W/INDEX: A PREDICTIVE MODEL OF OPERATOR WORKLOAD

Robert A. North and Victor A. Riley

Honeywell Inc.
Systems and Research Center
Minneapolis, MN

INTRODUCTION

One of the major developments in weapon system design over the past decade is the emergence of technologies that enable single crewmembers to operate very complex systems in highly dynamic environments. However, the growing complexity of the tactical environment is keeping pace with the ability of new automation technologies to deal with it, producing a constant tension between the economic and practical forces that drive crew size reduction and the mission performance requirements that favor larger crew sizes.

The crucial factor in the middle of this tension is crew workload. When all of the decision-making, systems monitoring, planning, situation assessment, and control responsibilities have been allocated between human and machine, the question of whether the operator's workload capacity has been exceeded is likely to determine the feasibility of the design. In the most complex systems operating in the most complex environments, where a single automation failure may make the operator's job unmanageable, it is especially important to consider factors that will make the job unnecessarily difficult by contributing to the operator's workload.

System designers need a tool that allows them to derive the crewstation design and automation configuration that produces the most manageable workload levels. This tool should permit rapid consideration of a wide range of design options, and should be useful at any stage of the design process, from high level concept generation through advanced development. In order to be useful in analyzing a complex environment, where an operator's attention is likely to be shared between multiple tasks over much of the mission, the tool must employ a realistic model of attentional timesharing and impose appropriate levels of workload penalties to account for different levels of conflicts between multiple tasks.

Finally, the tool should facilitate detailed and systematic consideration of all the major task or design attributes that contribute to workload. As new automation technologies mature and the operator's task becomes more management and decision-making than control, this should include consideration of the operator's cognitive processing as well as manual and sensory demands.

WORKLOAD INDEX (W/INDEX)

Honeywell Systems and Research Center has developed a computer-based tool to predict operator workload produced by specific crewstation designs over the course of representative mission scenarios. The Workload Index (W/INDEX) software tool, which runs in both MS-DOS and VAX-VMS environments, allows system designers to consider the workload consequences of decisions involving the physical layout of the crewstation, the application of automation to specific crew tasks, the use of various human-machine interface technologies, and the sequence of crew task loading. It is intended to be used iteratively at any stage of the design process to identify the best crewstation configuration, that is, the best combination of geometry, automation, and crew interface technologies that produces the lowest predicted workload over a wide range of mission conditions.

To use W/INDEX, the analyst must supply information for three W/INDEX databases: a task timeline, an interface/activity matrix, and an interface conflict matrix. These are indicated in Figure 1, and will be described in the following paragraphs.

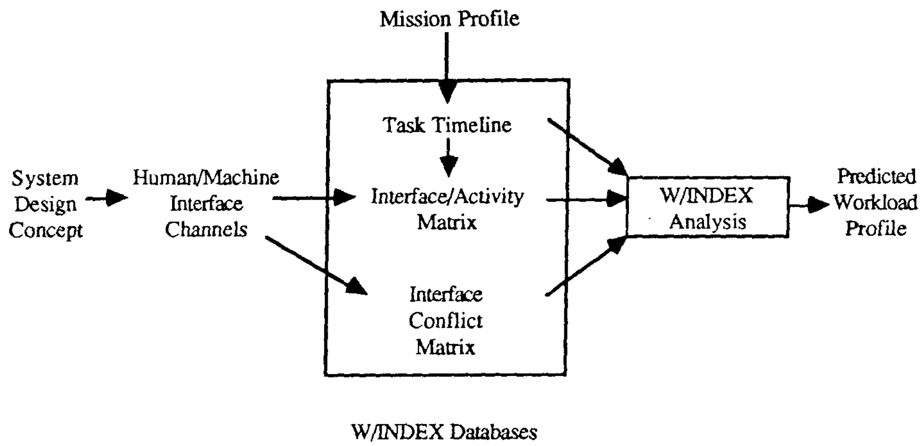

Figure 1. W/INDEX Data Flow

The task timeline is derived from a mission profile. It represents the specific tasks performed by the operator during mission performance and when they occur on the timeline. Each task may be assigned up to 20 start and stop times.

The interface/activity matrix is derived from the task timeline and from the human/machine interface channels, which represent the design concept. In the early stages of design, these interface channels may be very high level, such as "visual", "auditory", "manual", and "verbal". This permits the designer to predict potential overtaxing of one of these types of operator resource. During detailed design, specific controls and displays may be represented as channels. What the interface/activity matrix does is specify the amount of attention the operator must pay to each channel in the performance of each task. An example is shown in Figure 2. The numbers in the matrix are on a five point subjective scale, with 1 being very low attentional demand and 5 being very high. The two-letter identifier preceding the activity name referred to an aircraft system (e.g., FC = flight control, etc.)

The conflict matrix is derived solely from the design concept and represented in terms of interface channels. It specifies the degree of workload penalty that results when the operator must attend to multiple channels simultaneously, and gives the designer a means of considering the physical and cognitive capabilities of the operator in the context of the specific design being evaluated. The use of this matrix and the theory behind it will be discussed in detail in the next section.

Having provided each of these three types of data, the analyst simply commands W/INDEX to calculate a predicted workload level for each half second in the timeline. W/INDEX does this for each half second by summing the attentional demands in each

INTERFACE/ACTIVITY MATRIX

ACTIVITY NAME	CG	TH	TD	FC	FD	LT	LD	CD	RD	SP	HP
FC-AIR-TO-AIR	0	2	0	0	5	0	0	0	0	0	0
FC-HIGH-LEVEL	0	0	0	0	1	0	0	0	0	0	0
FC-INIT IFFC	0	0	2	0	0	0	0	0	0	0	0
NV-EVAL INTERCE	3	0	0	0	0	0	0	0	0	0	0
NV-FOLLOW STRG	2	0	0	0	0	0	0	0	0	0	0
NV-MONITOR INTE	1	0	0	0	0	0	0	0	0	0	0
NV-MONITOR PATH	2	0	0	0	0	0	0	2	0	0	0
NV-RECOG JM EFF	4	0	0	0	0	0	0	0	0	0	0
NV-RECOG TRACK	2	0	0	0	0	0	0	0	0	0	0
NV-VER FUEL	0	0	0	0	0	0	0	0	0	0	0
CO-ACC ASSIGN	1	0	0	0	0	0	0	2	0	0	0
CO-ACC MESSAGE	0	0	0	0	0	0	0	2	0	0	0
CO-AEW ALERT	3	0	0	0	0	0	0	0	0	0	0
CO-OB WING STAT	0	0	2	0	0	0	0	0	0	2	0
CO-REQ MESSAGE	0	0	0	0	0	0	0	3	0	0	0
CO-REV MESSAGE	5	0	0	0	0	0	0	0	0	0	0
CO-VER WING JAM	0	0	2	0	0	0	0	0	0	3	0
ASE-DETECT JAM	3	0	0	0	0	0	0	0	0	0	0
ASE-VER ECM	0	0	0	0	0	0	0	0	0	0	0
FRC-MAS ARM ON	0	0	0	0	0	0	0	2	0	0	0
FRC-VER WEAPON	0	0	0	0	0	0	0	0	0	2	0
MM-COMP FENCE	2	0	0	0	0	0	0	0	0	0	0
MM-EVAL SENS RN	2	0	0	0	0	0	0	0	2	0	0
MM-INIT PAS SRC	1	0	0	0	0	0	0	3	0	0	0
MM-MONITOR SYS	1	0	0	0	0	0	0	0	0	0	0
MM-PERFORM ID	2	0	0	0	0	0	0	2	0	0	0
MM-SELECT A/A	0	0	0	0	3	0	0	0	0	0	0
MM-VER DISPLAY	1	0	0	0	0	0	0	0	0	0	0

Figure 2. Example Interface/Activity Matrix.

interface channel across all the tasks that occur during that half second. These interface channel attentional demand totals are then summed to produce an additive workload estimate. Then, the additional costs of timesharing between multiple channels are determined by multiplying the sum of each pairwise combination of channels with the corresponding conflict matrix value. This is done both within channels and between channels. These timesharing costs are then added on to the current half second total to produce an instantaneous workload level for that time period.

The algorithm can be expressed as follows:

$$W_T = \sum_{i=1}^{l} \sum_{t=1}^{m} a_{t,i} + \sum_{i=1}^{l} \left[(n_{t,i} - 1)\, c_{ii} \sum_{t=1}^{m} a_{t,i} \right] + \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} c_{ij} \sum_{t=1}^{m} (a_{t,i} + a_{t,j})$$

Where:

W_t = instantaneous workload at time T

i,j = 1..l are the interface channels

t = 1..m are the operator's tasks or activities

$n_{t,i}$ = number of tasks occurring at time t with nonzero attention to channel i

$a_{t,i}$ = attention to channel i to perform task t

$c_{i,j}$ = conflict between channels i and j

and where:

$a_{t,i}$ and $a_{t,j}$ are both nonzero.

The first term of this formula provides the purely additive workload level, while the second term provides the penalty due to demand conflicts within channels and the third term provides the penalty due to demand conflicts between channels.

Having calculated instantaneous workload for each half second in the mission timeline, W/INDEX then provides a five-second average. This average extends to two seconds before time t to two seconds after time T, and accounts for the operator's discretion to anticipate or delay tasks to smooth instantaneous workload changes. This provides a smooth workload profile over the mission timeline.

A sample of W/INDEX output is shown in Figure 3. This chart, (a Macintosh-drawn compilation of four W/INDEX runs), shows four different combinations of automation configurations and interface technology options. The minimum path represents the lowest workload crewstation design. In this case, it switches between design conditions, suggesting that the lowest workload can be achieved when several interface options are available to the operator and the operator can choose which one to use based on other concurrent activities.

In this way, W/INDEX considers the difficulties of the tasks, the geometry of the crewstation, the availability of interface technologies and automation options, and the physical and cognitive capabilities of the operator. This summarizes the capabilities of the current version of W/INDEX, as it has been used on a variety of programs. At the end of this paper, we will describe how W/INDEX is being modified to meet new sets of challenges while improving the usability, accuracy, and reliability of the tool. The next section describes the theory and use of the conflict matrix.

CONFLICT MATRIX

The interface conflict matrix is one of the major features that separates W/INDEX from other timeline workload analysis tools. In the matrix, each interface channel is paired off with itself and all the other channels, and for each combination, a conflict factor from 0 to 1 is assigned.

This approach recognizes that some types of attention combinations are more difficult to timeshare than others. For example, it is easy to speak and drive at the same time, but difficult to speak and comprehend speech at the same time. Similarly, it is easier to avoid traffic and monitor an oncoming stoplight simultaneously than to avoid traffic while adjusting the radio frequency. Both task combinations require simultaneous visual attention, but the latter one requires more divided attention.

To guide the assignment of these conflict values, W/INDEX uses a model of multiple attentional resources developed by Dr. Christopher Wickens and researchers at the University of Illinois. The basic idea of this model is illustrated in Figure 4. It represents a space in which task demands can be placed according to whether they are verbal or spatial, their input form visual or auditory, and their output form manual or vocal. The closer together two tasks or interface channels are in this space, the more they draw on the same attentional resources and, therefore, the more difficult they are to timeshare.

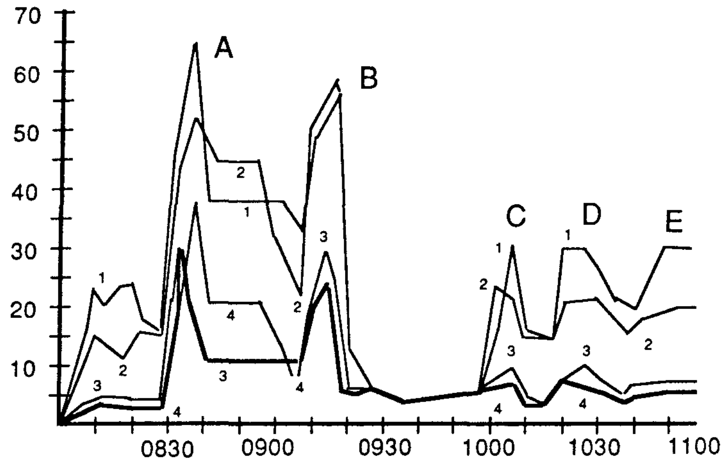

Figure 3. A Macintosh-Drawn Compilation of Four W/INDEX Workload Profiles. Each of the four profiles represents a different combination of automation and interface technology options.

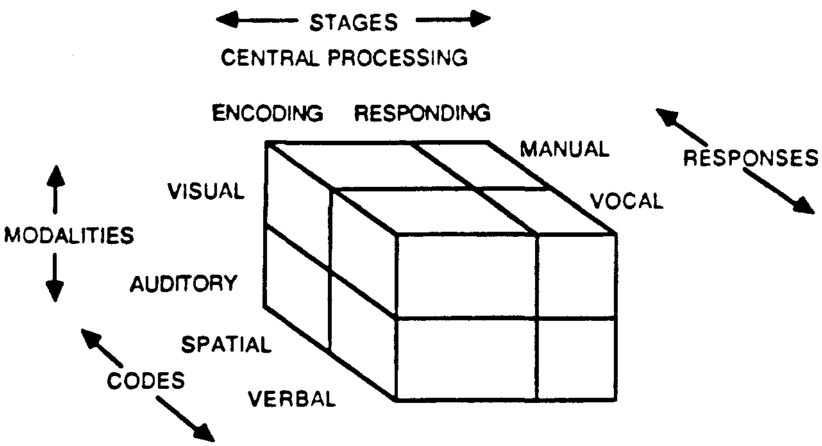

Figure 4. A Multiple Resources Model of Human Attentional Allocation

Figure 5 shows how this concept is used to assign conflict ranges to categories of interface channel combinations. Note that visual/visual conflicts are much higher than visual/speech conflicts. The analyst may then adjust the conflict value within the range given to account for physical separation or integration of controls and/or displays.

One of the primary strategies for reducing workload peaks, when they are produced, is to reduce the impacts of interchannel conflicts by substituting new interface channels for critical tasks. For example, a pilot required to operate a front panel switch while in air-to-air combat may experience an unmanageable demand conflict. If the switch task can be replaced by a speech command, the conflict is substantially reduced, and the result is lower workload.

APPLICATIONS

Since its inception in 1983, W/INDEX has been applied to a wide range of systems and questions. We used it to evaluate early concepts for the Army's LHX attack scout helicopter, then to evaluate degraded operations and pilot-vehicle interface alternatives later in LHX design. We also used W/INDEX to explore issues of one- versus two-man crews for the Apache advanced helicopter.

In 1986, we evaluated an Advanced Tactical Fighter (ATF) design against representative mission scenarios, and produced recommendations for changes in crewstation layout as well as automation and crew interface options.

Our latest effort looked at the feasibility of a crewstation design for the National Aerospace Plane (NASP). The purpose of this analysis was to compare workload predicted for the NASP cockpit with that for an existing transport, and thereby to estimate the feasibility of the proposed crewstation design and and task allocation. This analysis was able to address questions posed by the Air Force and NASA customers that were not approachable through any other means.

FREQUENTLY ASKED QUESTIONS ABOUT W/INDEX

One of the major issues accompanying any effort to predict workload is the degree to which the analysis can predict actual workload in the real world. This has proven to be an elusive goal, primarily because workload measurement itself is an inexact science even under the best experimental conditions. Several studies have been performed (Casali and Wierwille, 1983; Wierwille and Connor, 1983; and Wierwille, Rahimi, and Casali, 1985) to compare the accuracy of about fifteen different methods of measuring real workload during the performance of experimental tasks, including subjective ratings by the subjects, physiological measures, and dual task measures. Of these, only the subjective subject ratings have proven reliable in all conditions tested.

Furthermore, workload capacity can vary widely between individuals and even within individuals at different times. For all these reasons, it has not been possible to calibrate W/INDEX to a reliable measure of real workload, nor has it been possible to establish some predicted score as a reliable upper limit above which workload is unmanageable.

The use of subjective judgments of attentional demand raises questions about inter-rater reliability, since different analysts are likely to assign different ratings to these demand levels. Fatigue, which W/INDEX does not consider, is also an issue, as is the realism of conceptual mission timelines.

Honeywell is currently taking steps to address each of these issues. Table 1 shows how our current efforts relate to each concern raised above. Some of these approaches will be detailed in the next section. However, it is important to remember that the real goal of a W/INDEX analysis is to determine the best crewstation configuration, not to prove the ultimate feasibility of the design concept. As long as the analysis for a given design is internally consistent and appropriately representative and exhaustive, it should result in finding that best design solution, which is the tool's primary goal.

Task "A" Resources	Response	Task "B" Resources: Visual	Auditory	Manual	Verbal
	Visual	HIGH CONFLICT (.7-.9) Directly competing resources (e.g., two search tasks; less if tasks are adjacent or on same display areas.			
	Auditory	LOW CONFLICT (.2-.4) Noncompeting resources (e.g., search and listening).	HIGH CONFLICT (.7-.9) Highly competitive resources; some time-sharing if discriminability between inputs is high.		
	Manual	LOW CONFLICT (.1-.3) Noncompeting resources.	LOW CONFLICT (.1-.3) Noncompeting resources.	HIGH CONFLICT (.7-.9) Competing resources such as two tracking tasks or discrete choice tasks have shown high dual-task decrements.	
	Verbal	LOW CONFLICT (.1-.3) Noncompeting resources.	MEDIUM CONFLICT (.4-.6) More interfering if task requires voiced output.	LOW CONFLICT (.2-.4) Nonoverlapping resources showing little dual-task decrement in studies of tracking and voice input.	HIGH CONFLICT (1.0) Requires complete serial output; e.g., giving two messages or voice commands.

Figure 5. Prototype Conflict Matrix, used to Provide Conflict Levels Based on Resources Category.

W/INDEX PLANNED IMPROVEMENTS

Honeywell is currently working to improve W/INDEX in several areas. These efforts should provide better resolution to W/INDEX analyses while enabling subject matter experts to use W/INDEX without needing special knowledge of the tool or of the psychological theories behind it.

TABLE 1. RELATION OF W/INDEX IMPROVEMENTS TO WORKLOAD ANALYSIS ISSUES

Issue	Approach
Predictive Accuracy	Include consideration of multiple task dimensions that may affect workload
Cognitive Issues	Extend Multiple Resources Model to cognitive domain and verify inclusion of variable cognitive conflicts
Inter-rater Reliability	Automate database production by using question-answer dialogue
Timeline Realism	Connection with SAINT to generate and modify task schedules
Role in Design	Integrate W/INDEX with set of existing and emerging design tools to provide integrated analysis capability

W/INDEX's task difficulty rating method is being improved by incorporating a cognitive taxonomy into its structure that provides more detailed consideration of the variable conflict levels that may arise between tasks due to their cognitive differences or similarities. This extends the Multiple Resources Model into the strictly cognitive domain.

Furthermore, the next generation of W/INDEX will facilitate systematic consideration of the numerous task and interface channel characteristics that can contribute to workload. By taking more factors into account, W/INDEX should be able to represent specific design issues, such as display fidelity and fixation requirements, thereby improving its specificity.

Since the consideration of these factors will be systematic, the next generation of W/INDEX will be automated. Analysts will interact with the tool using a question-and-answer dialog, and W/INDEX will automatically construct the Interface/Activity and Conflict matrices based on the analyst's responses. Since the analyst will not have to directly insert values into these matrices, no special knowledge of the tool will be required to use it. Furthermore, systematic weighting of the various factors that contribute to workload will improve the tool's reliability between analysts and between analyses.

Finally, W/INDEX will be integrated into a systematic process Honeywell is developing to address a wide range of system design issues throughout the design process. This integrated approach will link tools, both Honeywell-developed tools and tools already available from other sources, so that the data flow from one tool to another follows the system development process. These tools will permit designers to elicit and prioritize function allocation issues and tradeoffs, predict the effects of automation on system performance, develop appropriate teaming strategies and information networks, develop mission profiles for individual operators, and determine individual operator workload levels.

The first step to integrating W/INDEX into this process has been to link the output from the Systems Analysis of Integrated Networks of Tasks (SAINT) tool developed by the Air Force to W/INDEX. SAINT considers the availabilities of operators in a team and schedules tasks for them so as to optimize the performance of the whole team network. Since each operator's task schedule constitutes a mission timeline for that operator, SAINT schedule output for a single operator can be fed to W/INDEX for workload estimation. Recently, we achieved this one-way linkage. Ultimately, W/INDEX and SAINT will interact, so that if undesirable workload levels result from one task schedule, the resulting peak periods can be sent back to SAINT for rescheduling.

CONCLUSION

We have described Honeywell's Workload Index (W/INDEX) crewstation design tool, both as it currently exists and as we foresee its future development. The currently available tool has been applied to a range of crewstation design problems with good results.

REFERENCES

Casali, J. G., and Wierwille, W. W. (1983) "A Comparison of Rating Scale, Secondary-Task, Physiological, and Primary-Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load." Human Factors, Vol. 25, No. 6,, pp. 623-642.

Wickens, C. D. (1984) "The Multiple Resources Model of Human Performance: Implications for Display Design," in the AGARD/NATO Proceedings, Williamsburg, VA.

Wierwille, W. W., and Connor, S. A. (1983) "Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator." Human Factors, Vol. 25, No. 1, pp. 1-16.

Wierwille, W. W., Rahimi, M., and Casali, J. G. (1985) "Evaluation of 16 Measures of Mental Workload Using a Simulated Flight Task Emphasizing Mediational Activity." Human Factors, Vol. 27, No. 5.

Applications of Human Performance Models to System Design

Edited by

GRANT R. McMILLAN
Harry G. Armstrong Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio

DAVID BEEVIS
Defence and Civil Institute of Environmental Medicine
Downsview, Ontario, Canada

EDUARDO SALAS
Naval Training Systems Center
Orlando, Florida

MICHAEL H. STRUB
U.S. Army Research Institute for the Behavioral and Social Sciences
Fort Bliss, Texas

ROBERT SUTTON
Royal Naval Engineering College
Manadon, Plymouth, Devon, United Kingdom

and

LEO VAN BREDA
TNO Institute for Perception
Soesterberg, The Netherlands

(1988)

PLENUM PRESS • NEW YORK AND LONDON
Published in cooperation with NATO Defense Research Group