

Workload Modeling using Time Windows and Utilization in an Air Traffic Control Task

Pei-Ju Lee, Andreas Kolling, and Michael Lewis

School of Information Sciences

University of Pittsburgh

Pittsburgh, PA, 15260 U.S.A

pel30@pitt.edu, andreas.kolling@gmail.com, ml@sis.pitt.edu

In this paper, we show how to assess human workload for continuous tasks and describe how operator performance is affected by variations in break-work intervals and by different utilizations. A study was conducted examining the effects of different break-work intervals and utilization as a factor in a mental workload model. We investigated the impact of operator performance on operational error while performing continuous event-driven air traffic control tasks with multiple aircraft. To this end we have developed a simple air traffic control (ATC) model aimed at distributing breaks to form different configurations with the same utilization. The presented approach extends prior concepts of workload and utilization, which are based on a simple average utilization, and considers the specific patterns of break-work intervals.

INTRODUCTION

A substantial body of research has accumulated on the development and application of tools that can predict and/or measure human mental workload. This research demonstrates that human performance can be affected by a variety of factors, including fatigue and physical or mental stress (Carnahan et al., 2000). This mental stress, or cognitive workload, can degrade the performance of a system when the capacity of the operator is exceeded and the operator can no longer perform the task adequately or complete all its required components. Gopher et al. (1984) present the workload construct as an attempt to explain why human operators are unable to cope with the requirements of a task, and workload measures as the effort to characterize the performance of a task relative to the operator's capability.

While mental workload cannot be directly observed, it may be inferred from operator behavior, measured indirectly through its effects on system performance, psychological or physiological processes, elicited through subjective judgments or predicted from task parameters (Cain, 2007). The study of mental workload can be of scientific interest in its own right but most often has been used to gauge or compare the difficulty of tasks or as a basis for adapting system characteristics to maintain workload within the operator's capacity (Parasuraman 2010). These approaches to adaptive automation (AA), referred to as adjustable or adaptive autonomy in robotics, may be either open or closed loop. In closed-loop AA some physiological or task-related aspect of operator performance is monitored to provide a basis for increasing or decreasing automation to maintain workload within a desired range. In open-loop AA the operator's performance is not measured. Excessive workload is predicted, instead, from system parameters and adjustments made to maintain these parameters within predefined limits without reference to operator performance. A governor designed to keep a truck within a user controllable speed would be an example of open-loop AA.

Using classical methods the time course of workload over an experimental or work session is rarely observed. In some tasks such as tracking, the relation between track difficulty and performance (serving as a measure of workload) is followed continuously but for analysis is aggregated across the session. Subjective measures of necessity are for the entire session rather than instantaneous. While closed-loop AA monitors workload continuously the adjustments it makes in response contaminate the measure making it difficult to determine whether the effects of workload are accumulating or only in response to immediate demands. While this distinction makes little difference to researchers wanting to compare task difficulty or design closed-loop AA, it is important to designers of open-loop AA systems because they must predict rather than merely adapt to any accumulative component of workload.

Existing research into cumulative effects of work, has largely been concerned with physiological effects such as muscle stress and fatigue which can be used as a basis for designing job rotation schedules—but these schedules do not consider cognitive status as a factor (Paul et al, 1999, and Emmett et al, 2009). Emmett et al. (2009) reviewed a wide variety of studies using scheduling theory in human factors finding the vast majority to involve shift and work studies considering time scaled in hours or days. There is, however, a strand of human factors research that has been concerned with cumulative effects of work on a much finer time scale. Hendy et al. (1997) proposed a two dimensional description of workload based on intensity (how much) and time (how long). In Neerincx's (2003) three-dimensional control task load (CTL) model, the first factor is the percentage of time occupied determined through timeline assessment. The occupied time is also used to assess workload by Reed (1992) and Beevis (1996). Occupied time, referred to as *utilization* in queuing theory, has proven to be a natural way to assess human effort at tasks in which the operator is required to complete many brief tasks in sequence. It has been particularly popular for characterizing air traffic control (ATC) tasks in which difficulty may be equated with the arrival rate

and density of air craft (Schmidt 1978) and multi robot control where difficulty is related to the frequency and duration of interaction each robot requires (Crandall et. al. 2010).

Schmidt (1978) examined the maximum allowable workload thresholds to reallocate the operator's traffic-handling capacity through a queuing model. He used ATC simulations to quantify service time in order to examine the relation between the time required and the time available to accomplish tasks. His results showed that the average task delay, or the workload, rapidly increased as the percentage of the time the operator was busy (utilization) went from 0.65-0.75. Cummings and Guerlain (2007) found a similar result with performance decaying significantly when operators exceeded 70% utilization at a control task redirecting cruise missiles. The paper which aimed to develop better capacity predictions with varying numbers of control targets, workload level, and complexity of scenarios demonstrated the robustness of this threshold. The 70% utilization effect cropped back up in a more recent study (Cummings et al., 2010) in which operators interacted with planning algorithms to control multiple UAVs. Despite the robustness of these finding they are difficult to apply to design of an open-loop AA system because they beg the question of *where* over the experimental session the detrimental effects of high utilization are occurring. Because AA can monitor and intervene at any time it is important to know whether these effects are cumulative, operate within some time window, or are strictly local. A utilization of 70% over a 20 minute span can be achieved in many ways, for example by using many short or fewer long breaks such as 14 minutes work with 6 minutes break, twice 7 minutes work with 3 minutes break, etc. An open-loop AA system requires precise predictions as to the effect of these different choices. Our paper provides a first framework to determine the effect on performance of different schedules for break-work intervals for a given task.

The utilization approach in this paper is derived from a study using a simplified ATC system. An airspace control ATC task in which the operator performs direct routing and maintains separation between aircraft was used. The airspace control task demands that an air traffic controller identify the number of aircraft in transition through a sector, the number of aircraft changing altitude, and the number of potential conflicts (Loft et al. 2007). ATC requires operators to maintain continuous attention on the display and to strategically control aircraft during the entire task. In such a dynamic system future output depends on both past and human input (Rouse, 1981). These complex dependencies make it more difficult to construct an operator error model since the actual cause of the error is difficult to infer. Using operator error models that aim to capture all such dependencies becomes intractable. To circumvent this, one can use simpler error models that do not capture all the complex dependencies but make it easier to identify points of failure using a simplified workflow. Endsley (1998) presented such a model in which operator errors were divided into two classes based on the level of situation awareness implicated in the error. Lindsay (2002) developed a comprehensive operator choice model (OCM) precisely

characterizing possible errors for highly simplified version of the ATC task. The OCM describes the sequence flow of operators performing an ATC task: (1) the operator systematically scans the display searching for possible conflicts; (2) the operator focuses on a particular pair and classifying the pair as being in conflict or not; (3) if the monitored pair is in conflict, the operator determines an action plan for the pair; and (4) performing corrective actions on the pair (i.e. setting new speeds for one or both of the aircraft). This workflow can be used to identify errors and categorize them as failures of scanning, deferring an action too long, or for some other reason.

We propose a general approach to determine the effect of utilization on performance and errors to support the ongoing work on mental workload measurement techniques. Because we expect very different psychological effects involving fatigue, adaptation, or time stress from sequences of differing break lengths, the usage of simple average utilization is not psychologically well justified without reference to its basic break-work sequences. In this paper we address two basic questions for a given task: 1) are the effects of utilization independent or dependent on the length of break-work intervals, and 2) do effects of utilization reset after each break-work interval or are they cumulative across multiple intervals?

METHODS

Air Traffic Control Simulator

We propose a mental workload measurement approach based on operator utilization measured in break-work intervals, or time windows, of varying length. The ATC system used to investigate this approach presents the operator with complex cognitive tasks requiring situation awareness in order to make correct decisions and respond promptly. The primary task of the operator is to project the route of an incoming aircraft and set its speed so that the aircraft will not collide with other aircraft within the sector. An incoming aircraft requiring a decision is marked in magenta to acquire the operator's attention. All other aircraft are marked in gray. These are either aircraft for which the operator already set a speed or aircraft that do not require decisions. The latter are spawned in order to keep a certain number of airplanes in the air to maintain the difficulty of each task at a constant level. The trajectories of aircraft could be projected by their heading and speed and are linear, as seen in Figure. 1. The conflict or separation violation between aircraft is defined as a loss of separation of 50 pixels on the display, and the minimum separation distance of each aircraft is indicated as a circle outside the aircraft.

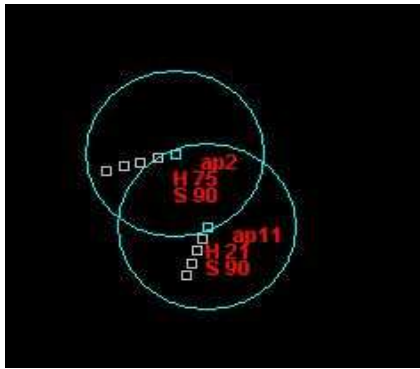


Figure 1. Two airplanes in a separation violation. The white squares mark the trace of each airplane.

Aircraft are spawned with a random configuration, i.e. speed, heading, and start coordinates, with the restriction that one of the speed settings the operator can choose leads to a conflict free path through the sector, so that every task is in fact solvable. The operator has a one-time chance, right after spawning, to modify the aircraft’s speed by clicking the right button when the cursor is pointed over it.

The speed menu has three options; they are speed decrease (-10 knots), maintain current speed (current), and speed increase (+10 knots). An operator can make two types of errors. The first error is choosing a setting for the speed that leads to a collision with an airplane already in the sector. The second error is delaying the decision to a point at which the aircraft collides with another. Each aircraft's trajectory may have multiple cross-points with other aircraft, but operators are advised to focus on potential conflicts that occur earlier. The canonical scenario resembles an actual ATC system in which an air traffic controller must acquire details of the aircraft, which include the type of aircraft, the path that the aircraft will travel, and the trail that the aircraft enters; must control the flight by adjusting the speed, altitude, climb/descend rate and route planning; and must monitor the status of the aircraft. The present simplified scenario of our ATC system maintains the perceptual challenges of a real system, but drastically reduces task complexity allowing exhaustive classification of errors. These simplifications allow a typical participant to perform the task without any prior air traffic controller experience.

Experimental Design

The experiment consists of break-work intervals each with a break followed by a work period. Work periods start with multiple aircraft entering the sector, one of which requires a speed setting from the user. After the operator chooses the speed of an aircraft a new aircraft is spawned to keep the operator working continuously. During breaks no new aircraft are spawned until the break period is complete, but all aircraft on the display continue on their paths.

Results from a pilot study indicated that operators have a similar task completion time regardless of configuration of aircraft on the screen. Operators need an average of 15.8 seconds to deal with 5 aircraft, so we use 15 seconds as one standard unit of work time and vary the break lengths of

intervals by 0.5, 1.0, 1.5, and 2.0 units of work time. This corresponds to four different interval types A, B, C, and D with break lengths of 8 seconds, 15 seconds, 22 seconds, and 30 seconds, respectively.

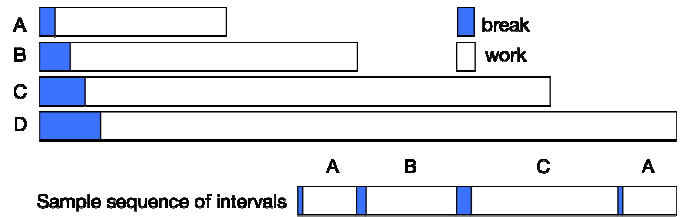


Figure 2. An example of four basic break-work intervals A,B,C, and D at 90% average utilization with 8,15, 22, and 30 second breaks.

As an operator enters a break-work interval he rests for the duration of the break. At the point he completes the first increment of work the current utilization within intervals of type A, B, C, and D would be 66%, 50%, 40%, and 33% respectively. After the second unit of work each of these utilization percentages increase, until at the end of the entire interval the user returns to the overall average utilization. For our experiments we chose the size of the work periods for each interval to yield a final 90% average utilization across all intervals. In the different intervals an operator is hence exposed to very different patterns of utilization while maintaining the same overall average utilization. Table 1 describes the four types of break-work intervals in detail.

Table 1. Break-work interval

Break-Work Interval	Break length(sec)	Work length(sec)	Interval length (sec)
A	7.5	75	82.5
B	15	135	150
C	22.5	195	217.5
D	30	240	270

To create longer experimental trials we concatenate the four basic types into longer sequences. There are four trials corresponding to sequences ABCADB, BCDBAC, CDACBD, and DABDCA. This covers a wide range of possible sequences. Every type of break appears at every position in the sequence. Additionally, every interval type precedes every other interval type at least once (for example, type A occurs at least once after B, C, and D). Obviously with a sequence length of six one cannot cover all 4⁶ possible sequences for our analysis. Every participant receives three trials each taking no longer than 20 minutes.

Participants

Twenty-eight paid participants were recruited from the University of Pittsburgh student community. None had prior experience with air traffic control, although they had all been frequent computer users.

Procedure

After reading instructions on how to modify aircraft speed via the ATC simulation interface, all participants in all

conditions practiced the speed change operation in a five-minute training session. After the training session, participants began the experimental session in which they performed the scanning and decision-making control for multiple aircraft. Each participant received three trial sequences, and each took no longer than 20 minutes. After each sequence, the participants were asked to complete the NASA-TLX workload survey and could rest for up to five minutes.

RESULTS AND DISCUSSION

During the simulation we recorded every choice of the operator and whether it led to an error, either by choosing late or by choosing the wrong option. From this we computed the performance measure, the probability that an operator fails to prevent a collision for a task. One estimate of the probability of an error is the fraction of correct to erroneous decisions within one unit of work time. Intervals A, B, C, and D have 5, 9, 13, and 16 units of work time respectively. Each unit of work time also has its associated utilization from approximately 0.33 up to 0.90 with overall 33 different utilizations. Figure 3 shows the distribution of work time units across utilization levels for the average operator and Figure 4 shows the distribution of utilization across intervals.

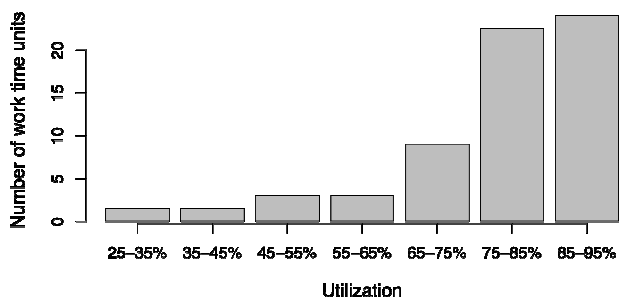


Figure 3. The number of units of work time at different utilizations that the average operator encountered (see also Fig. 4) across all intervals and trials.

A one-way analysis of variance of the error probability across different utilization levels, grouped in bins of size 10% as shown in Figure 3, revealed significant differences across utilization ($p < 0.01$). A further regression analysis on the effect of utilization on the error probability, in this case without grouping, determined a linear coefficient $b=0.13584$ ($t(2671)=3.932$, $p < 0.001$) with an intercept of $a=0.23519$. The average error probability across operators for each of the 33 differently utilized units of work time, as well as the linear model from the regression, is shown in Figure 5. These results suggest a linear relationship between utilization and error probabilities *within* a break-work interval.

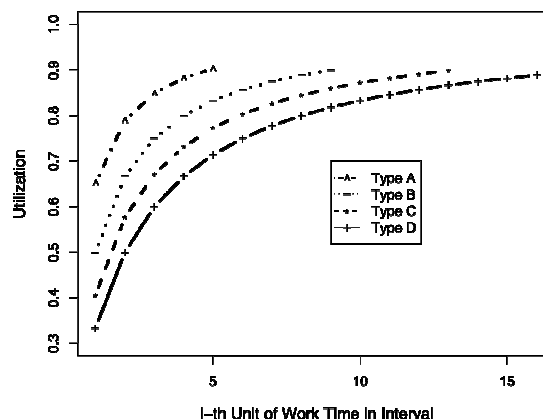


Figure 4. The utilization level of the i-th interval for all types.

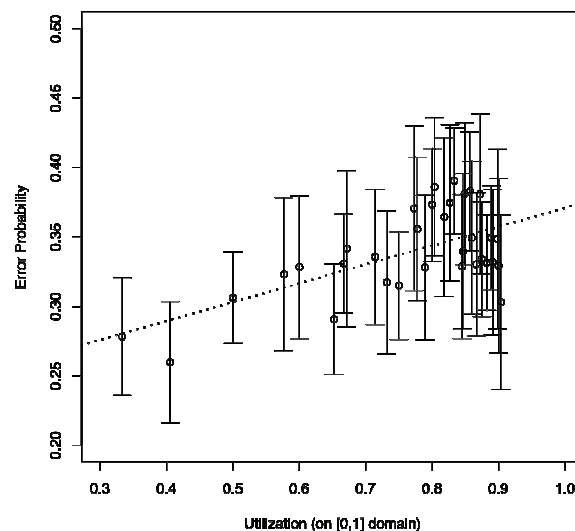


Figure 5. The error probability is plotted for differently utilized units of work time, averaged across all operators. The linear model is plotted as a dashed line. The standard deviation of the means is illustrated with whiskers.

The above poses another immediate question whether performance differs across types of break-work intervals since they are composed of units of work time with different utilization levels. To answer this we computed the error probability for every type of interval for every operator. For this error probability a one-way analysis of variance across different break-work interval types shows no significant differences ($p = 0.853$). The interval types A, B, C, and D performed with average error probabilities of 0.328, 0.336, 0.347, and 0.340, respectively. This suggests that the effects of utilization *across* break-work intervals are similar, despite the differences discussed above. In fact, longer intervals, such as C and D have lower utilized units of work time but also more higher utilized towards the end. But notice that the utilization within a break-work interval approaches 0.9 (90%) according to:

$$u(t, I) = 1 - \frac{b(I)}{t},$$

with t being the time passed within interval I with break length $b(I)$. Hence, if one were to increase break-work interval sizes even further than feasible in our experiments, then time work units with high utilization will occur much more frequently and may affect performance.

The second question to address is whether we have long term cumulative effects across break-work intervals. A correlation analysis between error probability and the start time of a unit of work time in a trial revealed no significant correlation between time and error probability ($p = 0.2728$). Notice that at the end of every break-work interval the overall average utilization is always 0.9 (90%) which supports the interpretation of this result as confirming that the effects of utilization for our ATC tasks do not accumulate across break-work intervals. But we can also perform a more detailed investigation by determining the effect of preceding break-work intervals onto a given break-work interval. For example, break-work intervals following an interval type D may perform worse if the effect of the high utilization values towards then end of D carries over the break. We would also expect type A intervals to be more affected since they have shorter breaks. A one-way analysis of variance across with preceding intervals as independent variable for each of the interval types revealed no significant effect of preceding type onto type A, B, C, or D (for all tests: $p > 0.4$). Additionally, an analysis of variance with type of interval as independent variable for all preceding types revealed no significant differences in performance (for all tests: $p > 0.6$). Hence, the type of preceding interval does not influence the performance in the subsequent interval and intervals are not affected differently by preceding intervals. This additionally supports the hypothesis that utilization effects do not accumulate across break-work interval within the context of our experimental setting.

CONCLUSION

We have presented an approach that enables a detailed analysis of workload based on break-work intervals and utilization patterns relating to overall performance in terms of error probabilities. We exemplified the use of this approach on an ATC system and answered two basic questions about the effect of utilization on performance while also providing the means to identify additional details of the relation between utilization and performance. From our analysis we conclude that in order for the design of an AA system for ATC one can focus on local effects within break-work intervals. Additionally, the linear relationship that we identified also allows designer to determine the optimal production rate of the system. Our approach offers the potential for much more detailed analysis and extensions, such as the inclusion of elaborate cognitive models of operators. For all intents and purposes the presented study is only a first, albeit significant, step towards the design of truly capable open-loop adaptive automation systems.

ACKNOWLEDGMENT

This research has been sponsored in part by AFOSR FA955008-10356 and ONR Grant N0001409-10680.

REFERENCES

- Bechtold, S.E., Janaro, R.E., Sumners, D.L. (1984). Maximization of labor productivity through multi-rest break scheduling. *Management Science* 30, 1442–1458.
- Beevis, D., & Essens, P. (1996). State -of-the -art report; Improving function allocation for integrated systems design. ALLFN'97D Revisiting the allocation of functions issue, Galway, IEAPress, Louisville .
- Cain, B. (2007). *Review of the mental workload literature*. Report #RTO-TR-HFM-121-Part-II. Defense Research and Development Canada, Toronto.
- Carnahan, B.J., Redfern, M.S., Norman, B., (2000). Designing safe job rotation schedules using optimization and heuristic search. *Ergonomics* 43 (4), 543–560.
- Cummings, M., Clare, A., Hart, C. (2010). The role of human-automation consensus in multiple unmanned vehicle scheduling, *Human Factors*, 52(1), 17-27.
- Cummings, M. & Guerlain, S. (2007). Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles, *Human Factors*, 49(1), pp. 1-15.
- Crandall, J. W., Cummings, M. L., Penna, M. D., & deJong, P. M. A. (2010). Computing the Effects of Operator Attention Allocation in Human Control of Multiple Robots. *IEEE Transactions on Systems, Man, and Cybernetics -- Part-A*.
- Endsley, M. R., & Rodgers, M. D. (1998). Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. *Air Traffic Control Quarterly*, 6(1), 21-44.
- Gopher, D. and Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures?, *Human Factors* 26(5): 519-532.
- Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information processing load. *Human Factors*, 39, 30–47.
- Lindsay P. and Connelly S. (2002). Modelling erroneous operator behaviours for an airtraffic control task. In J. Grundy and P. Calder, editors, *Third Australasian User Interfaces Conference (AUI2002)*, volume 7 of *Conferences in Research and Practice in Information Technology*, pages 43–54. Australian Computer Society, Inc.
- Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, 49, 376–399.
- Neerinx, M. A. (2003). Cognitive task load design: model, methods and examples. In Hollnagel, E. (Ed.), *Handbook of Cognitive Task Design*: 283-305. Mahwah, NJ: Lawrence Erlbaum Associates.
- Paul, P., Kuijter, F.M., Visser, B., Han, C., Kemper, G., 1999. Job rotation as a factor in reducing physical workload at a refuse collecting department. *Ergonomics* 42(8), 1167–1178.
- Parasuraman, R. and Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration, *Human Factors*, 52(3), 381-410.
- Reed, J. (1992). A plant local panel review. In: Kirwan, B., Ainsworth, L.K. (Eds.), *A Guide to Task Analysis*. Taylor & Francis, London, pp. 267–288.
- Rouse, W. B. (1981). Human-computer interaction in the control of dynamic systems. *ACM Comput. Surv.* 13, 1 (Mar.), 99.
- Schmidt, D. K. (1978). A queuing analysis of the air traffic controller's workload. *IEEE Transactions on Systems, Man, and Cybernetics*, 8, 492–498.