FINAL REPORT

VALIDATION OF VARIATIONS IN MENTAL WORKLOAD AS A FUNCTION OF SCENARIO DIFFICULTY: TRAFFIC DENSITY AND VISIBILITY

Department of Psychology
Old Dominion University
Norfolk, VA USA

Summary: Fluctuations in mental workload can be expected as a function of traffic density and visibility. The aim of the current investigation was to establish simulation scenarios that differed in attentional processing requirements. Four scenarios were created and tested representing two levels of traffic density (urban versus freeway) and two levels of visibility (clear versus foggy). An array of mental workload assessment measures were used to exam changes in attentional processing requirements in each scenario. The assessment array consisted of physiological (P300 amplitude and latency) and behavioral (RT and accuracy) indices. Preliminary results indicate that workload differs significantly as a function of traffic density in rural versus freeway scenarios. Workload also differs significantly in rural versus freeway scenarios as a function of visibility as observed by a significant interaction between the two variables of interest. Results are discussed in terms of their application for validating the difficulty level of simulation scenarios as a format for examining mental workload.

INTRODUCTION

Methods of assessing human performance in complex environments such as aviation have undergone dramatic changes in recent years. Recent advances in technological capabilities, analytical techniques, and the increasing availability of equipment for noninvasive, real-time assessment of human brain function, have led to revolutionary advances in mental workload measurement.

Measures of brain function have several advantages over alternative measurement techniques. These advantages include increased sensitivity to both transient and continuous fluctuations in mental demand without the need to introduce an additional task as well as the ability to discern the relative contributions of various brain mechanisms as a result of task dynamics. The ability of neurophysiological measures to provide real-time assessment of mental workload can facilitate both examination of task components and environment variables leading to compromised performance as well as the potential to develop adaptive automation interfaces to offset workload in high demand situations.

Neurophysiological techniques vary along several dimensions including: 1) level of invasiveness; 2) spatial resolution capabilities; and, 3) temporal resolution capabilities. EEG/ERP methods are relatively low in level of invasiveness and have strong temporal resolution capabilities. However, EEG/ERP methods lack strong spatial resolution capabilities. Conversely, PET scans and fMRI, which have strong spatial resolution capabilities, are highly invasive (lacking portably which makes them unsuitable for assessing performance in dynamic complex environments). Further, PET scans and fMRI methods lack the temporal resolution capabilities suitable for assessing workload for real time applications. Emerging optical brain imaging systems such as NIRS offer the combined advantage of low levels of invasiveness, strong temporal resolution and relatively strong spatial resolution, as compared to EEG/ERP techniques. However, further testing is required to determine the suitability of implanting NIRS assessment techniques in complex environments.

The aim of the current investigation was to establish simulation scenarios that imposed different levels of mental workload on operators. A driving simulation environment was chosen for initial examination. Similar to aviation operations, driving an automobile involves extreme fluctuations in mental workload. Attentional processing demands are placed on the driver both from outside the vehicle (i.e., traffic density and visibility) and within the vehicle (i.e., mobile phones and invehicle displays).

Mental Workload Assessment

The methods of assessing mental workload are perhaps as multifaceted as the various terms used to define it. It is not within the scope of the current report to present yet another definition, but rather simply to point out that mental workload is comprised of different aspects. These aspects include neurophysiological processes as well as perceptual and cognitive processes. Physiological, behavioral and subjective indices can be expected to assess different aspects of mental workload stemming from potentially different sources (i.e., situational and internal demands). For example, ERP components such as the P300 may be sensitive to the perceptual aspects of stimulus evaluation in a particular task (i.e., intensity, contrast) but relatively insensitive to response characteristics (Coles, Smid, Scheffers, & Otten, 1995) while response time (RT) can be expected to be sensitive to both aspects. Stimulus intensity and target probability have previously been demonstrated to increase P300 amplitude in both auditory and visual oddball sensory detection paradigms (Polich, Ellerson, & Cohen, 1996). Conversely, Kramer et al., (1987) found that performance data (RT, accuracy) generated by an auditory detection task were not sensitive to changes in simulated flight task difficulty; however, peak amplitude for the P300 component was sensitive to manipulations in flight task difficulty. If sensitive to external task demands, ERP components have the potential advantage of not requiring an overt response by the participant (Mangun & Hillyard, 1995). However, both ERP components and performance measures may fail to assess the operators' internal perception of workload.

Subjective assessment techniques provide a window into internal sources of workload. Task relevant internal aspects are captured in a definition provided by Hancock and Caird (Hancock & Caird, 1993) in which workload is conceptualized as having three dimensions. The three dimensions include: 1) time for action; 2) perceived distance from the desired goal; and 3) the level of effort required to achieve the desired goal. According to this model then, mental workload increases as the distance to the goal and time constraints increase.

The current investigation examines the sensitivity of physiological and behavioral measures to differential task demands. A dual task paradigm involving a visual detection task was performed in conjunction with simulated driving. The detection task required participants to discriminate between two stimuli of different probability. Mental workload was assessed using physiological and behavioral indices. The physiological measure included analysis of ERP components, namely P300 amplitude to target stimuli in the detection task and behavioral measures included RT and accuracy to targets. It was predicted that P300 amplitude would be greater for target versus distractor stimuli in the sensory detection task but that this amplitude differential would decrease as the mental workload of the driving task increased due to increased traffic density and poor weather conditions. It was further predicted detection accuracy would decrease and RT would increase as traffic density increased and weather conditions worsened.

METHODS

Participants

Data from fourteen participants from the university subject pool who volunteered for this investigation were included in the analysis. Participants ranged in age from 18-35 years. All were licensed drivers and reported normal hearing and visual abilities.

Simulator and Stimulus Generation Apparatus

The General Electric Capital I-Sim$^{@}$ was used to present the simulated driving task. As illustrated in Figure 1, the simulator is equipped with a full set of operational controls including steering wheel, brake and accelerator pedals as well as lights, turn signal indicators, etc. The side screens allow presentation of a 180 horizontal field of view. Side rearview mirrors allow the driver to monitor traffic from all directions.

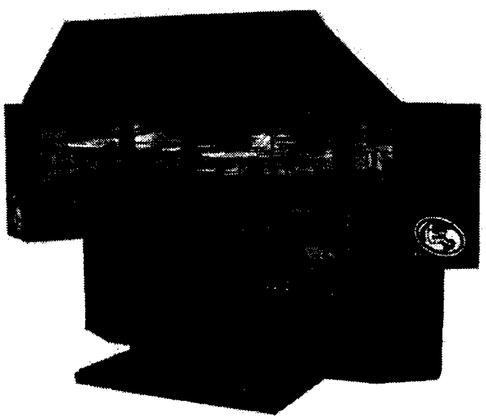

Figure 1: GE Capitol I-sim Driving Simulator

Driving Scenarios

Four urban roadway scenarios consisting of two levels of traffic density, low and moderate, and two levels of weather, clear and foggy, were presented. Each scenario lasted approximately 5 minutes.

Sensory Detection Task

Visual Detection. For the visual detection task, a color was briefly presented on the entire 15-inch viewable area of a laptop computer located to the right of the participant as he/she was seated in the driver seat (refer to Figure 1). The distractor color, green, was presented 70% of the time and the target color, red, was presented the remaining 30% of the time at a rate of 1 per 1700 ms for a duration of 50 ms. Participants were instructed to indicate the presence of the target (red) color by pressing a response button.

ERP Recording Equipment

Electroencephalographic (EEG) activity was recorded from three sites (Fz, Cz, and Pz according to the International 10/20 system) and linked mastoids were used as references. Electroocular activity recorded from electrodes placed above and below the left eye was evaluated and used to edit ocular artifacts from the EEG data file. Electrode impedances were maintained below 5 kohms. A NuAmps amplifier system in conjunction with Neuroscan 4.0 software was used for data collection and analysis.

Procedure

Participants completed an orientation scenario to acquaint them with the controls and handling characteristics of the simulated vehicle. They were then given practice on both the driving task by itself and then the detection task alone and then practice performing both tasks concurrently. Following the completion of all practice trials, participants completed the experimental blocks in counterbalanced order. Baseline trials of each of the detection tasks were included at the beginning and end of the dual task trials. Baseline driving trials (driving only trials) of both low traffic density and moderate traffic density scenarios factorially combined with the clear and foggy weather conditions were included in the counterbalanced experimental blocks. The entire experimental paradigm lasted approximately two hours.

RESULTS

Secondary Detection Task.

Event-Related Potentials. The primary focus of the ERP analysis was amplitude of the P300 component. The P300 is characterized by a large positive deflection that peaks between 300 to 500 ms. Peak amplitude was defined as the largest positive deflection between this range. A 2 (event type: target, distractor) by 3 (electrode site: Fz, Cz, Pz) repeated measures ANOVA design was used to analyze the P300 amplitude data.

ERP Detection Task Baseline Trials-P300 Amplitude. A significant main effect for event type (targets versus distractors) during baseline (ERP detection task only) trials was observed, $F(1,13) = 15.26$, $p < .05$. All subsequent analyses are for target stimuli only. P300 amplitude to target events was significantly larger than to distractor events for the visual baseline trials. There was also a significant main effect for electrode site (Fz, Cz, and Pz), $F(2,26) = 14.68$, $p < .05$. P300 amplitude increased from Fz, Cz, to Pz in the ERP detection task. The interaction of event type (target versus distractor) and electrode site was also significant, $F(2,26) = 3.51$, $p < .05$.

ERP Dection Task Workload Index-P300 Amplitude. P300 amplitude for target stimuli was used as an index of mental workload to compare the difficulty of driving in low (freeway) versus moderate (urban) traffic density combined with clear versus foggy weather conditions. For site Pz, P300 amplitude in this paradigm was not sensitive to increased task load resulting from driving in rural versus freeway traffic $F(1,10) = .076$, $p > .05$. There was a marginal effect for weather $F(1,10)=3.52$, $p< .10$, and no interaction $F(1,10)= .88$, $p > .05$. For site Cz, there was a marginal effect for road $F(1,10)= 3.47$, $p< .10$, but not for weather or the road by weather interaction $ps> .10$.

ERP Dection Task Workload Index-P300 Latency. For both site Pz and Cz, there were no significant effects for any of the variables.

Secondary-Task Performance Accuracy and response time (RT) measures were recorded for the ERP detection task. There was a significant main effect for road type (traffic density) for accuracy to the detection task, $F(1,10) = 123.9, p < .001$ and for the road type by weather interaction, $F(1,10)= 13.75, p< .005$. As seen in Figure 2, participants did best in the freeway-clear weather condition. Interestingly, they did poorest in the urban-clear weather condition, which was slightly poorer than the urban-foggy condition.

For RT there was an effect for road type, $F(1,10)= 13.44, p< .005$, with participants performing better on the freeway. There was no effect of weather or a road by weather interaction.

DISCUSSION

Preliminary results indicate that the four simulation scenarios used in the current investigation vary in attentional processing requirements and are therefore well suited for examining the feasibility of using new optical brain imaging techniques (NIRS) for assessing mental workload in complex environments. However, only the results from the performance data on the oddball task reflected these attentional differences. Accuracy in responding to the oddball stimuli was significantly better under low density, highway traffic than under higher density urban traffic. Furthermore, traffic density interacted with weather conditions such that participants performed better for the highway clear weather condition than for the poor weather condition. Interestingly, this effect was reversed for urban driving. Possibly, the poor weather condition increased the participants overall arousal level, making them more vigilant.

The ERP data was somewhat equivocal. Although there were marginal effects for weather condition (site Pz) and for traffic density (site Cz) for P300 amplitude, there was no interaction. Further, there were no effects found for P300 latency. The exact reason for these lack of effects is not clear. One potential problem may be the small number of participants. This, in part may be attributed to the loss of data from participants who did not finish the experiment due to motion sickness. Also, because of the nature of the task, there tended to be a great deal of movement artifact. While the motion sickness is a definite problem that will have to be resolved, the movement artifact would presumably be less of a problem for collecting the NIRS data since potential sources for EEG artifact (e.g. EMG) would not be a factor.

References

Coles, M. G., Smid, H. G. O. M., Scheffers, M. K., & Otten, L. J. (1995). Mental Chronmetry and the study of human information processing. In M. G. Coles (Ed.), *Electrophysiology of the Mind: Event-related brain potentials and cognition.* (pp. 86-131). Oxford: Oxford University Press.

Hancock, P. A., & Caird, J. K. (1993). Experimental evaluation of a model of mental workload. *Human Factors, 35*(3), 413-429.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assesment of operator workload during simulated flight missions. *Human Factors, 29*, 145-160.

Mangun, G. R., & Hillyard, S. A. (1995). Mechanisms and models of selective attention. In M. G. Coles (Ed.), *Electrophysiology of the Mind: Event-related Brain Potentials and Cognition* (No. 25 ed., pp. p. 40-85). Oxford: Oxford University Press.

Polich, J., Ellerson, P. C., & Cohen, J. (1996). P300, stimulus intensity, modality, and probability. *International Journal of Psychophysiology, 23*(1-2), 55-62.