N90-25539

PERFORMANCE-BASED WORKLOAD ASSESSMENT: ALLOCATION STRATEGY AND ADDED TASK SENSITIVITY

Michael A. Vidulich
Harry G. Armstrong Aerospace Medical Research Lab
AAMRL/HEG
Wright-Patterson AFB, Ohio 45433-6573

ABSTRACT

The paper reviews the preliminary results of a research program investigating the use of added tasks to evaluate mental workload. The focus of the first studies was a reappraisal of the traditional secondary task logic that encouraged the use of low-priority instructions for the added task. It was believed that such low-priority tasks would encourage subjects to split their available resources among the two tasks. The primary task would be assigned all the resources it needed, and any remaining reserve capacity would be assigned to the secondary task. If the model were correct, this approach was expected to combine sensitivity to primary task difficulty with unintrusiveness to primary task performance. The first studies of the current project demonstrated that a high-priority added task, although intrusive, could be more sensitive than the traditional low-priority secondary task. These results suggested that a more appropriate model of the attentional effects associated with added task performance might be based on capacity switching, rather than the traditional optimal allocation model.

INTRODUCTION

Overview of Paper

The goal of the research described in the present paper was to produce guidelines for the behavioral assessment of workload. Following a few definitions, the paper will begin with a review of the conceptual underpinnings of the secondary task technique. The development of the traditional view of the secondary task technique will be reviewed, along with refinements that have recently been proposed by some authors. This will lead to a proposed alternative to the secondary task technique. The results of two experiments comparing the traditional technique to the alternative will then be presented.

Terminology. It is important to be precise about terminology. Typically, in performing a secondary task evaluation, there are two types of tasks. The first is a primary task. It is called the primary task because the subjects are told to maintain its performance at a single-task level. Typically, the primary task is the task that the researcher is interested in measuring. In an aircraft environment, a typical primary task might be maintaining flight control in a certain scenario with different possible display configurations. The secondary task is a task added by the researcher to perform the measurement. It is called secondary because the subjects are typically told to perform it as well as they can, without letting it interfere with the performance of the primary task. In other words, it is secondary in priority, relative to the primary task. The terms "primary" and "secondary" thus refer directly to the priorities the subject is instructed to assign the tasks.

Task Types. The most common pairing of task types in secondary task research (see Ogden, Levine, and Eisner (1979) for a review) has been a continuous primary task (for example, a tracking task in the lab or flight control in an aircraft simulator), paired with a discrete secondary task (for example, a Sternberg task in the lab or a communications task in a simulator). This is not the only possible combination of task types, and there is no logical reason that the assignments could not be reversed or that two tasks of the same type could not be combined. However, inasmuch as the continuous primary task and discrete secondary task has been the most common combination, it was selected for examination in the present research.

Early Single-Channel Explanations of the Secondary Task Technique

Knowles (1963) provided the first major review of secondary task research. Reflecting the dominant trend in attention theory of the time, Knowles' model of the secondary task paradigm was based on a single-channel, multiplexed information processor. In Figure 1, S1 represents the primary task's stimuli and D1 represents the primary task's decision processor. Due to instructions, or some comparable manipulation, the information processing channel would be switched to S1 and D1 whenever necessary to maintain the required level of performance. Whenever the channel was available, it would be switched to the secondary task (S2 and D2). Presumably, the more demanding the primary task, the less frequent and the shorter in duration the switches to the secondary task would be. Thus, it was expected that as primary task difficulty increased, its performance would be maintained (because of the priority instructions), but secondary task performance would degrade. In other words, the secondary task was expected to be both sensitive to primary task difficulty and unintrusive to primary task performance. Knowles acknowledged that such unintrusiveness was not common, but apparently did not think it was a major problem, so long as it was kept at a minimal level.

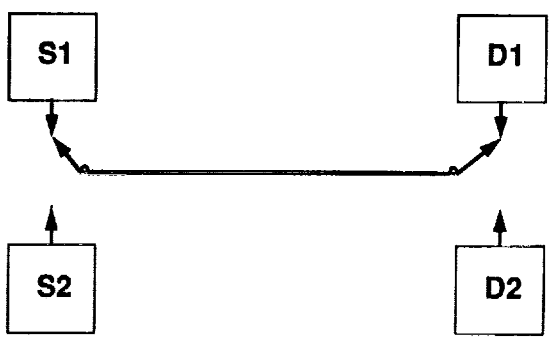

Figure 1 - Knowles' (1963) single-channel multiplexer model of the secondary task paradigm.

In 1971, Rolfe performed a review of a greatly-expanded data-base of secondary task research. As did Knowles, Rolfe explicitly argued for a single-channel model of secondary task operation. However, Rolfe also used the term "capacity," and discussed it as a divisible commodity. He described the secondary task technique as an attempt to measure the reserve capacity that was available whilst performing a primary task. Nevertheless, given Rolfe's (1971) strong endorsement of the single-channel "which must be allowed a finite time to process one stimulus-response before a second can be accepted" (p. 135), it is not clear that reserve capacity was thought to be anything more than spare time. This was consistent with Knowles' (1963) multiplexer model. Rolfe also pointed out that secondary task intrusiveness was a pervasive problem, and cautioned that primary task performance should be monitored to ensure that parity of primary task performance was maintained.

The Traditional Resource Interpretation of the Secondary Task

A major event in the interpretation of the secondary task paradigm was the publication of Kahneman's 1973 book on attention theory. The multiplexed single-channel was supplemented with the possibility of simultaneously sharing the available capacity (or "resources") among different tasks. In a number of different guises, capacity theory (or as it is also known, "resource theory") has become the central concept in discussions of the secondary task technique.

The traditional explanation of the secondary task (e.g., Williges and Wierwille, 1979) is based on the following logic: The human possesses a store of information processing resources (represented by the circle in Figure 2) that can be divided among tasks. The performance of the primary task demands some certain level of allocation from the store (represented by the shaded area). The traditional priority instructions are intended to ensure that the primary task always gets the amount of resources that it needs. Hence the name "primary." Whatever reserve capacity is left-over (indicated by the unshaded area) is allocated to the performance of the secondary task. The priority instructions are intended to ensure that the secondary task gets allocated no more than the reserve capacity of resources. Assuming the priority instructions are effective, the quality of the secondary task performance should be proportional to the size of the reserve capacity; the more reserve capacity the better secondary task performance should be. Thus, secondary task performance is expected to be sensitive to primary task demand. Also, the performance of the secondary task, if it uses only the reserve capacity, should be unintrusive to primary task performance. However, virtually every review of secondary task research since Knowles (1963) has bemoaned the fact that secondary task intrusiveness has been pervasive (e.g., Ogden et al., 1979; Rolfe, 1971; Williges and Wierwille, 1979).

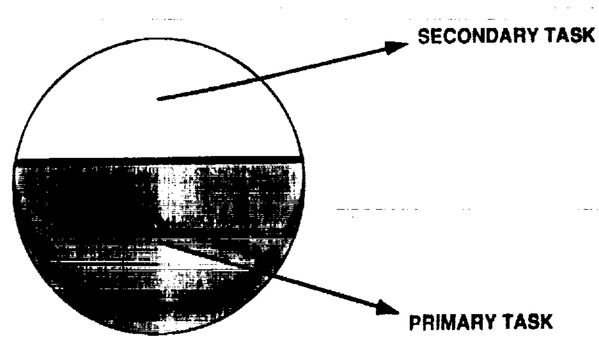

Figure 2 - The traditional capacity model of the secondary task.

Attempts to Counter Intrusiveness

Recently, there has been increasing concern about the consequences of secondary task intrusiveness. For example, Wickens (1984) cautioned that a simple comparison of the secondary task decrements that resulted from pairing with different primary tasks would be inadequate and potentially misleading. Were the subjects to differentially favor one version of the primary task less than the other primary task(s), they might inappropriately sacrifice primary task performance to maintain or improve secondary task performance. To counter this, Wickens (1984) argued that the joint performance on both the primary and secondary tasks must be compared within a Performance-Operating Characteristic (POC) space. A POC plot would have secondary task performance along one axis, and primary task performance along the other axis. Poor performance on either task would be near the origin of the figure. Thus, as joint performance of a task pair improves, the point representing that performance would move further from the origin. Also, the as the relative priorities between the time-shared tasks changes, the point will move closer to the axis representing the favored task. Therefore, plotting within the POC space provides an opportunity to compare the overall difficulty of the different primary tasks, even in situations in which parity of intrusiveness was not maintained.

Going even further with this logic, Gopher and Donchin (1986, p. 41-26) suggested that,

> Maximization of interference appears to be more consistent with the original secondary task rationale, in which the second task is added to saturate the capacity of the system, create an overload, and enable one to scale the demands of the primary task. It is, therefore, somewhat surprising that a lack of obtrusiveness of the introduction of a secondary task to the performance of a primary task has been identified as a highly desired property of a good secondary task.... How can this aspiration coexist with the main thrust of a technique that advocates the study of interference patterns as its main tool?

Gopher and Donchin (1986) pointed out that the traditional view of the secondary task made the strong, but questionable, assumption that subjects had full voluntary control over their resource allocation. As a result of problems with this, and other assumptions underlying the secondary task technique, Gopher and Donchin proposed replacing the traditional secondary task approach with full POC methodology. That is, instead of specifying a single set of priorities, subjects would be instructed to perform trials with a variety of relative priorities between the tasks. The use of multiple priorities would generate a sufficient number of data points for the generation of complete POCs outlining the performance tradeoff function between the two tasks. In contrast, Wickens' (1984) discussion of the secondary task technique was limited to the usual instructions, but evaluated the results within a POC space.

However, there are potential problems with the use of POC methodology as a workload assessment tool. For example, the use of multiple instruction sets requires an expansion of the experimental design. The minimum number would be two levels of relative priority for each task pair, which would double the size of any experiment relative to a single-point secondary task evaluation. This would, of course, be very expensive and time-consuming. Second, there is the issue of complexity. Complexity can refer to both the more complicated experimental design, which is likely to be unattractive to system evaluators who do not come from a background of attention research, and also to the more complicated data demands of full POC methodology. Unlike the traditional secondary task technique, plotting within a POC space requires a measure of primary task performance. Such a measure is certainly desirable, whenever it is available. But in complex operational tasks a measure of primary task performance might be unobtainable, or perhaps even difficult to define. Also, with the increasing trend towards automated systems, it is likely that there may be a need to assess workload in environments in which very few responses are ever made to the "primary" task. A third problem is operator acceptance. In operational settings, researchers often try to "hide" the secondary task as one of the tasks within a complex, but realistic, set of tasks. If subjects are told that one aspect of a task that they normally perform is changing priorities over trials, it might compromise the realism and reduce the operator acceptance.

Of course, if Gopher and Donchin are correct about the deficiencies of the traditional secondary task technique, then all of the problems associated with POC methodology will have to be accepted and applications constrained to environments allowing full POC methodology. But, the problems with using POC methodology as the standard technique are dire enough to provide serious encouragement for investigating other alternatives.

An Alternative Model of Secondary Task Operation

The present research was designed to study a possible alternative to POC methodology. But, before getting into the alternate approach or the research that was performed, it is advisable to alter the terminology somewhat. Instead of the usual primary/secondary distinction, the designation of the tasks will be separated from any explicit indication of their relative priorities. One task will be referred to as the "measured" task (i.e., the task whose demand must be measured). The other task will be referred to as the "added" task (i.e., the task that was added to provide a measure of the measured task's workload). The reason for this change should become obvious in a moment.

The traditional secondary task paradigm assumed that subjects could voluntarily control their resource allocation finely enough that the secondary task would have access to only the reserve capacity. However, the pervasiveness of intrusions in the previous secondary task research indicates that the subjects' control is not as flexible as expected. There have been numerous demonstrations that a sudden discrete stimulus can pre-empt the processing of a time-shared continuous task (e.g., Klapp, Kelly, and Netick, 1987; McLeod, 1977). So perhaps, instead of using only the reserve capacity for an added task, the subjects momentarily withdraw their resources from the measured task and reallocate them to the added task. This is illustrated in Figure 3 by the arrow moving from the measured task to the added task, implying that the discrete added task momentarily gains access to all resources.

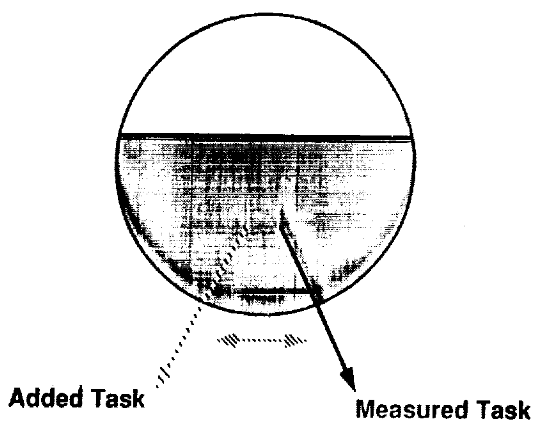

Figure 3 - An alternative model of added task methodology based on capacity switching.

This may appear to be very reminiscent of Knowles' (1963) single-channel interpretation, but there is a difference. In Knowles' model the switch was a structural feature, located outside the processors. The time taken for the switch to move from one task to another would be expected to be independent of the individual tasks' parameters. But, within what could be called a "capacity switch" model, it is reasonable to imagine that the amount of time required to reallocate resources to a new task might be influenced by the level of allocation to the current task. So, the more difficult the measured task, the longer it would take for the added task's pre-emption to be accomplished. Within this viewpoint, measured task intrusiveness is inescapable, but does not mitigate against sensitivity. The best way to measure the time required by the switch would be to instruct subjects to switch as quickly as possible to the added task, in the expectation that the switch would be slower as measured task difficulty increased. In other words, rather than the traditional low-priority secondary task, there would be a high-priority added task.

EXPERIMENT ONE

These two sets of priority instructions were tested in a laboratory experiment. The tasks were selected from the Air Force Criterion Task Set (Shingledecker, 1984). The measured task was a continuous unstable tracking task. An unstable tracking task is akin to trying to balance a vertical stick on the end of your fingertips. As the tracking task's lambda parameter is increased, the task becomes more unstable. The effectiveness of lambda as a workload manipulation, had been demonstrated in earlier laboratory experiments that studies the lambda effects on root mean square (RMS) error and subjective workload ratings (Shingledecker, 1984).

The added task was a Sternberg memory search task. The memory set was either 2 or 4 letters. Probe letters appeared periodically during the course of the trial. The average interstimulus interval (ISI) was 3 s, but the ISI varied randomly from 2.5s to 3.5s to produce some temporal uncertainty.

There were 20 subjects in the experiment. The subjects were recruited from local colleges and paid for their participation.

Each subject performed the dual-tasks under two sets of instructions. In one session the Sternberg task was the high-priority task. The subjects were instructed to track as well as they could when no Sternberg probe was present, but to consider the Sternberg the most important task whenever one of its probes appeared. In another session, the subjects were told the Sternberg was the low-priority task. They were asked to perform the Sternberg as well as they could, without letting it interfere with tracking performance. This corresponds to the usual secondary task instructions.

The question the experiment was designed to address was; "Which instructions would produce the Sternberg performance that was most sensitive to tracking difficulty?" That is, under which set of instructions would Sternberg reaction time be most affected by tracking lambda?

Figure 4 displays the reaction time (RT) results as a function of lambda, for the two sets of instructions: The high-priority Sternberg results are on the left, and the low-priority Sternberg results are on the right. In neither case, were the Sternberg reaction time data sensitive to the tracking task's lambda. This result was not anticipated. However, an interpretation of this result is possible within multiple resource theory.

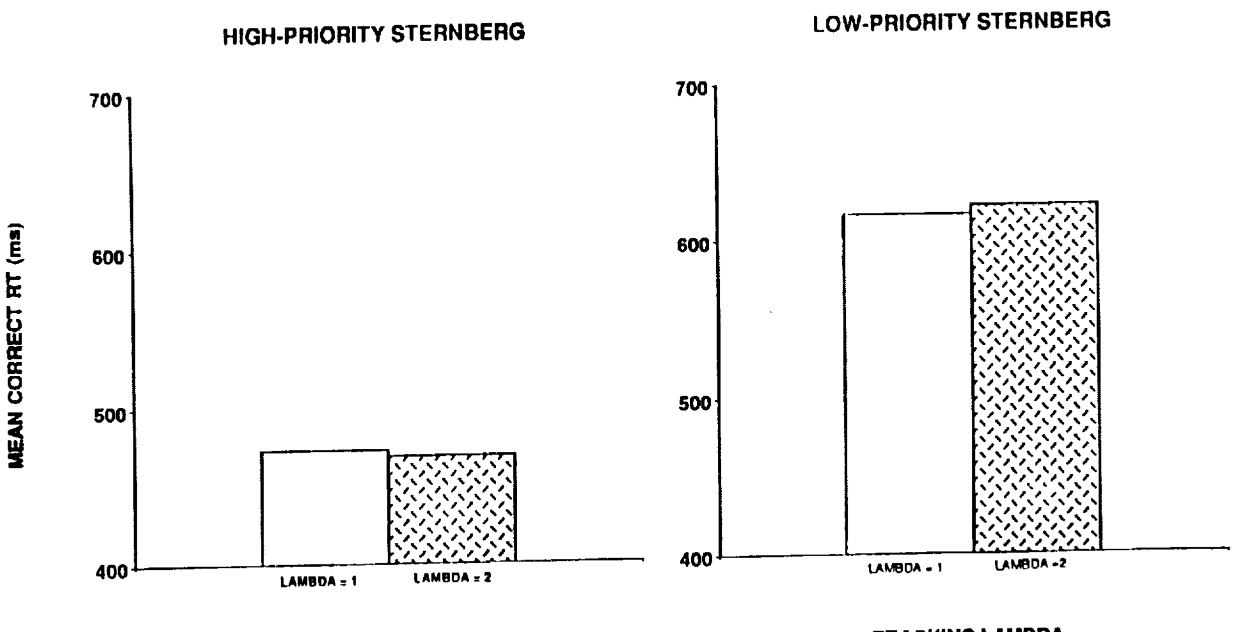

Figure 4 - Experiment 1, RT of correct Sternberg task responses as a function of priority and tracking lambda.

Wickens' (1980) multiple resource model distinguishes between early, perceptual/central processing resources and late, response execution processing resources. The tracking task used an easily-seen, highly-compatible display with no real need for predicting future cursor actions (as would be required in higher-order tracking). So, the perceptual/central demands of this tracking task were probably very low, regardless of the lambda level. But, the tracking did require the generation of frequent graded responses. Increasing lambda to 2 would increase the frequency of these responses, adding even more to the high response execution demands, but possibly without any significant increase in the perceptual/central demands.

On the other hand, the response execution demands of the Sternberg responses were probably trivial, because the subjects only needed to press a button. But, the scanning of the memory set would be expected to place relatively heavier demands on perceptual/central resources, and these demands would be expected to increase as the memory set was expanded from 2 to 4 items.

So, it is plausible that the two tasks simply failed to overtax any single source of resources. To test this hypothesis, the Sternberg task was redesigned to compete more directly with the tracking.

EXPERIMENT TWO

In Experiment 2, the measured task was still the unstable tracking task with two levels of lambda. However, the Sternberg task was changed. Instead of manipulating the Sternberg task's memory set size (it was set to 3), the response device was manipulated. On half of the trials, the subjects used the same button box that was used in Experiment 1, but on the other half of the trials the subjects had to deflect a joystick in the appropriate direction and press a button on top. Not only did this joystick response require a greater muscular involvement than the button-press; but also, because the joystick was on an easily-tipped stand, it required a very precise movement, as well. Changing the Sternberg response device from button to joystick was intended to directly compete with the tracking task's demand for response execution resources.

The same priority manipulation used in Experiment 1 was used in Experiment 2. The experimental questions were now twofold: One, would the control manipulation increase inter-task interference enough to generate sensitivity? And, Two, if it did, which instructions would produce the most sensitive Sternberg task performance?

There were 26 subjects in the experiment. The subjects were recruited from local colleges and paid for their participation.

In Figure 5, it can be seen that the manipulation was effective. On the left side of Figure 5 there is a small, but statistically significant, effect of lambda on the Sternberg reaction time. When subjects were given the high-priority Sternberg instructions, RT to the Sternberg task was 11 ms faster when time-shared with the easy lambda tracking task than with the difficult lambda tracking task ($F(1,25) = 4.69$, $p < 0.04$). On the right side of Figure 5, it is clear that the low-priority Sternberg is still insensitive to tracking demands.

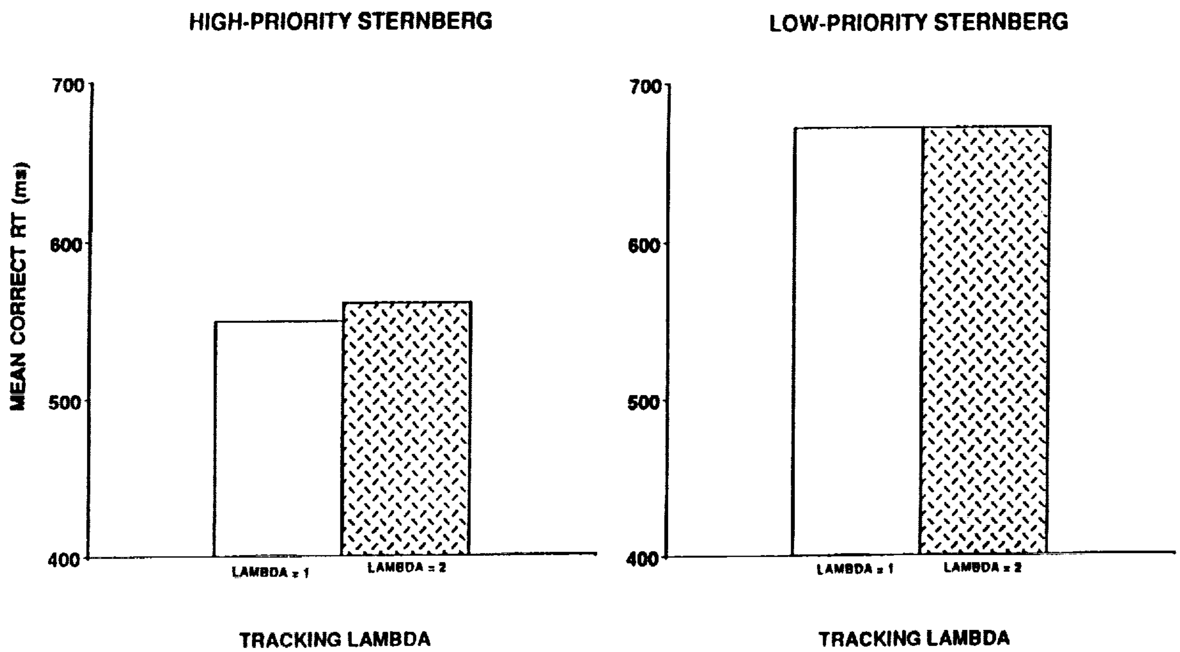

Figure 5 - Experiment 2, RT of correct Sternberg task responses as a function of priority and tracking lambda.

But, what about intrusiveness? To compare the intrusiveness caused in tracking performance by the occurrence of the Sternberg probe under the different priority instructions, the tracking RMS error was analyzed by intervals around the Sternberg probe occurrence time (see Figure 6). In the center of each figure, is the one-second interval during which the Sternberg probe was presented. The two one-second intervals directly before each Sternberg probe presentation, and the two intervals directly after were also recorded for analysis. In the analysis, a significant Priority x Sternberg Control x Interval interaction ($F(4,100) = 5.78$, $p < 0.001$) illustrated the effects of the variables.

In Figure 6, it can be seen there was little evidence of intrusiveness in the low-priority Sternberg condition (on the right side of the figure). The RMS error was virtually flat across intervals in the button trials (the solid line), and showed only a slight effect in the joystick trials (the dotted line). The left side of the figure, displaying the high-priority Sternberg results, looks quite different. The overall rise in the RMS error was a by-product of the priority instructions emphasizing the Sternberg task, at the expense of the tracking. But, more important is the shape of the curves. The "U" shape may seem a bit peculiar at first, but the heightening of each "U's" left side was probably an artifact of too short an ISI. The average ISI was only 3 s, and it appears that the subjects were just recovering from one Sternberg stimulus when the next stimulus was presented. In any case, it can be seen that immediately following a Sternberg stimulus presentation, there was a rise in RMS error, and this rise appears to be steeper when the Sternberg response control was the joystick. So, the Sternberg control manipulation was successful in increasing interference with the tracking task, and this increased interference was correlated with greater sensitivity in the high-priority Sternberg condition.

In Experiment 1, this three way interaction was not statistically significant. The Experiment 1 data, when plotted in the same manner as Figure 6, generated flat lines across the five intervals, regardless of priority or memory set size. In Experiment 1, the two tasks' difficulty manipulations simply did not conflict with each other.

(Note - The method and results for Experiment 2 have been also discussed in Vidulich (in press)).

CONCLUSIONS

Taking all of the results into consideration encourages the following conclusions:

(1) The high-priority added task appears to be a viable approach for workload assessment. It appears to be more sensitive than the traditional low-priority secondary task approach, and is much less demanding to implement than full POC methodology.

(2) The sensitivity of an added task appears to be directly related to its intrusiveness. In cases where the added task did not intrude on the measured task's performance, it was also insensitive to the measured task's difficulty.

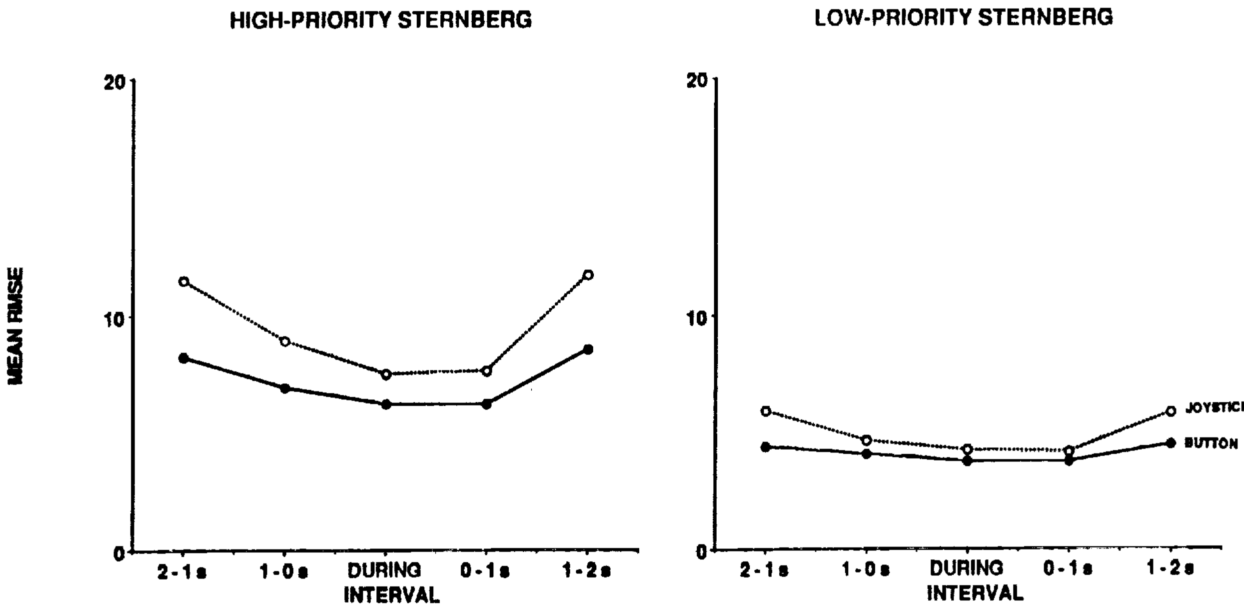

Figure 6 - Experiment 2, Sternberg task intrusiveness in tracking RMS error as a function of interval, priority, and Sternberg task control device.

And, (3) Intrusiveness seems to depend on specific forms of interference. In Experiment 1, the two tasks were too different in their processing needs, and the level of interference was insufficient to generate sensitivity, regardless of instructions.

These two studies are an insufficient basis for an unqualified endorsement of the high-priority added task alternative. But, the results are certainly encouraging enough to encourage further research. A third experiment in this series is currently being planned, in Experiment 3 an attempt will be made to modify the tracking task to make it compete more with the original Sternberg memory set size manipulation. Planned follow-ups will include study of the effects of predictability of added task time of occurrence and continuous added tasks. Either of these manipulations might reduce the unexpected abruptness of the added task occurrence and make the situation more amenable to finer voluntary control of resource allocation. If so, the traditional low-priority secondary approach may perform better.

REFERENCES

Gopher, D., and Donchin, E. (1986). Workload - A examination of the concept. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), Handbook of perception and human performance - Volume II, Cognitive process and performance (pp. 41-1 - 41-49). New York: Wiley.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Klapp, S.T., Kelly, P.A., and Netick, A. (1987). Hesitations in continuous tracking induced by a concurrent discrete task. Human Factors, 29, 327-337.

Knowles, W.B. (1963). Operator loading tasks. Human Factors, 5, 155-161.

McLeod, P. (1977). A dual-task response modality effect: Support for multiprocessor models of attention. Quarterly Journal of Experimental Psychology, 29, 651-667.

Ogden, G.R., Levine, J.M., and Eisner, E.J. (1979). Measurement of workload by secondary tasks. Human Factors, 21, 529-548.

Rolfe, J.M. (1971). The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox, and D. Whitfield (Eds.), Measurement of man at work (pp. 135-148). London: Taylor and Francis.

Shingledecker, C.A. (1984, November). A task battery for applied human performance assessment research (Tech. Rep. No. AFAMRL-TR-84-071). Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Lab.

Vidulich, M.A. (in press). Objective measures of workload: Should a secondary task be secondary? In Proceedings of the Fifth International Symposium on Aviation Psychology.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention & performance VIII (pp. 239-257). Hillsdale, NJ: Erlbaum.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, OH: Merrill.

Williges, R.C., and Wierwille, W.W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.