perhaps, these diagnostic criteria would be based on the experienced workload for each of a number of different processing components. At present, however, on-task analysis of workload for specific processing components is not possible. Thus versions of load-mediating interfaces in the near future would use inferential tools to suggest which aspects of the current task are overloading specific processing components.

VI. CONCLUSION

In the present framework, mental workload is described in terms of time, distance to desired goal, and effort. This initial description is linked to a spatial representation that promises some advantage in resolving the problem of assessment. The quantification of the axes presented is a vital component in development, as without such information the present suggestion remains only a descriptive framework within which to incorporate empirical results. Although the scaling of salient points along the axes will be idiosyncratic to a certain extent, it may be possible to extract significant commonalities or *invariants* across effort directed toward task resolution. This can emerge through the use of a consistent "simple" task. Using a vector representation imposed upon Fig. 1 for such a common task and having access to optimal effort solutions may provide *prediction* of workload, rather than the traditional *post hoc* assessment.

The relation between stress, workload, and subsequent performance is a complex topic that remains under investigation [32]–[34]. In complex human-machine tasks, stress may have a critical effect on performance. We can distinguish between stress that is a routine part of a task and unusual stress that may qualitatively change the way in which the task is performed. The type of adaptive interface considered here applied to scenarios where stress remains within manageable bounds and where mental workload effects stem from the joint influence of task demands and temporal constraints upon satisfaction of the task requirements.

We also outlined the role of vertical models of task performance in the new approach to mental workload based on the addition of the temporal dimension. This should lead to a methodology for adaptive interface design which assess the types of behavior required by the task and matches the task demands to the profile of available behavior given the temporal constraints and stressors. Elsewhere we have discussed in more detail the architecture for adaptive interfaces based on mental workload assessment and machine reasoning [22]. In our earlier work, we assumed that an explicit accounting of attentional resource utilization would be required in an adaptive interface. However, we have also reviewed the difficulties associated with basing the development of task loading strategies on a multiple attentional resource approach [39]. It now appears that the revised perspective on mental workload and the simpler model of behavior used in this paper will prove more tractable as a guide in the design of interfaces for adaptive task loading [49]. Useful adaptive processes should be possible even with fairly simple models of tasks providing that these models can be linked to mechanisms for controlling the levels of stress and mental workload in the operational environment.

The concepts of stress and mental workload are often applied to human-machine interactions where catastrophic failure can potentially occur. However, workload and stress have an impact on human-computer tasks in general, not only on long term health effects [50] but also in terms of task performance. The development of adaptive interfaces allow human-computer interaction to occur in an environment where the effects of mental workload and stress have been modulated and controlled.

ACKNOWLEDGMENT

The authors would like to thank Professor Daniel Gopher, and an unknown reviewer for their insightful comments on an earlier version of this work.

REFERENCES

- [1] W. B. Rouse, "Optimal allocation of system development resources to reduce and/or tolerate human error," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, pp. 620-630, Sept./Oct. 1985.
- [2] N. M. Morris, W. B. Rouse, and S. L. Ward, "Information requirements for effective human decision making in dynamic task allocation," in *Proc. IEEE Conf. Syst. Man Cybern.*, 1985, pp. 720-724.
- [3] R. A. Hess, "A qualitative model of human interaction with complex dynamic systems," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-17, pp. 33-51, Jan./Feb. 1987.
- [4] N. Moray, Ed., *Mental Workload: Its Theory and Measurement*. New York: Plenum, 1979.
- [5] D. Gopher and E. Donchin, "Workload: An examination of the concept," in *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman, and J. P. Young, Eds. New York: Wiley, 1986.
- [6] R. D. O'Donnell and F. T. Eggemeier, "Workload assessment methodology," in *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman, and J. P. Young, Eds. New York: Wiley, 1986.
- [7] P. A. Hancock and N. Meshkati, Eds., *Human Mental Workload*. Amsterdam, The Netherlands: North-Holland, 1988.
- [8] S. K. Card, "Human factors and artificial intelligence," in *Intelligent Interfaces: Theory, Research, and Design*, P. A. Hancock and M. H. Chignell, Eds. Amsterdam, The Netherlands: North-Holland, in press.
- [9] M. H. Chignell, P. A. Hancock, and A. Loewenthal, "Intelligent interface design," in *Proc. IEEE Conf. Syst. Man Cybern.* 1985, pp. 620-623.
- [10] E. Rich, "Users are individuals: Individualizing user models," *Int. J. Man-Machine Studies*, vol. 18, pp. 199-214, 1983.
- [11] R. C. Williges, B. H. Williges, and J. Elkerton, "Software interface design," in *Handbook of Human Factors*, G. Salvendy, Ed. New York: Wiley.
- [12] A. Gershman and T. Wolf, "The dynamics of man-machine interaction in a conversational advisory system," in *Proc. Conf. IEEE Syst. Man Cybern.*, 1986, pp. 721-725.
- [13] R. E. Cullingford and J. L. Kolodner, "Interactive advice-giving," in *Proc. IEEE Syst. Man Cybern.*, 1986, pp. 709-714.
- [14] D. A. Norman, "Cognitive science and cognitive engineering," in *Interfacing Thought*, J. B. Carroll, Ed. Cambridge, MA: MIT Press, 1987.
- [15] M. H. Chignell and P. A. Hancock, "Integration of the cognitive and Physical aspects of the human-machine interface," *Proc. Hum. Fact. Soc.*, vol. 30, pp. 1007-1011, 1986.
- [16] C. Smith, C. Irby, R. Kimball, W. Verplank, and E. Harslem, "Designing the Star user interface," *Byte*, vol. 7, pp. 242-282, 1982.
- [17] E. L. Hutchins, J. D. Hollan, and D. A. Norman, "Direct manipulation interfaces," in *User Centered System Design: New perspectives on Human-Computer Interaction*, D. A. Norman and S. W. Draper, Eds. Hillsdale, NJ: Erlbaum, 1986.
- [18] R. J. K. Jacob, "Direct manipulation in the intelligent interface," in *Intelligent Interfaces: Theory, Research, and Design*, P. A. Hancock and M. H. Chignell, Eds. Amsterdam, The Netherlands: North-Holland, in press.
- [19] D. Gentner and A. Stevens, *Mental Models*. Hillsdale, NJ: Erlbaum, 1983.
- [20] M. H. Chignell and P. A. Hancock, "Knowledge-based load leveling and task allocation in human-machine systems," in *Proc. Ann. Conf. Manual Control*, vol. 21, 1985, pp. 9.1-9.11.
- [21] S. K. Car, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum, 1983.
- [22] P. G. Polson, "Qualitative and quantitative models of human-computer interaction," in *Interfacing Thought*, J. B. Carroll, Ed. Cambridge, MA: MIT Press, 1987.
- [23] M. H. Chignell and T. J. Higgins, "An intelligent warning system for instrument landings," *Int. J. Indust. Erg.*, 1987, vol. 2, pp. 77-89.
- [24] W. B. Rouse and W. J. Cody, "Function allocation in manned system design," in *Proc. IEEE Conf. Syst. Man Cybern.*, 1986, pp. 1600-1606.
- [25] P. A. Hancock and D. M. Carson, "The time trap: Temporal incongruencies under stressful conditions," *Aviat. Space. Environ. Med.*, 1988.
- [26] E. C. Zeeman, *Catastrophe Theory: Selected Papers 1972-1977*. Reading, MA: Addison-Wesley, 1977.
- [27] P. A. Hancock, "The effect of skill on performance under an environmental stressor," *Aviat. Space Environ. Med.*, vol. 57, pp. 59-64, 1986.
- [28] R. Lachman, J. L. Lachman, and E. C. Butterfield, *Cognitive Psychology and Information Processing: An Introduction*. Hillsdale, NJ: Erlbaum, 1979.
- [29] P. A. Hancock and M. H. Chignell, "Functional synchronization: A theory for human-computer interaction," in *Proc. 2nd Int. Conf. Human-Computer Interaction*, Aug. 1987.
- [30] W. Schneider and R. M. Shiffrin, "Controlled and automatic human information processing: Detection, search and attention," *Psych. Rev.*, vol. 84, pp. 1-66, 1977.
- [31] S. Sternberg and R. L. Knoll, "Perception, production, and imitation of time ratios by skilled musicians," *Ann. NY Acad. Sci.*, vol. 423, pp. 29-44, 1984.
- [32] P. A. Hancock, "The role of temporal factors in workload prediction," in *Proc. IEEE Conf. Syst. Man Cybern.*, 1986, pp. 1049-1053.
- [33] —, "Stress and adaptability," in *Energetics and Human Information Processing*, G. R. J. Hockey, A. W. K. Gaillard, and M. G. H. Coles, Eds. Berlin: Nijhoff, 1986.
- [34] P. A. Hancock and M. H. Chignell, "Toward a theory of mental workload: Stress and adaptability in human-machine systems," in *Proc. IEEE Conf. Syst. Man and Cybern.*, 1986, pp. 378-383.

- [35] —, "The principle of maximal adaptability in setting stress tolerance standards," in *Trends in Ergonomics/Human Factors II*, R. Eberts and C. Eberts, Eds. Amsterdam, The Netherlands: North-Holland, 1985.
- [36] —, "Adaptive control in human-machine systems," in *Human Factors Psychology*, P. A. Hancock, Ed. Amsterdam, The Netherlands: North-Holland, 1987.
- [37] S. S. Stevens, "On the psychophysical law," *Psych. Rev.*, vol. 64, pp. 153-181, 1957.
- [38] M. L. Matthews, "The influence of visual workload history on visual performance," *Hum. Fact.*, vol. 28, pp. 623-632, 1986.
- [39] M. H. Chignell and P. A. Hancock, "A comparison of mental workload and available capacity in complex person-machine systems," in *Applications of Fuzzy Set Theory in Human Factors*, W. Karwowski and A. Mital, Eds. Amsterdam, The Netherlands: Elsevier, 1986, pp. 271-288.
- [40] D. Kahneman, *Attention and Effort*, Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [41] J. B. Israel, C. D. Wickens, G. L. Chesney, and E. Donchin, "The event-related brain potential as an index of display monitoring workload," *Hum. Fact.*, vol. 22, pp. 259-273, 1980.
- [42] C. D. Wickens, *Engineering Psychology and Human Performance*. Columbus, OH: Merrill, 1984.
- [43] —, "The structure of processing resources," in *Attention and Performance VIII*, R. Nickerson and R. Pew, Eds. Hillsdale, NJ: Erlbaum, 1980.
- [44] J. L. McClelland, "On the relations of mental processes: An examination of systems of processes in cascade," *Psych. Rev.*, vol. 86, pp. 287-330, 1981.
- [45] J. T. Townsend and G. Ashby, *Mathematical Models of Response Times and Stochastic Processes*. London: Cambridge Univ. Press, 1984.
- [46] D. Navon and D. Gopher, "On the economy of the human processing system," *Psych. Rev.*, vol. 86, pp. 214-255, 1979.
- [47] J. Rassmussen, "Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, pp. 257-266, 1983.
- [48] L. Bainbridge, "Problems in the assessment of mental load," *Trav. Hum.*, vol. 37, pp. 279-302, 1974.
- [49] M. H. Chignell and P. A. Hancock, "Intelligent interfaces," in *Handbook of Human-Computer Interaction*, M. Helander, Ed. Amsterdam, The Netherlands: North-Holland, in press.
- [50] R. A. Karasek, R. S. Russell, and T. Theorell, "Physiology of stress and regeneration in job related cardiovascular illness," *J. Hum. Stress.*, pp. 29-42, Mar. 1981.



Peter A. Hancock (A'87) received the Teacher Certificate and B. Ed. degrees from the Loughborough College of Education, Leicestershire, England, the M.Sc. degree from Loughborough University in 1978 and the Ph.D. degree from the University of Illinois, Urbana, in 1983.

He currently holds a joint appointment as Assistant Professor in the Departments of Safety Science and Human Factors at the Institute of Safety and Systems Management, the University of Southern California, Los Angeles. His re-

search concerns include the study of time and the effects of stress and mental load upon operator performance.



Mark H. Chignell (PM'87) received the M.S. degree in industrial and systems engineering from Ohio State University, Columbus, and the Ph.D. degree in psychology from the University of Canterbury, New Zealand.

He is currently an Assistant Professor of industrial and systems engineering at the University of Southern California, Los Angeles. His research involves the development of intelligent interfaces between humans and machines.

Dr. Chignell is the coauthor of a book on expert systems and has published articles in the areas of engineering design, aviation, and on-line information retrieval.