To appear in the *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica: Human Factors and Ergonomics Society, 2004.

MEASUREMENT OF TASKLOAD AND PERFORMANCE IN A DYNAMIC MULTI-TASK EXPERIMENT

Brian R. Levinthal and Esa M. Rantanen,
Aviation Human Factors Division, Institute of Aviation
University of Illinois at Urbana-Champaign

This paper discusses an experimental paradigm for measuring human performance under time pressure. Participants were presented with four simultaneous number-entry tasks. Entry could occur only within a discrete window of opportunity, represented visually by a target range within a variable-speed progress bar. Participants could view only one task at a time, thus performance required scanning and sampling of all four tasks. Sampling behavior (mouse movements to the vicinity of one of the four tasks) was manipulated via access effort; blocks of trials presented either a half or full second lag before a target task would become visible. Participants' performance was evaluated by the proportion of responses completed within the required window of opportunity as well as the proportion of a window of opportunity that elapsed before the onset of a response. Results indicate a decrease in such performance as a result of display lag manipulation. The potential for the use of this paradigm in developing predictive measures of performance is discussed.

INTRODUCTION

Successful control of dynamic systems implies that the users have a "mental model" of the system, allowing them to predict its behavior and the consequences of their inputs to it. The necessity of prediction in the control of even the simplest devices or in the performance of the most mundane everyday tasks is obvious, and bespeaks of the congenital role of temporal mental models in human behavior and performance. Time is hence an integral dimension of mental models as well as an inherent component and constraint in nearly every human activity.

In addition being relevant to anticipatory behavior in control of dynamic systems, time offers attractive methods for the measurement of covert mental models. Timing data (e.g., reaction times) are relatively easy to obtain under both experimental and naturalistic conditions, and time is a variable that is common to the human, the task, and the environment. Time offers a common unit of measurement of human performance in the context of the task, and can be used to infer the "goodness" of the temporal dimension of the operator's mental model of the task or system being controlled.

The specific demands imposed on humans by control of complex and dynamic systems can be manifested objectively as taskload and subjectively as workload. Taskload generally exists as a combinatorial result of all actions taking place and situations developing in a given time. Workload refers to the subjective experience of the operator in response to the taskload, and cannot be measured directly. Both are often inversely related to performance, which can be quantified. Thus, direct measurement of taskload and performance will allow for inferences to be made about workload experienced by the operator.

It may be hypothesized that the operator's timing performance is a "product" of two parallel processes, an internal "clock" or some timekeeping mechanism, and the attentional and perceptual processes that sample the external environment (c.f., Neisser's perceptual cycle). It may be further hypothesized that human sampling behavior is driven by the goodness of the temporal mental model, and in particular, three distinct aspects of it: (1) correct time to act or update the mental model from cues available in the environment, (2) the time available for action or checking of cues, and (3) the time required to perform action or check cues. The first depends on prospective memory and the latter two would be characterized by a mental model of information access cost.

We make several key assumptions for development of time-based measures of taskload and performance: (1) A critical variable determining taskload is time pressure, which may be expressed as Time Available – Time Required. (2) An individual's ability to perform under time pressure is dependent on an accurate temporal mental model of the task demands (Time Available) and own performance (Time Required). (3) The goodness of this temporal mental model is manifested in proper calibration of Time Required vs. Time Available, resulting in accurate and consistent timing of overt actions and correct action sequencing and scheduling. (4) The temporal mental model is affected by taskload as well as individual differences. (5) Measurable timing performance will allow for inferences to be made of both taskload and individual differences in coping with it.

During any time-critical task, an individual will be confronted by three relevant features: (1) The correct time to act, or opening of a window of opportunity, (2) time available, or duration of a window of opportunity, and (3) time required to complete the task within a window of opportunity. In order to complete all tasks in an efficient manner, the operator must be aware of all three for each task. Although optimal sequence of tasks to be completed can be computed, it is a resource-intensive task that is often avoided (Moray, Dessouky, Kijowski, and Adapathya, 1991). It is likely that human operators will opt for simple strategies, placing salient visual features of the timing task above the more accurate information available from mental arithmetic or duration estimation. There is evidence that participants, given visually different stimuli, tend to underestimate their window of opportunity with faster-moving trials over slower-moving trials, even when the duration of trials is equal (Rantanen & Xu, 2001). Furthermore, there are substantial performance decrements when visual stimuli are removed, forcing participants to rely on duration estimation (Xu & Rantanen, 2003).

This paper presents a multiple-task time pressure experiment that manipulated taskload using a time available/time required model, and measured the subsequent effect of taskload on performance. Time available was manipulated by varied temporal windows of opportunity, and time required was manipulated via increased scanning costs.

METHOD

Participants

Twelve students from the University of Illinois at Urbana-Champaign, 20–30 years of age, participated in the study. Participants were compensated $8 for an hour-long experiment, consisting of four 10-minute blocks during which participants were presented with a time pressure simulation involving tasks of low cognitive demand.

Apparatus

A computer program developed specifically for this experiment presented participants with an abstract time-sharing task. The task was performance-dependent; speed and accuracy of performance in one trial affected the onset of subsequent trials. In this way, the task mimicked the dynamic nature of real-world scenarios (e.g., Air Traffic Control). Participants viewed a computer screen divided into four panes, of which only one was visible at a time. To view other quadrants, participants were required to move a cursor (using a mouse) to the desired quadrant. The previous portion of the screen would become blank, and the desired quadrant would become visible.

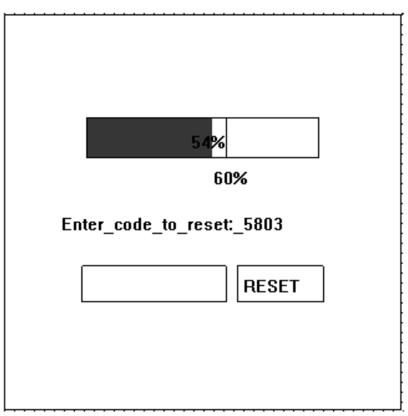

Figure 1. A sample pane in the Task Scheduler program.

Each quadrant contained a red progress bar, an indication of the current percentage progressed, a mark on the bar indicating the window of opportunity to reset the timer (the required task), and instructions for resetting the timer by typing a four-digit code (Figure 1). The participants were required to observe the progress bar to reach a strict window of opportunity and enter the code before the bar reached 100%.

Parameters

The progress bars, present in each quadrant, moved at one of three different speeds: 60, 30, and 15 seconds to progress from 0 to 100% (a ratio of 1x : 2x : 4x, respectively). The "opening" of the windows of opportunity to reset the timers corresponded to the critical percentage point for a given trial, which occurred at 60, 80, and 90% (creating a window size of 40, 20, or 10%). While there were nine possible combinations of speeds and window sizes, there were only five possible durations of the window of opportunity. The five window durations were 1.5 seconds, 3 seconds, 6 seconds, 12 seconds, and 24 seconds. Pilot testing indicated that the 1.5-second durations were typically too short to complete the task, and that the 24-second durations were excessively long. Both window durations were removed from this experiment, thus only seven combinations of speed and percent were used (Table 1). Trials were sampled randomly from an even distribution of these seven speed/window size pairings, such that past trials held no predictive value for subsequent trials. During a block of trials, the delay parameter was set either at .5 or 1 second.

	Speed			
		1x	2x	4x
Window size (%)	10%	6 sec	3 sec	1.5 sec
	20%	12 sec	6 sec	3 sec
	40%	24 sec	12 sec	6 sec

Table 1. Window durations resulting from Speed/Window size pairings. The 1.5 and 24-second durations were eliminated for the experiment.

Procedure

The participants were instructed to type the four-digit sequence in each quadrant to reset the timer, and to do so after its progress bar reached the critical percentage, but before the bar reached 100%. Once a trial was successfully completed, the red progress bar reset to 0%, marking the beginning of the next trial. Participants were presented with four 10-minute blocks of such trials. While no penalties were enforced for failure to complete the task within this window of opportunity, participants were asked to be as diligent as possible.

Output

The program recorded a timeline (accurate to .01 seconds) of all events that transpired during the experiment, including movements from one quadrant to another, the opening and closing of windows of opportunity, time of task completion, keystrokes made by the participant, and the speed and critical percentage of the bar. These data were sufficient for recreating the participant's actions during the experiment. Several taskload and performance metrics were derived from these raw data.

RESULTS

Data Reduction and Coding

Two metrics for performance were used to judge individual performance during the time pressure task simulation. Performance was judged based on the proportion of trials that were successfully completed within their window of opportunity, as well as the percentage of the window of opportunity that elapsed prior to the initiation of a response (first key pressed). Good performance was manifested by the execution of tasks within the window of opportunity, as well as a minimization of lag between the opening of a window of opportunity and the onset of a response (i.e., timely performance).

The raw output of the program used in our paradigm also allowed for the derivation of several objective taskload and performance metrics. This analysis will deal primarily with two such metrics whose difference resulted in our primary measure of taskload. The first metric, model A, relates to individual performance. Model A is calculated as the ratio of time used to complete all trials in a given epoch (two minutes) to the average time required to complete that number of trials. "Time used" was the amount of time that elapsed from the opening of a window of opportunity until the completion of the task by the participant, and was summed across all tasks that occurred during an epoch. This is primarily a measure of efficiency, as lower values of A would reflect more optimal performance of the tasks that occurred during a specific epoch.

The second metric, model B, is calculated as the ratio of the summed window durations for tasks that occurred during an epoch to the average amount of time required for the participant to complete those tasks. Low values of B reflect minimal excess time to perform tasks, while higher values approximately represent the amount of buffer participants had within that epoch to perform the necessary tasks. Thus, model B functions primarily as a measure of task demand.

The difference between model A, efficiency, and model B, task demand, results in the primary taskload metric. A negative value of model A – model B indicates that there was more time available than was used during that epoch, and a positive value indicates that less time was available than was required.

Window Size and Response Initiation

Windows of opportunity varied based on the critical percentage and the speed of the moving bar, resulting in three discrete window durations (12 seconds, 6 seconds, and 3 seconds). The response initiation percentages were analyzed across the primary delay condition and speed/window size pairings by ANOVA. The main effect of delay condition was significant, $F(1, 3638) = 58.65$, $p < .001$, as were the main effects of speed and window size, $F(2, 3638) = 268.63$, $p < .001$, and $F(2, 3638) = 105.221$, $p < .001$, respectively. Response initiation percentages were elevated with greater delay, and also became larger as window size and speed increased (Figure 2).

Post-hoc comparisons of speed/window size pairs were performed to determine differences in performance within the three window durations. In the .5-second delay condition, response initiation percentage differed significantly for the two 12-second pairings, $t(489) = 4.499$, $p < .005$. Three comparisons were necessary for the 6-second window pairings as it was possible to form a 6-second window for all three speeds. The effects of pairing on performance between the 1x / 10% window and the 2x / 20 % window were significant, $t(471) = 4.470$, $p < .005$, as were 1x/10% and 4x/40% pairings, $t(578) = 7.057$, $p <.005$, and 2x/20% and 4x/40% pairings, $t(507) = 2.843$, $p < .005$. There was no significant effect of pairing on performance for the shortest window duration within the .5-second delay condition. For the 1-second delay condition, 12-second window comparisons demonstrated a significant effect on performance, $t(470) = 3.605$, $p < .005$. The three 6-second pairings produced a significant effect on performance as well, $t(470) = 4.732$, $p<.005$, $t(592) = 8.751$, $p < .005$, and $t(518) = 3.563$, $p < .005$, respectively. As in the .5-second delay condition, there was no significant effect of pairing on performance for the 3-second window.

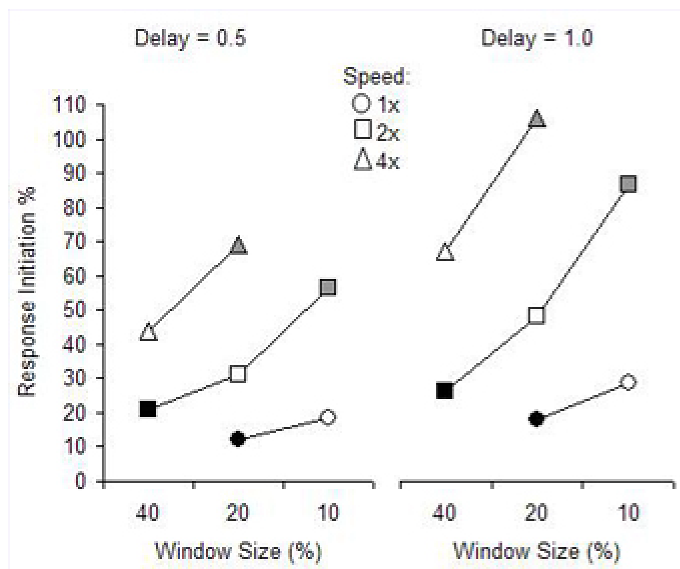

Figure 2. Average percentage within the window of opportunity before the initiation of a response, separated by speed/window size pairing. The three window durations are grouped by color (Black: 12 second | White: 6 second | Grey: 3 second).

Taskload and Performance

A linear regression analysis was performed to relate individual performance to level of taskload. The mean proportion of tasks that were successfully reset within their window of opportunity decreased reliably as taskload increased, $R^2 = .614$ (Figure 3). The mean percentage of elapsed window of opportunity prior to response onset increased with taskload, $R^2 = .608$ (Figure 4). Mean initiation times greater than 1.0 indicate that responses began after the closing of the window of opportunity.

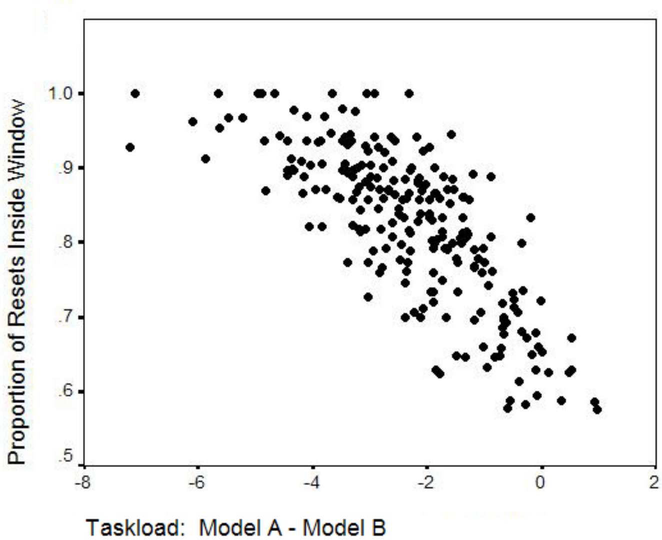

Figure 3. The proportion of responses completed within the window of opportunity decrease as taskload increases.

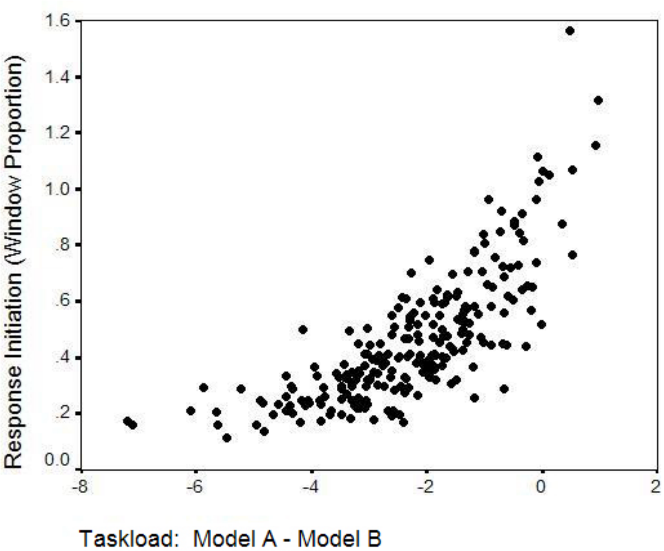

Figure 4. The proportion of the window of opportunity that elapsed prior to the initiation of a response increases as taskload increases.

DISCUSSION

Participants were less accurate in their responses when the presented delay was increased and for both delay conditions, participants' accuracy was below the ceiling level. Increasing the sensitivity of our paradigm was of primary importance in this experiment, and the addition of the delay condition (at both .5 and 1 second levels) appeared to achieve this goal.

Previous studies have indicated that, under time pressure, participants will opt for strategies employing salient visual features of windows of opportunity when planning behavior (Rantanen & Xu, 2001). This finding was replicated in our experiment, as for both delay conditions visually different pairings of temporally identical windows of opportunity were treated differently. This was true for the longer duration pairings (6 or 12 seconds), but not for the shortest 3-second windows. It is possible that participants were more restricted in their performance for shorter windows of opportunity due to limitations in typing speed or other factors related to task completion time.

This experiment also demonstrated the feasibility of time-based metrics to accurately describe and quantify both operator taskload and performance in dynamic tasks. The development of multiple metrics allows for the examination of diverse aspects of operator performance in a wide variety of situations.

ACKNOWLEDGMENTS

This research was supported in part by a collaborative agreement no. DOT 02-G-019. Dr. Carol Manning was the technical monitor. We also wish to thank Sharon Yeakel for the development and modification of the software used in this experiment.

REFERENCES

Chapanis, A. (1959). *Research techniques in human factors*. Baltimore, MD: Johns Hopkins University Press.

Moray, N., Dessouky, M.I., Kijowski, B.A., Adapathya, R. (1991). Strategic Behavior, Workload, and Performance in Task Scheduling. *Human Factors, 33*(6), 607-629.

Rantanen, E. M., & Xu, X. (2001). Human performance in timing of discrete actions. *Proceedings of the 45th HFES Meeting* (pp. 527-531). Santa Monica, CA: HFES.

Xu, X., & Rantanen, E. M. (2003). Timing of discrete actions: Evidence for perceptual dominance in human performance. *Proceedings of the 47th HFES Meeting* (pp. 581-585). Santa Monica, CA: HFES.

Zakay, D., & Shub, J. (1998). Concurrent duration production as a workload measure. *Ergonomics, 41*(8), 1115-1128.