

WORKLOAD PREDICTION, DIAGNOSIS, AND
CONTINUING CHALLENGES*

Donald L. Parks and
George P. Boucek, Jr.

The Boeing Company

INTRODUCTION

Many methods and models have evolved in recent years with the objective to improve our ability to predict and measure workload, based on the desire to assure operators can perform all tasks as required. With this evolution has come a better recognition of differences in need, language has become more precisely defined and there now is less confusion of concepts and purpose than 10 years ago. There is now a wide assortment of workload "tools" with widely varying degrees of complexity. Additionally, more attention is being given to whether the tools measure what they purport to measure, and whether the variety of tools now proposed actually do measure the same thing.

However, there is need for an increasing recognition of the designer's requirement to use workload evaluation results quite early for design diagnostics...to identify specific interfaces for attention in a given design approach, even in preliminary design. Even the earliest decisions are often too far reaching in impact to neglect an early appraisal...the simplest of time lines have been known to expose significant system problems. Furthermore, early and effective design integration of solutions is becoming more critical as systems become more and more complex, in order to avoid major change late in the design cycle for seemingly simple issues. Many of the tools that have emerged are unable to support this need.

Finally, a new issue is emerging as our ability to automate system operations improves. We need methods to appraise and control the consequences of underload as well as overload, and to produce accepted definitions of the upper and lower load limits - the "red lines" for marginal crew performance conditions.

* Originally released as Boeing Document D180-31116-1

Thus, although the reasons have changed, early and quantitatively defendable predictions of crew workload remains one of the most important issues for crew station design and development. In the past, the main interest was to avoid hardware changes late in the design-development cycle. Now, we keep hearing that it is "only" software. But consider the software nightmare that we face. There will be major issues to resolve in system and information integration, information management and display-control formatting, and in turn major workload questions as we attack the problem of comprehending and integrating: (a) an explosion of information associated with more sophisticated systems (b) a complex information network, and (c) complicated information integration display requirements. We need early simple-to-apply tools for this phase that set us up to transition to use of other tools as appropriate.

This paper reports the results of a continuing effort in the use of a timeline technique for workload prediction. It summarizes an evolution of our approach to timeline analysis and prediction (TLAP), and diagnostic applications. It presents methods to isolate the causes of high demand load conditions, including cause factors in peak load conditions and demand loads imposed by operation of each individual subsystem. It includes an approach to appraising cognitive workload as done in the past, and one using a new analytic technique. Efforts to demonstrate validity have been conducted and validation data are presented. Also, data is presented that demonstrates the degree to which results from use of the other tools correlate with this technique.

BACKGROUND AND PROBLEM SUMMARY

In cockpit design, we continue to emphasize the need for and use of an analytic predictive tool as a metric for workload evaluation. This is a critical aspect of our need to appraise workload before costly design, subsystem developments and cockpit integration have been committed. The cost of changes late in development is a major issue, but it is not until then that system development has progressed to the point where the context of the scenario can be realistically simulated for use of many of the other techniques now being proposed. Accordingly, we want to be able to forecast the effects of given design concepts and resolve related issues before spending extensive time and resources to design, develop and test such concepts.

In order to accomplish such objectives, typically, we want to compare new design concepts with existing designs in order to confirm that they are equal or better in terms of crew workload demands. This is especially important if major changes in design concept are being considered. Results of analyses at this level will provide an overall appraisal of relative workload for a new design concept.

However, the new design may feature areas of excessive workload. In turn, then, one of our strongest needs is for workload diagnostic techniques. In recent literature, there has been a lot of attention given to overall workload prediction and measurement, but little heed has been paid to techniques for resolving workload problems, once found. If there is a workload problem, how can it be resolved? The ability to identify heavy contributors to out-of-bounds conditions remains one of the major issues for the whole field of workload techniques. If a technique does not help to identify and focus on specific issues for resolving workload problems, it is not very useful to the designer. You've told him "There's a needle in the haystack somewhere. Go find it." So now, he has to redesign his concepts and needs to focus where the higher payoffs are indicated. He thus needs to follow some systematic procedure to identify, reduce or eliminate the problem. The more quantitative and defendable, the better.

If the problem is a workload peak, it should be made readily possible to identify the specific procedures, tasks or subsystems involved, and in turn possible solutions in the form of reallocation, redesign or both.

The point is, by the time a new airplane gets to the point where mission simulations and evaluations required for realistic use of many of the new tools can be made with a high level of confidence, a great number of design decisions, developments and subcontracts will have been made that go far beyond the cockpit in their impact. In fact, a great deal of interdisciplinary integration is involved in modern vehicles; a few of the areas impacted are sensors, avionics systems, details of subsystem design and levels of automation. As design and development progresses, the degrees of freedom for the cockpit designer become less and less until a late change of any consequence has a far reaching impact that is unacceptable. So we want to commence a new system with tools that can be used from the very early stages and proceed through design and development with increasing confidence that workload is manageable over the wide variety of circumstances that are of concern and will be demonstrated to be so in acceptance and certification testing.

We face, potentially, an explosion of information management problems, with greater attention to automation concepts. Such issues will become much more complex in the near future and workload management on our part will become more difficult. It is unlikely that we will change our requirements for pilot situation awareness and control, in order to assure he can monitor status and provide active and backup modes of operation as required. However, with new systems information management and flight management requirements, we can expect more extensive use of automation, expert system and artificial intelligence techniques as well as new approaches to information integration and display formatting in cockpit management. Adaptability and/or compatibility of our methodology with this part of the future should become part of our criteria for the workload tools of today.

More recently, another question is becoming relevant that was of little concern in the past. In the past, we have emphasized the need to minimize workload. With the old dials and gages, combined with limited technological ability to automate, we wanted to minimize workload. Now that we can better integrate and automate the display-control design, there is an increasing concern about the possibility of too little workload. Complacency and boredom could become increasingly important factors in cockpit operation. Accordingly, our metrics must feature the capability to establish upper and lower bounds for workload. Workload management to assure that pilot can readily accommodate surprises and that he will remain alert in a heavily automated system is a serious issue for the near future.

By now it should be clear that we feel a strong need for predictive, diagnostic and viable analytic tools. We also have a continuing interest in empirical performance measurement techniques. The reasons are two fold: (1) Our predictions must eventually be demonstrated in test. (2) There should be a correlation between empirical techniques and the predictive approach. If high correlations exist, predictive design and development tools can be used with greater confidence that they lead us down the right path.

TECHNICAL DISCUSSION

This section of this report presents an approach for timeline analysis and prediction (TLAP) of workload, and for diagnostic applications of the technique. The approach is based on time required to perform a task

vs. time available within the task sequence. To clarify any misconceptions from past presentations, a step by step evolution of the techniques will be presented, with results as appropriate. Approaches for using this tool for workload diagnostics will be described. Techniques that have been applied to estimate the more elusive parameter---cognition---will also be discussed. Finally, data representing how robust this technique is for its purpose will be presented, including validation data and the degree to which other techniques have been found to correlate with it. Finally, discussion will address selected technological challenges that remain to be exploited.

Timeline Analysis, Prediction and Diagnosis

Timeline Analysis as an approach to workload evaluation started many years ago as a simple, manually developed layout of a task sequence on a timeline. This was then appraised for whether tasks could be performed within the time available, for stimulus-response compatibility and disparity, for any equipment performance characteristics that could impact completion of a given task series and for the ability to complete the task in time.

This approach to workload analysis has now progressed to a far more sophisticated level. While original features are still evident, data details and processing methodology has changed and application features have grown. Accordingly, and in order to minimize misconceptions about the context herein, an elaboration of the present approach and how it is used will be presented.

The first step in evolving the timeline based workload analysis is to develop a generic mission scenario for a typical flight, including major worst-case conditions. The scenario defines flight segments, altitudes, speeds, key events, typical operations requirements and both normal and degraded modes of operation. Constraints of operation, environmental variations, timing considerations and particularly critical situations are all outlined, and representative variations in the mission are also identified and injected into this generic scenario. The key objective is to develop a generic mission definition that encompasses all conditions and thus avoid oversights. The resulting scenario may be brief, serving as the baseline for an extensive task definition. Alternatively, it may be very extensive and detailed in which case it provides much of the structure for systems selection and detailing of the task definition.

In turn, the scenario is used to develop a detailed task definition and sequence. Given a scenario and a candidate cockpit configuration, operations procedures and specific display-control operation requirements are identified. Most typically, this is accomplished for both a baseline airplane with existing display-control concepts as a point of departure and for the new cockpit concept under development. This permits appraisal of workload on a relative rather than an absolute basis, for a higher level of confidence in the result.

Personal experience has demonstrated that the most convenient way to organize the task requirements is to construct a timeline that relates the display-control operations to the scenario and its key events as illustrated in figure 1. This can and has been done at varying levels of detailed definition depending on the level of cockpit definition. The main issue is to define when tasks have to occur, and when related sequencing must start in order to complete on time. This format is relatively easy for an analyst to develop and can become increasingly detailed as task analysis continues. It also provides a convenient outline for defining and inputting detailed data into the computer, as well as a checkpoint of the computer results against input data.

UNSHIFTED

MISSION - FROM TETILA FIVE
 TEST CASE
 CONFIGURATION - CONFIG. A
 FLIGHT PHASE - ENG START AND TAXI
 CREWMEMBER - CAPTAIN

JUN 1, 1987

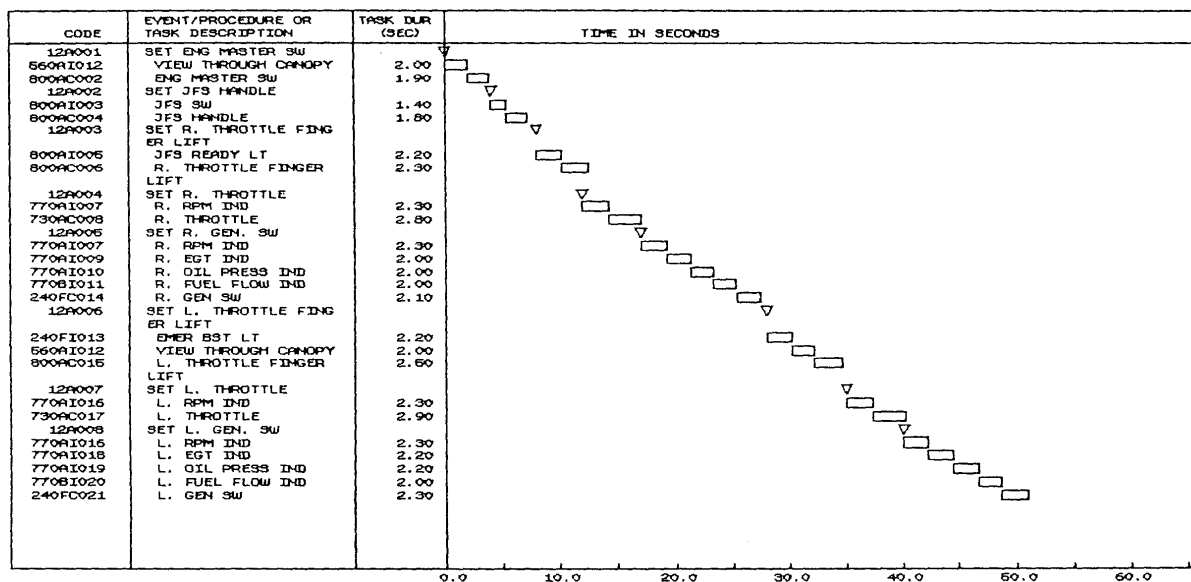


Figure 1. Timeline Definition Format

Other task characteristics that might be relevant can also be treated. These tasks that must be performed at a critical time can be identified. Those that can be slipped for earlier or later performance can also be defined. If tasks must be performed in a given sequence, this too can be specified. Other variations are feasible; many have been explored over the years, including task performance probabilities and task performance time variations (e.g., standard deviations).

The timeline is then further refined. Our development of timeline analysis recognizes that humans can do more than one thing at a time. For example, in driving a car, one processes visual and auditory information, makes decisions, steers, and operates the accelerator, brake (and clutch)---discrete, parallel, sequential and coordinated tasks are performed as an integrated network of responses. We use a human performance channel allocation scheme to handle such variations. This involves defining which aspects of a task are performed by vision (internal, external), hands (right, left), feet (right, left), audition (hearing, speech) and cognition. (This part can be automated if we have a data bank of task performance times and a definition of display-control task demands. For first use, derivation is manual. For initial application, the sources of data for estimating task performance times include information scattered through the literature, time and motion methods, appraisals in mockups, and simulation. Once in the data base, the task requirement is catalogued for future use.)

Most performance times are straight forward. However, cognitive workload has been another issue. Analyst estimates have been used by some analysts, but more quantitative approaches exist. One evolved from a flight test program some years ago wherein a consistent quirk was indicated in eye movement data. Pilots would look at a display, then there would be a pause before they took action. The action could be clearly related to the presented information, suggesting that processing was involved. This processing time came to be accepted as demonstrating cognitive operations. Furthermore, the pause time was essentially a constant percentage of task performance time. This relationship became the basis for a more quantitative estimate of cognitive workload in our approach. More recently, another estimating technique based on information theory has been developed. This latter technique will be discussed in a later portion of this paper.

At this point, data is ready for final computation and processing. By inspection, it can be determined if any task conflicts exist, in terms of conflicting or incompatible demands (like rubbing the head and patting the stomach at the same time). Processing through the computer produces a series of plots and reports for appraising workload demand and for detecting and diagnosing problem areas.

Workload estimates are produced by solving the equation:

$$W/L = \frac{R}{T_A} = \frac{\text{Time Required}}{\text{Time Available}} = xx\%$$

over small increments of time (in order to avoid a leveling effect). Processing produces a percent workload figure which can be plotted over time to produce a workload time history, as is illustrated in figure 2.

UNSHIFTED
JUN 1, 1987

WORKLOAD HISTOGRAM
CREWMEMBER- CAPTAIN
CHANNEL- RIGHT HAND
CONFIGURATION- CONFIG. A

MISSION
TEST CASE

FLIGHT PHASE
ENG START AND TAXI

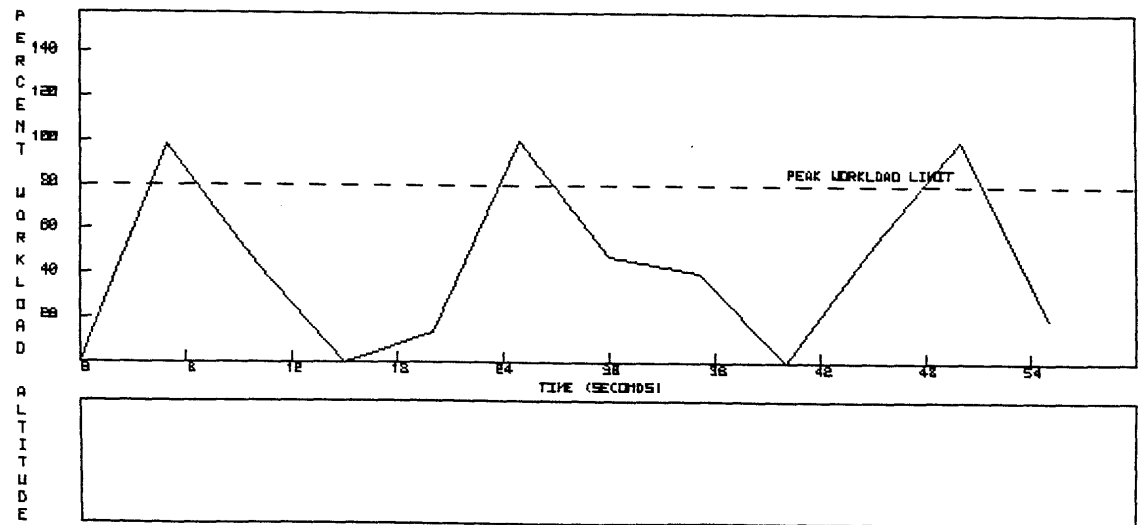


Figure 2. Example of Workload Time History Profile as Produced by TLAP

Referring to figure 2, an limit factor can be set in that is used as a cut-off for peak workload considerations. In between this case, the cut-off is 80%, which is that level of time demand where pilots have been observed to start dropping tasks in order to continue performing at a self imposed acceptable level, perhaps to allow for between-task transition time. For purposes of this paper, the 80% time might be considered to be the fully loaded condition.

As peaks exceeding the 80% time occur (or whatever limit might be chosen), diagnostics to determine the cause of the workload excesses are in order. There are several alternatives available to the analyst for this purpose.

- o First, the task timeline (figure 1) for the particular time period can be inspected in detail for those tasks and events that cause the peaks. Analysis of attendant situations and conditions will provide necessary insights as to cause factors and alternative approaches.
- o Second, a task sliding feature might be applied to determine if an artifact is present---if an analytic allocation requires performance at unduly stringent times and thus would be averted by a pilot in a real life condition. Some discretion may be desirable in interpretation here, to assume that the task sliding feature is reasonable.
- o Third, a subsystem utilization feature can be used. Since the task time line involves given subsystems, the computer can accumulate the record of subsystem interface operations to produce a subsystems "time demand" summary, and can rank order the demand summary to facilitate diagnostics. This feature is illustrated in figure 3. The figure clearly illustrates where the heaviest demand loads exist and where the highest payoff could be attained in terms of design changes. Of course, it may be as convenient to automate some monotonous and less critical chore that applies throughout the mission (such as subsystem monitoring) to lower the overall workload level.

Other characteristics may also be of interest, such as the degree of variability in demands placed on a channel---it may be desirable to narrow the range of variability in demand for some channel in order to control a widely varying demand. Figure 4 illustrates a one sigma estimate of demand variability for each performance capability group for a mission segment. Casual inspection suggests there might be an advantage to reappraising the visual task allocations and making some adjustments in design. Admittedly, this is seldom feasible in practice, but the diagnostic implications are evident.

In overview, then, each of the diagnostic techniques offers a straight forward approach to supportable recommendations for changing procedures or design, or even for reallocating the tasks assigned to the crewmen.

Cognition

One of the most difficult components of workload to include in any analytic assessment technique is the cognitive component. Most early analytic techniques either ignored the cognitive factor and concentrated on the physical parameters or considered cognition as an all or none component. Newer techniques attempted to get at the variable nature of cognition.

UNSHOFTED

SUBSYSTEM ACTIVITY SUMMATION BARGRAPH
MISSION - TEST CASE
CONFIGURATION - CONFIG. A
FLIGHT PHASE - ENG START AND TAXI
CREWMEMBER - CAPTAIN

JUN 1, 1987

CHANNEL - ALL

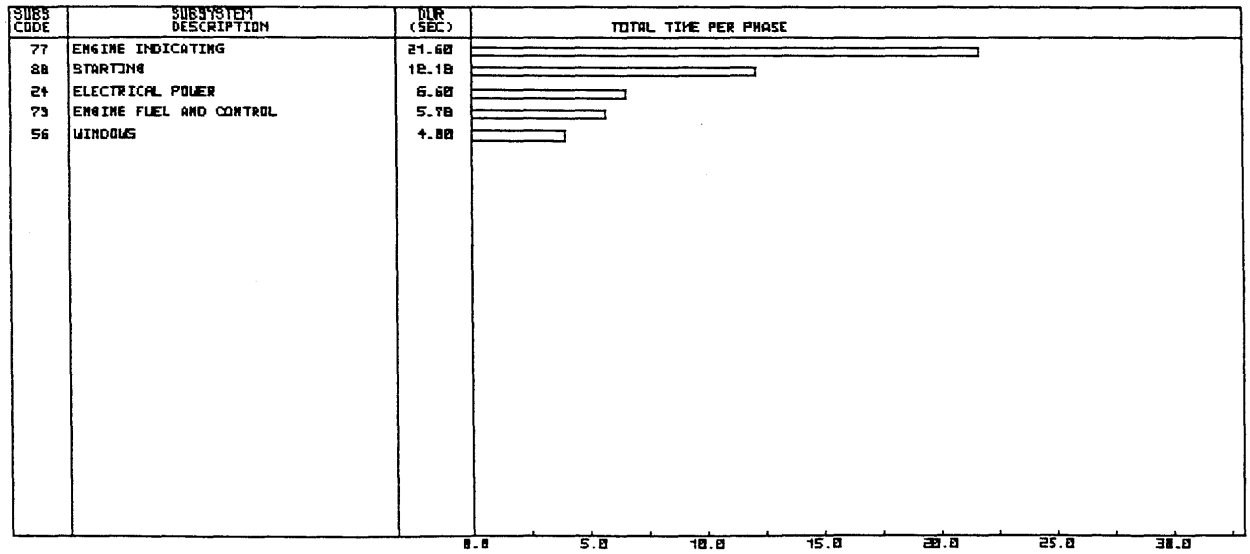
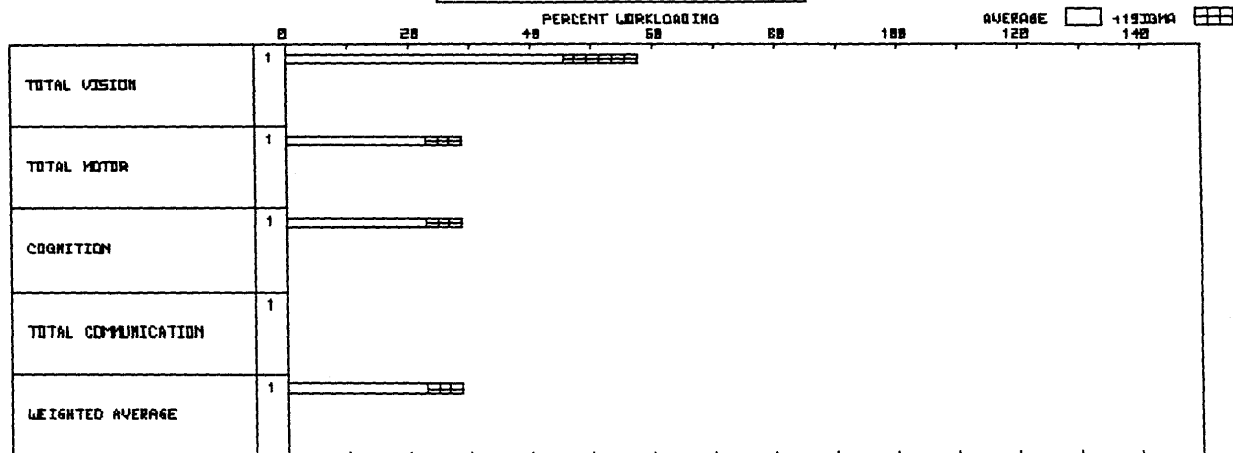


Figure 3. Subsystem Utilization Summary Produced by TLAP

UNSHIFTED

WORKLOAD SUMMARY
CREWMEMBER - CAPTAIN

JUN 1, 1927



MISSION
TEST CASE

CONFIGURATION
CONFIG. A

FLIGHT PHASE
ENG START AND TAXI

Figure 4. Summary of Channel Parameter Loads....Demand Average and Standard Deviation

One method used within Boeing and currently being used for a joint USAF/FAA program investigating workload assessment techniques uses a device complexity score as a basis for estimating cognitive workload. The device complexity score is based upon the information content of the possible states that the device presents; it is derived using the concepts of Information Theory. Information has been defined as the number of binary digits (BITS) into which an event can be encoded depending upon the number of alternatives it presents (figure 5). In this context, each display (i.e. gage, dial, flight display, etc.) is encoded based on the numerical units that have to be used. Each control is also encoded depending upon the number of choices (alternatives) that it enables. Procedure complexity is then equal to the sum of device complexity scores for all steps of a procedure. Summing device complexity scores could thus produce a figure representing total subsystem control complexity, i.e.,

$$\text{complexity} = \sum_i b_i$$

where b_i is the individual complexity score

$$\text{and cognition} = \sum_i B_i$$

where B_i is the corresponding cognitive workload score in seconds, from the equation

$$B_i = 0.27216 + 0.12456 b_i \text{ if } b_i \leq 8 \text{ bits, or}$$

$$B_i = 0.27216 + 0.22968 b_i \text{ if } b_i > 8 \text{ bits}$$

This technique includes information processing in the workload analysis as a major component of cognitive work. The advantages of this type of approach are that it: provides a common measure for relating the workload associated with different control and display devices; helps avoid the limitations associated with simply summing the numbers of devices; permits numerical evaluations to be made between different methods of interfacing a system or procedure with the crew. The cognition score for the timeline analysis is based upon the procedure complexity score and the communication channel time.

Timeline Robustness: Validation and Reliability Considerations

Although it is convenient to think of people as experiencing a similar level of workload in response to a set of fixed task demands (e.g. an average workload level), "constant" workload levels are not necessarily constant due to the individual nature of each person's actions. Certainly, many questions could be asked about the magnitude and influence of individual differences; these are typically resolved by allowing for variability in analyses and confirming the allowance in simulation. However, the current objective of the assessment task is to detect differences in workload levels, identify excessive peak loads and proceed into diagnosis and resolution. Manufacturers also need to assure that techniques to be used reliably discriminate between high and low levels of workload in order to make general conclusions. A brief discussion of status regarding some of the key questions is in order.

There are many types of validity, each affecting the ultimate usefulness and acceptability of the analytic technique. The questions to be answered are whether the timeline analysis technique really predicts what it says it predicts and how does its output relate to the actual workload being experienced in the procedures being followed by the crew? The most direct way to establish the answer to the first question is to perform the timeline analysis and then validate it by taking data in a

flight test program. This was the method of choice for the Boeing 757 and 767 airplane programs. The results from these test programs indicate that the timeline analysis can accurately predict (with greater than 90 percent accuracy) the flight data for all phases of flight. Thus the technique is predicting what it says it is predicting. The second question is much more difficult to answer; discussion follows.

The Boeing Company teamed with Douglas Aircraft Company to work a program sponsored jointly by the U.S. Air Force and the Federal Aviation Administration to evaluate workload assessment techniques (Boucek, et al, 1987). One of the comparisons that was made in this study was the correlation between the predicted task loading (generated by timeline analysis) and the results of objective (physiological measures and secondary task performance) and subjective (NASA's TLX and USAF SWAT) measurement techniques. The purpose of this correlation analysis was to evaluate how much the workload scores overlap. If workload assessment techniques have been shown to be valid and reliable and they do not correlate with each other, then they are thought to be assessing different aspects of workload.

The results of this analysis show that the timeline data has not only consistent internal correlations but also interpretable correlations with the other assessment data (see figure 6). The results of the internal TLAP comparisons indicated that the visual channel data correlated well with everything (manual left .74, manual right .64, auditory .55, and cognitive .65) except the verbal channel. The manual channel data did not correlate with any other scale except the visual channel. Verbal and auditory channels were highly correlated (.78). Finally, the cognitive channel was highly correlated with the visual (.65), auditory (.85) and verbal (.65) channels.

The results of the comparison of the timeline data with the data from the other assessment techniques indicate that: the average inter-beat interval for the heart was correlated with the visual (-.55), manual left (-.64) and manual right (-.56) channels; the standard deviation for the interbeat interval correlated with the right manual (.54) channel. The Mulder Spectral Analysis Blood Pressure component correlates with the visual (.91) and the cognitive (.66) components. Wheel and column inputs correlate with the manual right (.55, .60) and the pedal inputs with the manual left (.59). Both of the subjective scales correlate with the verbal channel (SWAT .51, TLX .63). Finally, the accuracy of the response to the probe in the secondary task correlates with the manual right (.51), auditory (.51) and cognitive (.55).

SUMMARY AND CONCLUSIONS

This paper presents a progress report on use of Timeline Analysis and Prediction (TLAP) methods for workload analysis and diagnosis. It addresses questions that have been posed for the timeline approach to workload and demonstrates the credibility of the approach.

The report includes an outline of the approach and rationale to using timeline analysis as a cockpit development tool. The tool is adaptable in that it can be used in stages factored to the level of system definition. It can progress from a relatively simple level early in cockpit design when preliminary estimates might be all that are available, to very sophisticated levels with detailed design. A step by step procedure was described and methods of use were discussed. Approaches for incorporating cognition in a timeline framework were presented for further use and exploration. Validation experience was summarized, as was the extent to which other techniques correlated with applications of the tool.

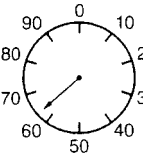
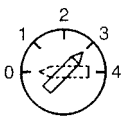
Definition	Examples
Device complexity is the number of binary digits required to encode the possible number of alternatives associated with the device	 Instrument Reading Number of Alternatives = $\frac{\text{Range}}{(0.5)(\text{Scale Unit})}$ Device Complexity = $\log_2 \frac{10}{(0.5)(1)} = 4.35 \text{ bits}$ Rotary Selector Switch  Discrete Control Number of Alternatives = 5 Device Complexity = $\log_2 (5) = 2.32 \text{ bits}$
Device Complexity = $\log_2 (\text{Number of Alternatives}) \text{ bits}$	

Figure 5. Timeline Analysis Device Complexity Measure

a)	Visual	Manleft	Manright	Verbal	Auditory	Cognitive
VISUAL	1.00	<u>.74</u>	<u>.64</u>	.10	<u>.55</u>	<u>.65</u>
MANLEFT	-	1.00	.28	-.01	.25	.42
MANRIGHT	-	-	1.00	.21	.38	.49
VERBAL	-	-	-	1.00	<u>.78</u>	<u>.65</u>
AUDITORY	-	-	-	-	1.00	<u>.85</u>
COGNITION	-	-	-	-	-	1.00

b)	Visual	Manleft	Manright	Verbal	Auditory	Cognitive
HRM	<u>-.55</u>	<u>-.64</u>	<u>-.56</u>	.17	.14	-.02
HRSD	.14	.10	<u>.54</u>	-.13	-.13	-.31
APB	<u>-.91</u>	<u>-.87</u>	-.39	-.04	-.44	<u>-.66</u>
ARS	-.05	.12	.32	.11	-.08	-.26
EBK	-.11	-.03	.11	-.39	-.47	-.42
WHL	.34	.36	<u>.55</u>	.09	.15	-.01
STK	.38	.43	<u>.60</u>	-.13	-.07	-.12
PDL	<u>.54</u>	<u>.59</u>	.39	-.47	-.30	-.04
SWAT	.04	.05	.49	<u>.51</u>	.21	.20
TLX	-.02	.03	.40	<u>.63</u>	.28	.25
STRT	-.39	-.41	.14	.03	-.13	-.32
STRT%	.48	.19	<u>.51</u>	.31	<u>.51</u>	<u>.55</u>

Correlations greater $\pm .50$ are in bold.

Legend	Task Channels	Legend	Task Channels
Visual	Eyes	Verbal	Spoken Communication
Manleft	Manual Left Hand Tasks	Auditory	Listening
Manright	Manual Right Hand Tasks	Cognitive	Cognitive Tasks

Legend	
HRM	Average Inter-beat Interval
HRSD	Standard Deviation for the Average Inter-beat Interval
APB	Mulder Spectral Analysis Blood Pressure Component from Heart Rate
ARS	Mulder Spectral Analysis Respiration Component from Heart Rate
EBK	Eye Blinks per Minute
WHL	Wheel Control Inputs per Minute
STK	Stick Control Inputs per Minute
PDL	Pedal Control Inputs per Minute
SWAT	Subjective Workload Assessment Technique
TLX	NASA Task Load Index
STRT	Secondary Task Reaction Time
STRT%	Probe Accuracy to Positive Probes for the Secondary Task

Figure 6. Correlation Matrices for the TLAP Variables:
(a) Internally, and (b) with Other Methods

Both intra-tool correlations and correlations of results from applying other tools in the same context were encouraging. Magnitudes and trends were sufficiently consistent to warrant continued refinement and use of this tool. There was a very high degree of predictive accuracy compared to actual flight data and a respectable set of correlations between TLAP predictions and measurements using interbeat heartrate interval, the Mulder Blood Pressure Spectral Analysis, SWAT, and secondary task techniques.

In overview, then, continued research and development has led to beneficial refinements in the timeline approach to workload analysis, and demonstrated that the method is quite robust. It is a realistic, useful tool that has a large and growing experience base. It is useful early in design when data may be fragmentary and is easily adaptable as system changes occur. Continued evolution and refinement is warranted.

However, foreseeable changes in our work-a-day requirements will require that this and all other techniques be continually reexamined in the context of fast changing technology and issues. It will become more and more important for all techniques to be adaptable to this new environment in order to continue to produce diagnostic workload information. Experience to date with TLAP indicates that it has this capability.

Acknowledgement

This paper summarizes many years of progress on the timeline based approach to workload, as can easily be inferred from perusal of the bibliography. In addition to continuing company sponsorship, the most recent activity (Boucek, et al, 1987) was sponsored by USAF's Flight Dynamics Laboratory and the FAA, under the guidance of H. G. Britten-Austin, USAF's AFWAL/FIGR and P. K. Hwoschinsky, FAA's APS 450. Earlier efforts were sponsored by NASA, LRC (Miller, 1976), and USN, NADC (Parks and Springer, 1975).

REFERENCES

- Auffret, R. (ed.), 1977, Studies on Pilot Workload, AGARD Conference Proceedings No. 217, Papers presented at the Aerospace Medical Panel Specialists' Meeting, Koln, West Germany, April.
- Boucek, G. P., Jr., Sandry-Garza, D. L., and Logan, A. L., (Boeing) Biferno, M. A., Corwin, W. H. and Metalis, S., (Douglas), 1987, Proceedings of the Workshop on the Assessment of Crew Workload Measurement Methods, Techniques, and Procedures: Part Task Simulation Data Summary, AFWAL-TR-87-3103, Sept. 15-16.
- Boyd, S. P., 1983, Assessing The Validity of SWAT as a Workload Measurement Instrument, United States Air Force Academy. Proceedings of the Human Factors Society, 27th Annual Meeting, 124-128.
- Chiles, W. D., 1977, Objective Methods of Developing Indices of Pilot Workload, Civil Aeromedical Institute, Federal Aviation Administration, Oklahoma City, OK, July.
- Cooper, G. E. and Harper, R. P., Jr., 1969, The Use of Pilot Ratings in the Evaluation of Aircraft Handling Qualities, Moffett Field, CA: NASA Ames Research Center, NASA TN D-5153.
- Eggemeier, F. T., Crabtree, M. S., and Reid, G. B., 1982, Subjective Workload Assessment in a Memory Update Task, Proceedings of the Human Factors Society, 26th Annual Meeting.
- Eggleston, R. G. and Kulwicki, P. V., 1984, A Technology Forecasting and Assessment Method for Evaluating System Utility and Operator Workload, Presented at the 1984 Annual Meeting of the Human Factors Society.

- Fadden, D. M., 1982, Boeing Model 767 Flight Deck Workload Assessment Methodology, Paper presented at the SAE Guidance and Control System Meeting, Williamsburg, VA, November.
- Gerathewohl, S. J., 1977, Inflight Measurement of Pilot Workload, Panelists: E. L. Brown, Douglas Aircraft Co., J. E. Burke, Vought Corporation, K. A. Kimball, USAMRL, S. P. Stackhouse, Honeywell, Inc. and W. Long, Bell Helicopter Co., Aerospace Medical Association Annual Scientific Meeting, Las Vegas.
- Hart S. G. and Bortolussi, M. R., 1983, Pilot Errors as a Source of Workload, Paper presented at the Second Symposium on Aviation Psychology, Columbus, OH.
- Hart S. G., Childress, M. E., and Hauser, J. R., 1982, Individual Definitions of the Term Workload, Paper presented at the 1982 Psychology in the DOD Symposium.
- Hart S. G., and Staveland, L. E., In press, Development of a Multi-dimensional Workload Rating Scale: Results of Empirical and Theoretical Research, To appear in P. A. Hancock and N. Meshkati (Eds.), Human Mental Workload, Amsterdam: North Holland Press.
- Hay, G. C. House, C. D., and Sulzer, R. L., 1978, Summary Report of 1977-1978 Task Force on Crew Workload. Report No. FAA-EM-78-15, U. S. Department of Transportation, Washington, D. C., December.
- Linton, P. M., Jahns, D. W. and Chatelier, P. R., 1977, Operator Workload Assessment Model: An Evolution of a VF/VA-V/STOL System, AGARD-CPP-216, Aerospace Medical Panel Specialist Meeting, Koln, Germany, 18-22 April.
- McLucas, J. L., Drinkwater, F. J. and Leaf, H. W., 1981, Report of the President's Task Force on Aircraft Crew Complement, Douglas Aircraft Company.
- Miller, K. M., 1976, Timeline Analysis Program (TLA-1), Final Report, Boeing Document D6-42377-5, Prepared for National Aeronautics and Space Administration, Langley Research Center (NASA-CR-144942) April.
- Moray, N. (Ed.), 1979, Mental Workload: Its Theory and Measurement, Plenum Press, New York.
- Parks, D. L., 1978, Current Workload Methods and Emerging Challenges, in Mental Workload: Its Theory and Measurement, N. Moray (Ed.), New York: Plenum Press.
- Parks, D. L. and Springer, W. E., 1975, Human Factors Engineering Analytic Process Definition and Criterion Development for CAFES, Boeing Document D180-18750-1, Prepared for Naval Air Development Center, Contract N62269-74-C-0693, June.
- Parks, D. L. and Stern, P. H. and Niwa, J. S., 1965, Crew Number Study: Supporting Documentation for Advanced Manned Strategic Aircraft, Crew Factors Study, Volume III - Task Allocation Report; Boeing Document D6-16224-3, Prepared for U.S. Air Force Systems Command, Contract AF33(657)-15339, October.
- Reid, G. B., Shingledecker, C. A., Nygren, T. E. and Eggemeier, F. T., 1981, Development of Multidimensional Subjective Measures of Workload, Proceedings of the 1981 IEFF International Conference of Cybernetics and Society, 403-406.
- Roscoe, A. H., 1978, Assessing Pilot Workload, AGARD-AG 233, AD A051 587, Paris: NATO.
- Roscoe, A. H., 1984, Assessing Pilot Workload in Flight, Paper reprinted from Conference Proceedings No. 373 Flight Test Techniques. AGARD, NATO, Neuilly-sur-Seine, France, April.
- Ruggiero, F. T. and Fadden, D. M., 1987, Pilot Subjective Evaluation of Workload During a Flight Test Certification Program, Paper reprinted from Conference Proceedings No. 282, The Practical Assessment of Pilot Workload, AGARD, NATO.
- Shingledecker, C. A. and Crabtree, M. S., 1982, Standardized Test for the Evaluation and Classification of Workload Metrics, Proceedings of the Human Factors Society, 26th Annual Meeting.

- Sheridan, T. B. and Simpson, R. W., 1979, Toward the Definition and Measurement of the Mental Workload of Transport Pilots. Massachusetts Institute of Technology, January.
- Siegel, A. I. and Wolf, J. J., 1961, Techniques for Evaluating Operator Loading in Man-Machine Systems, Applied Psychological Services.
- Stone, G., Gulick, R. K. and Gabriel, R. F., 1985, Use of Task/Timeline Analysis to Assess Crew Workload, Douglas Paper 7592, Douglas Aircraft Company, Long Beach, CA.
- Sulzer, R., Cox, W. J., and Mohler, S. R., 1981, Flight Crewmember Workload Evaluation, DOT/FAA/RD-82/83.
- Wickens, C. D., 1980, The Structure of Attentional Resources, in Attention and Performance VIII, R. Nickerson and R. Pew (Eds.), Englewood Cliffs, New Jersey: Erlbaum.
- Wierwille, W. W., 1979, Physiological Measures of Aircrew Mental Workload, Human Factors, 21, 575-593.
- Wierwille, W. W., and Casali, J. G., 1983, A Validated Rating Scale for Global Mental Workload Measurement Applications. Proceedings of the Human Factors Society, 27th Annual Meeting.
- Weiner, J. S., 1982, The Measurement of Human Workload, Ergonomics, 25, 953-965.
- Whitley, L. C., and Vaughn, R. R., 1968, Man-Machine Stochastic Simulator TEN-708, MMSS Volume I, Boeing Document D6-29184-TN-1, April.
- Whitley, L. C., and Vaughn, R. R., 1968, Man-Machine Stochastic Simulator TEN-708, MMSS Volume II, Boeing Document D6-29184-TN-II, April.
- Whitmore, D. C. and Parks, D. L., 1974, Computer Aided Function-Allocation Evaluation System (CAFES), Phase IV, Final Report, Boeing Document D180-18433-1, Prepared for Naval Air Development Center, Johnsville, (Contract N62269-74-C-0274), December.

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