Effects of Sensory Modality and Task Duration on Performance, Workload, and Stress in Sustained Attention

James L. Szalma, University of Central Florida, Orlando, Florida, **Joel S. Warm, Gerald Matthews, William N. Dember, Ernest M. Weiler,** and **Ashley Meier,** University of Cincinnati, Cincinnati, Ohio, and **F. Thomas Eggemeier,** University of Dayton, Dayton, Ohio

The workload and stress associated with a 40-min vigilance task were examined under conditions wherein observers monitored an auditory or a visual display for changes in signal duration. Global workload scores fell in the midrange of the NASA Task Load Index, with scores on the Frustration subscale increasing linearly over time. These effects were unrelated to the sensory modality of signals. However, sensory modality was a significant moderator variable for stress. Observers became more stressed over time as indexed by responses to the Dundee Stress State Questionnaire, with evidence of recovery in the auditory but not the visual condition toward the end of the watch. This result and the finding that signal detection accuracy – although equated for difficulty under alerted conditions – favored the auditory mode, indicate that display modality and time on task should be considered carefully in the design of operations requiring sustained attention in order to enhance performance and reduce stress. Actual or potential applications of this research include domains in which monitoring is a crucial part, such as baggage screening, security operations, medical monitoring, and power plant operations.

INTRODUCTION

The expanded use of automation has altered the role of operators in many work settings from that of active controllers to executive monitors who must stay attentive to displays and take action only in the event of imminent problems (Sheridan, 1987). Substantial evidence shows that the need to sustain attention or to remain vigilant imposes a high workload on operators, who also find tasks requiring such attention highly stressful (Warm, 1993). The workload and stress of monitoring tasks are of special concern for human factors specialists, given the negative impact these effects have on worker health and productivity (Nickerson, 1992).

The NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) has been the principal vehicle for assessing the workload of sustained attention. It is a multidimensional instrument that provides a reliable index of global workload and also identifies the relative contributions of six sources of workload. Three of those sources reflect the demands that tasks place upon operators (mental, physical, and temporal demand), whereas the remainder characterize the interaction between the operator and the task (performance, effort, and frustration). Global workload scores in vigilance typically fall in the middle to upper range of the scale, with Mental Demand and Frustration scales reflecting the principal workload components (Warm, Dember, & Hancock, 1996). Moreover, Dember et al. (1993) have demonstrated that the decline in performance efficiency typical of vigilance tasks – the vigilance decrement (see Davies & Parasuraman, 1982) – is accompanied by a linear rise in global workload over the course of the watch.

Although the NASA-TLX is a convenient and reliable instrument for measuring mental workload, it is essentially a subjective scale (Warm et al., 1996), and there is always some question

Address correspondence to James L. Szalma, Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390; jszalma@mail.ucf.edu. **HUMAN FACTORS,** Vol. 46, No. 2, Summer 2004, pp. 219–233.

as to whether any form of self-report accurately reflects respondents' "true" perceptual experiences (Natsoulas, 1967). In order to establish the validity of ratings of perceived workload in vigilance experiments, Warm and his associates (e.g., Becker, Warm, Dember, & Hancock, 1995; Hitchcock, Dember, Warm, Moroney, & See, 1999; Warm et al., 1996) sought to bring such workload ratings under experimental control by demonstrating that factors that degrade vigilance performance increase workload ratings, whereas factors that enhance performance diminish perceived workload. In addition to time on task, several stimulus factors that serve to degrade vigilance performance have also been found to elevate the workload of vigilance tasks. These include decreases in signal salience and increases in event rate, spatial uncertainty, and display complexity. In addition, factors that elevate signal detection, such as knowledge of results and forewarning about the imminent arrival of critical signals, reduce the workload associated with vigilance tasks.

To date, research on the workload of sustained attention has been conducted exclusively with visual tasks. However, monitoring tasks are also performed in the auditory modality, and the sensory modality of signals is not a matter of indifference where vigilance is concerned. The speed and accuracy of signal detections is greater for auditory than for visual signals, and the vigilance decrement is less pronounced for auditory than for visual tasks (Davies & Parasuraman, 1982; Warm & Jerison, 1984).

Hatfield and Loeb (1968) have shown that these sensory effects arise from differential "coupling" inherent in the auditory and visual modalities. Auditory tasks are "closely coupled" because monitors in those tasks are usually linked to a source of stimulation, either through headphones or through an enveloping sound field. Hence the monitor's physical orientation does not affect his or her receptiveness to stimuli. By contrast, monitors in visual tasks are "loosely coupled" because they are typically free to make head and eye movements, which can be incompatible with carefully observing the display; thus their physical orientation is crucial in determining their receptiveness to the stimuli to be monitored. To overcome coupling disparity, observers of visual displays, as compared with auditory displays, must maintain a relatively fixed posture that can lead to discomfort and restlessness as well as to eyestrain and tension (Galinsky, Rosa, Warm, & Dember, 1993). Efforts to work under such aversive conditions while continually observing displays for critical signals may result in an elevated level of workload for visual monitoring as compared with auditory monitoring. Accordingly, the sensory modality of signals could be an additional component to the compendium of factors responsible for the workload of sustained attention. One goal for the present study was to explore that possibility.

The stressful nature of vigilance tasks is revealed by observers' consistent reports that they feel less attentive and more bored, strained, irritated, and fatigued at the end of a vigil than prior to its start (Warm, 1993). Such reports, however, are based on instruments that tap only unidimensional aspects of stress state (e.g., fatigue, boredom), and such an approach does not adequately describe how different environmental stressors can induce differential patterns of cognitive and affective response (Hockey, 1984). Moreover, Matthews (2001) has argued that these patterns reflect the qualitatively different person-environment interactions that may occur in demanding task situations, which are supported by the different coping mechanisms that individuals use to adapt to environmental demands. In short, unidimensional measures of stress state fail to capture the multidimensional nature of the stress construct.

Recently, Matthews et al. (1999) developed the Dundee Stress State Questionnaire (DSSQ) for assessing transient states associated with mood, arousal, and fatigue. The DSSQ was designed explicitly to reflect the multidimensional nature of stress, expressed through affective, motivational, and cognitive change. This instrument consists of 77 items that yield 10 factor-analytically determined scales: Energetic Arousal (alertness-sluggishness), Tense Arousal (nervousness-relaxation), Hedonic Tone (general feelings of happiness-cheerfulness), Intrinsic Task Motivation, Self-Focused Attention (self-reflection), Self-Esteem, Concentration, Confidence and Control, Task-Relevant Cognitive Interference (worry about task performance), and Task-Irrelevant Cognitive Interference (worry about personal concerns). Validity studies have shown that the DSSQ scales are differentially sensitive to environmental stress factors and show different patterns of correlation with objective performance indices (Matthews et al., 1999; Matthews, Warm, Dember, Mizoguchi, & Smith, 2001; Matthews et al., 2002).

These scales were subjected to another, second-order factor analysis, yielding three secondary factors: task engagement, distress, and worry (Matthews et al., 1999). *Task engagement,* defined primarily by the Energy, Motivation, and Concentration scales, contrasts enthusiasm and interest with fatigue and apathy. The Tense Arousal, Hedonic Tone, and Confidence and Control scales define the *distress* factor, and the *worry* factor is defined by the Self-Focused Attention, Self-Esteem, and Concentration scales and both cognitive interference scales. Hence, stress states may be conceptualized either as 10 relatively narrowly defined scales (the approach taken in this article) or as three broader syndromes that interlink several discrete stress responses.

Several studies have employed the DSSQ to assess the stress of sustained attention. In the initial investigation, Matthews et al. (1999) reported that participation in vigilance tasks led to decreases in energetic arousal, concentration, and motivation, indicating a loss of task engagement. Hedonic tone and confidence also decreased in that study, indicating an increase in distress. These findings have been replicated in other vigilance experiments (Grier et al., 2004; Helton, Dember, Warm, & Matthews, 1999; Matthews et al., 2001; Parsons et al., 2000; Temple et al., 2000) and generalize to real-world tasks requiring sustained attention, such as prolonged simulated automobile driving (Matthews & Desmond, 2002).

Given its success in revealing the multidimensional nature of subjective stress in vigilance, the DSSQ was the instrument of choice in the present effort to explore the factors that control stress responses to sustained performance. It is likely that the stress of sustained attention arises, in part, from the need to make continuous signal/nonsignal decisions under conditions of great uncertainty and with little opportunity for situational control. Such a view is consistent with the account by Warm et al. (1996) of the origins of the workload of sustained attention. In addition, Hancock (1998) and Scerbo (1998) have suggested that the aversiveness of the task is exacerbated by the imposition of vigilance tasks upon monitors by an external authoritative agency (the experimenter in laboratory studies, management in operational settings) under conditions in which monitors feel obliged not to quit.

An account of the genesis of stress in sustained attention along these lines is global in character, however, and does little to pinpoint the specific elements of the vigilance task itself that may affect monitors' stress reactions. One task dimension that has been found to influence the subjective stress of vigilance is the sensory modality of signals. Thus Galinsky et al. (1993) found that observers engaged in a visual sustained attention task reported a greater level of stress than did those monitoring an auditory display. They suggested that the modality difference might arise because, in comparison with observers in the visual task, those in the auditory task were presumably free from the negative effects of postural constraint, tension, and eyestrain, factors discussed earlier in regard to modality differences in perceived workload.

It is worth noting, however, that the subjective stress effects described in the Galinsky et al. (1993) study were limited to self-reports of fatigue, and no attempt was made to examine whether modality-based effects extend to other dimensions of subjective stress. It may be that the modality effects on stress depend on the specific dimension of stress state that is measured, with differences emerging for some facets but not others. Accordingly, together with assessing the effects of the sensory modality of signals on perceived mental workload, the present study also sought to determine the multidimensionality of sensory-determined stress in sustained attention by means of the DSSQ. Finally, the present study was also designed to examine the manner in which self-reports of stress might change over time. To date, temporal changes in self-reported stress in vigilance have been entirely neglected. Given that time is a key element in vigilance performance, a complete understanding of the nature of self-reports of stress in vigilance requires information on their temporal course. As in the case of sensory modality, it may be that the effect of task duration on performance varies across the dimensions of stress state. For instance, over the course of a watch-keeping session, changes in scales associated with task engagement may progress at rates different from those of changes in scales associated with distress or worry.

METHOD

Participants

The participants, 256 undergraduates at the University of Cincinnati (128 men and 128 women), served in order to fulfill a course requirement. They ranged in age from 18 to 39 years, with a mean of 20.3 years. All observers had normal or corrected-to-normal vision and were free of known hearing impairments.

Experimental Design

Two sensory modalities (auditory and visual) were combined factorially with four vigil durations (10, 20, 30, or 40 min). Thirty-two observers were assigned at random to each of eight independent groups resulting from the combinations of sensory modality and vigil duration, with the restriction that the experimental conditions were equated for sex.

Apparatus and Measurement: Performance

In order to assess the effects of audio versus visual channels per se on vigilance performance, workload, and stress, it was necessary to utilize a dimension for discrimination that is common to the two modalities. Temporal discrimination was chosen for that purpose because audiovisual correlations in the discrimination of temporal intervals are substantial ($r = .90$; Loeb, Behar, & Warm, 1966), and skill in making precise temporal judgments acquired through training in one modality readily transfers to the other (Warm, Stutz, & Vassolo, 1975).

Observers monitoring the visual display viewed the repetitive presentation of a horizontally oriented 2 × 9 mm white bar that appeared against a gray background on a video display terminal (VDT). The transluminance of the bar was 37.8 cd/m^2 (as measured by a Spectra-Model UBD 1° Spot Meter), and that of the gray background was 3.49 cd/m^2. Neutral events (those requiring no overt response from the observer) were flashes lasting 247.5 ms. Observers monitoring the auditory display listened to 247.5-ms bursts of white noise presented binaurally via Grayson-Stadler TDH-39 headphones. In order to control for the effects of wearing headphones, observers in the visual conditions also wore them for the duration of the vigil.

For each participant, the apparent loudness of the noise was matched to the apparent brightness of the visual stimulus by means of a cross-modality matching procedure (Gescheider, 1997). In the visual case, critical signals for detection were brief (125-ms) flashes of the light bar. In the auditory case, critical signals were brief (200-ms) noise bursts. The disparity in the duration changes used to specify auditory and visual critical signals was necessary to compensate for the fact that temporal discrimination is more acute in the auditory mode (Dember & Warm, 1979). The values used for neutral events and critical signals in this study are identical to those used by Galinsky et al. (1993) and were verified as to their equal discriminability under alerted conditions by pilot work preceding this study. In the pilot work, a sample of 20 observers discriminated critical from neutral events over 20 trials using a two-alternative temporal forced-choice procedure.

In both modalities, stimuli were presented at a rate of 40/min by setting stimulus onset asynchrony at 1.5 s. Ten critical signals occurred within each 10-min period of watch (signal probability = .025). Observers responded to critical signals by pressing a key on a response pad attached to an Apple Power Mac computer keyboard. The computer orchestrated stimulus presentations and recorded observers' responses. Responses occurring within 1.5 s after the onset of a critical signal were recorded automatically as correct detections. All other responses were recorded as errors of commission or false alarms. Participants were tested individually in a 2.0 × 1.9 × 1.9 m Industrial Acoustics Sound Chamber. They were seated in front of the VDT, which was positioned at eye level approximately 35 cm from the observer. Ambient illumination in the chamber was 0.26 cd/m^2 and was provided by a 25-W bulb housed in a parabolic reflector located behind and to the right of the observer. The bulb served to diffuse the light evenly, minimizing glare on the VDT. Fresh air was supplied by a fan mounted in the chamber wall above and behind the observer.

Apparatus and Measurement: Perceived Workload and Stress

Perceived mental workload was measured by a computerized version of the NASA-TLX administered immediately after the vigilance session. The standard version of the NASA-TLX was used, in which participants first provided ratings on each of the six subscales and then engaged in a paired-comparison procedure to determine the relative contributions of the subscales to the global workload score. Perceived stress was measured using the DSSQ administered in two parts: a previgil questionnaire, completed prior to the initiation of the vigil, and a postvigil questionnaire, completed after the vigil. To avoid possible testing bias, half the observers in each sensory modality-vigil duration group (determined at random and equated for sex) were administered the NASA-TLX, and the remaining half completed the DSSQ.

Procedure

Upon entering the laboratory, each participant was seated in the testing chamber, and the cross-modality matching procedure was administered. The experimenter then read the instructions for the sustained attention task, and these were also presented simultaneously on the VDT. Following the instructions, half the participants completed the previgil version of the DSSQ and then engaged in a 5-min practice session with the sensory modality they would later encounter in the main portion of the session. The remaining participants proceeded directly to the practice session after reading the instructions. All participants were required to detect a minimum of 80% of the critical signals and to commit no more than 10% false alarms during practice. Observers failing to pass these criteria were provided a second 5-min practice session. The study included data from only those participants who passed the criteria on either the first or second practice session; 13 participants (5% of the total sample) could not meet the performance criteria and had to be replaced.

The mean percentage of correct detections at the end of practice was 95% for observers in each sensory modality. The main vigil commenced immediately after participants completed the qualifying practice period. Observers surrendered their wristwatches, pagers, and cell phones upon entering the laboratory and had no prior knowledge of the length of the vigil, other than that the entire session would not exceed 75 min.

RESULTS

Performance

Signal detections. Mean percentages of correct detections in the auditory and visual tasks are plotted as a function of periods of watch in Figure 1. In this analysis, and all subsequent analyses, the data for the 20-min, 30-min, and 40-min groups are based only on the final 10 min of the vigil and on independent groups of participants.

It can be seen in Figure 1 that overall detection probability was greater in the auditory (M = 89.9%) than in the visual condition (M = 70.6%), that detection probability generally declined over time, and that the vigilance decrement tended to be greater for visual as compared with auditory signals. These impressions were mostly confirmed by a 2 (modality) × 4 (time) between-groups analysis of variance (ANOVA), based on an arcsine transformation of the detection scores, which revealed significant main effects for modality, $F(1, 248) = 60.24$, $p < .001$, and time on watch, $F(3, 248) = 14.94$, $p < .001$. The interaction between these factors, however, was not significant ($p > .05$).

False alarms. False alarm rates in this study were low. The mean percentage of false alarms across all conditions was 1.1%, with five of the eight conditions having a mean false alarm rate below 1%. A 2 (modality) × 4 (time) between-groups ANOVA, based on an arcsine transformation of the false alarm scores, revealed no significant effects for modality, time, or their interaction ($p > .05$ in each case).

Workload

Global workload. Mean global workload scores on the NASA-TLX for the auditory and visual tasks are plotted in Figure 2 as a function of periods of watch. It can be seen in the figure

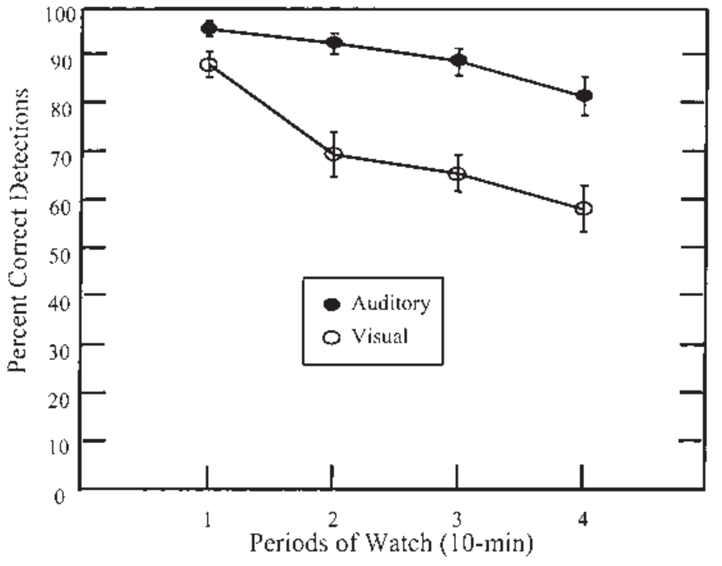

Figure 1. Percentages of correct detections in the auditory and visual tasks as a function of periods of watch. (Error bars are standard errors.)

that the global workload scores for both the auditory and the visual tasks were similar and substantial; the scores for both tasks fell within the midrange of the scale and were considerably higher than those for a 5-min card-sorting task, based on an additional 32 observers (16 male and 16 female undergraduates from the University of Cincinnati) carried out for comparison purposes. Observers sorted for suit at a rate of one card/s while timed by a computerized metronome. A 2 (modality) × 4 (time) between-groups ANOVA revealed no significant modality or time effects and no significant interaction between these factors in the global workload scores of the two vigilance tasks ($p > .05$ in each case). The finding that the vigilance-based

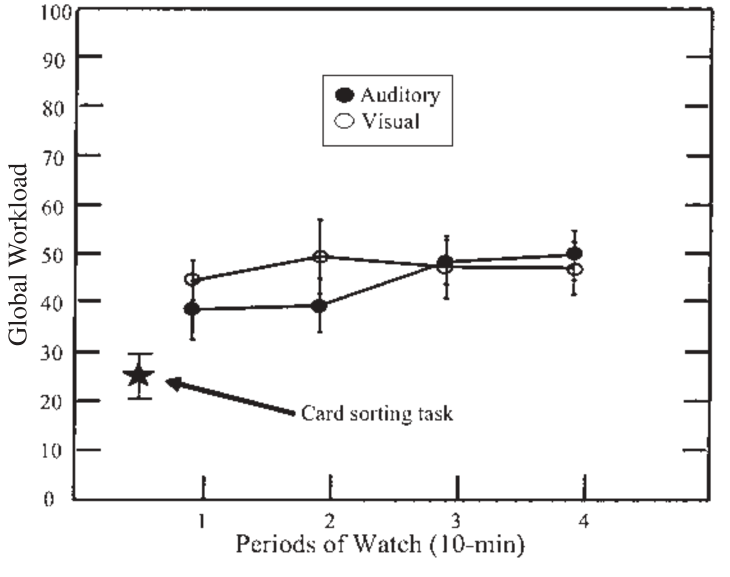

Figure 2. Global workload scores on the NASA-TLX for the auditory and visual tasks as a function of periods of watch. (Error bars are standard errors.)

workload scores in this study exceeded those of the card-sort task indicates that the NASA-TLX was sensitive to the workload imposed by the vigilance task. This in turn suggests that the lack of significant effects of modality and time in global workload scores does not reflect a general insensitivity of the NASA-TLX to the demands of the vigilance task.

Weighted workload ratings. In addition to global workload scores, mean weighted ratings on the subscales of the NASA-TLX were also determined in all experimental conditions. The Mental Demand (M = 165.4) and Frustration (M = 177.7) subscales contributed most to workload, and Physical Demand (M = 20.6) contributed least. The mean weighted workload ratings for the Temporal Demand, Performance, and Effort subscales were 118.2, 90.1, and 117.9, respectively.

The weighted ratings data were subjected to a 4 (periods) × 2 (modality) × 5 (subscales) mixed-ANOVA with repeated measures on the last factor. Because of the paired-comparison procedure used in determining the dimensional weightings, the Physical Demand subscale was dropped from the ANOVA in order to meet the independence assumption of the statistical procedure. In this analysis, and all other analyses in this study involving repeated measures, Box's epsilon was used in computing degrees of freedom to correct for violations of the sphericity assumption (Maxwell & Delaney, 2004). The overall difference among the subscales was statistically significant, $F(3, 417) = 14.75$, $p < .001$, and there was a significant Subscale × Time interaction, $F(10, 417) = 2.02$, $p < .05$. None of the remaining sources of variance in the analysis was statistically significant ($p > .05$).

Following a procedure recommended by Keppel (1991), we further examined the Subscale × Time interaction by testing the effects of time separately within each subscale. Significant time effects were found only for the Frustration subscale, $F(3, 124) = 3.92$, $p < .05$. In this case, a trend analysis indicated that the weighted ratings increased linearly over time, $F_{lin}(1, 124) = 11.42$, $p < .01$, and that there were no significant deviations from linearity, $F_{nonlin}(2, 124) < 1$, $p > .05$. The weighted ratings for the Frustration subscale are plotted as a function of periods of watch in Figure 3. A least-squares procedure was used to determine the line of best fit to the data. It can be seen in the figure that observers' weighted Frustration ratings increased at the rate of 35.9 units per 10-min period of watch.

Stress

Standard scores. All pre- and postvigil comparisons on the DSSQ were made in terms of standard scores (z scores) using the formula

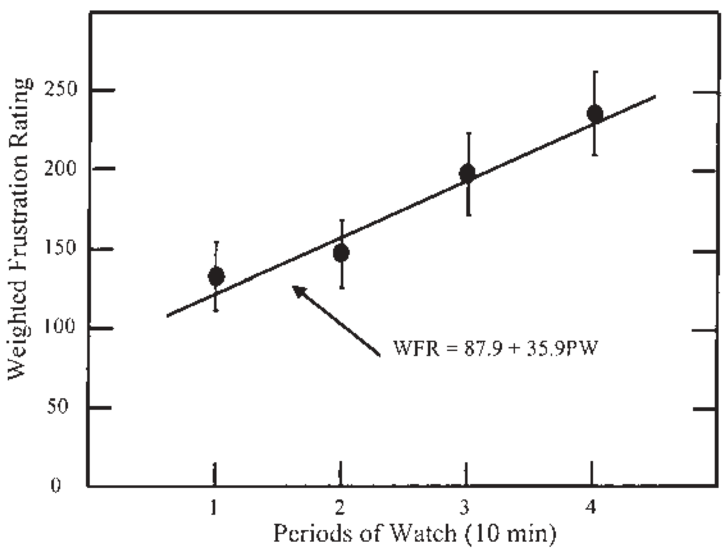

Figure 3. Weighted NASA-TLX frustration scores as a function of periods of watch. WFR = weighted Frustration rating; PW = period of watch. (Error bars are standard errors.)

(phase score – normative mean)/standard deviation of the normative sample. Normative means and standard deviations based on a large British sample were obtained from Matthews et al. (1999).

Coarse-grained analysis. Stress data from the DSSQ were analyzed at two levels: A coarse-grained analysis determined if the outcome of the present study generally conformed to previous vigilance findings with the DSSQ by comparing pre- and posttask states without analyzing time and modality effects. A fine-grained analysis examined the multidimensional effects of the sensory modality of signals and periods of watch on the stress of sustained attention within the posttask data. The two-level approach was necessary because the fine-grained temporal analysis used in the present study could potentially mask earlier findings based on a more gross analysis. Specifically, a given DSSQ scale might show a change early in the vigil but then remain constant over the watch. Such a result would emerge as a null effect on the basis of a fine-grained temporal analysis but as a meaningful effect if, as in prior studies, comparisons were made only on an overall previgil/postvigil basis. Modality was eliminated from the coarse analysis to avoid redundancy with the fine-grained examination of the data.

Mean previgil and postvigil standard scores collapsed across modalities and time periods are presented for the 10 scales of the DSSQ in Figure 4. It can be seen in Figure 4 that relative to their previgil reports, observers' postvigil scores revealed that they felt less energized, motivated, able to concentrate, happy, confident, and self-focused after the vigil than prior to its start. They also felt greater self-esteem postvigil. A 2 (phase) × 10 (scales) within-groups ANOVA revealed significant main effects for phase, $F(1, 127) = 128.01$, $p < .001$, and scales, $F(5, 611) = 20.52$, $p < .001$, and a significant Phase × Scales interaction, $F(6, 709) = 37.87$, $p < .001$. The nature of the interaction was probed by assessing the statistical significance of the phase differences on each scale by means of t tests, using an alpha level of .05 and the Bonferroni correction. All of the changes on the scales just noted reached significance; changes on the remaining scales were not significant. In terms of the secondary factors of the DSSQ, par-

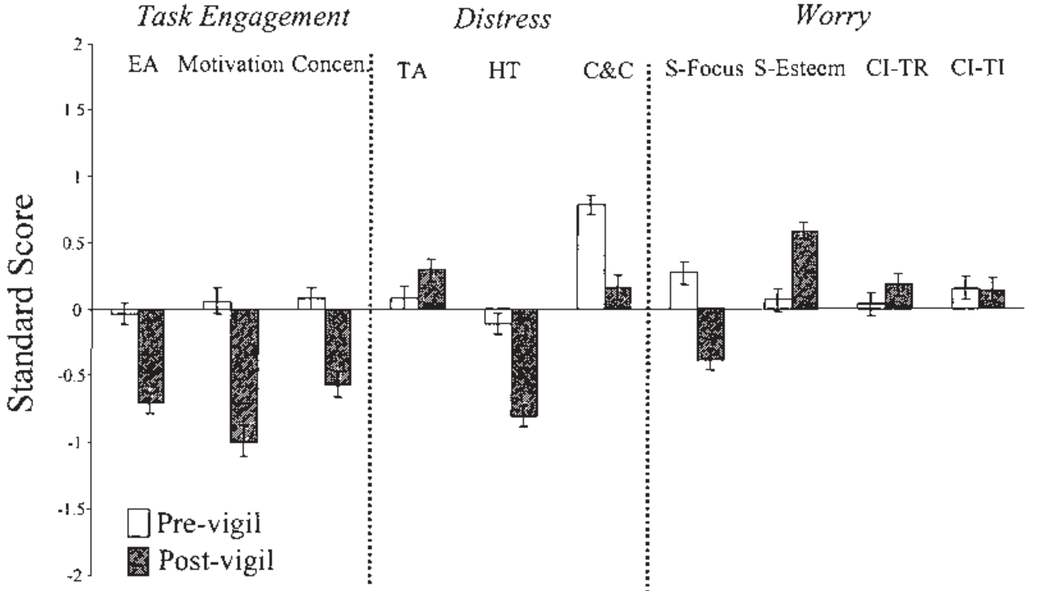

Figure 4. Mean previgil and postvigil DSSQ scores: Coarse–Grained analysis. EA = Energetic Arousal; Concen. = Concentration; TA = Tense Arousal; HT = Hedonic Tone; C&C = Confidence and Control; S-Focus = Self-Focused Attention; S-Esteem = Self-Esteem; CI-TR = Task-Related Cognitive Interference; CI-TI = Task-Irrelevant Cognitive Interference. (Error bars are standard errors.)

ticipants tended to feel more distressed (declines in the Hedonic Tone and Control and Confidence scales) and less task engaged (declines in the Energetic Arousal, Motivation, and Concentration scales) at the end of the vigil than before its start. For the most part, the coarse-grained examination of the DSSQ scores indicated that the vigilance task was stressful.

Fine-grained analysis. Prior to examining the effects of stimulus modality and time on task on the stress of sustained attention, it was necessary to demonstrate that observers in the modality and time conditions were similar to each other before the initiation of the vigil. Toward that end, mean previgil standard scores for observers who would later experience one of the two sensory modalities at one of the four task durations were examined by means of a 2 (modality) × 4 (duration) × 10 (DSSQ scales) mixed ANOVA, which showed no significant effects for modality or time, no significant interaction between these factors, and no higher-order interaction between these factors and the DSSQ scales ($p > .05$). Thus the requirement of previgil similarity in all combinations of sensory modality and task duration was clearly met.

Postvigil standard scores are presented in Figure 5. For each DSSQ scale, the scores for the two sensory modalities are plotted as a function of time on task. A 2 (modality) × 4 (time) × 10 (scales) mixed ANOVA for the postvigil data revealed a significant main effect for scales, $F(5, 604) = 42.46$, $p < .001$. Although there were no significant overall effects for sensory modality or time on task ($p > .05$ in each case), there was a significant two-way interaction between scales and time, $F(15, 604) = 2.64$, $p < .01$, and a significant three-way interaction among scales, time, and modality, $F(15, 604) = 2.11$, $p < .01$.

In order to explore the three-way interaction more fully, we performed separate Time × Modality between-groups ANOVAs on the data for each DSSQ scale. No significant main effects or interactions were observed with regard to the Energetic Arousal, Concentration, Self-Focused Attention, or Self-Esteem scales ($p > .05$ in each case). In the case of the Confidence and Control scale, the only significant effect was for time on task, $F(3, 120) = 2.77$, $p < .05$. As can be seen in Figure 5, observers' feelings of confidence and control declined over the course of the watch. In the case of the Motivation and Hedonic Tone scales, the only significant effect was for sensory modality, $F(1, 120) > 4.40$, $p < .05$ in each case. Figure 5 reveals that observers in the visual condition reported themselves as feeling less motivated ($M = -1.27$) and less happy ($M = -0.97$) postvigil than did those in the auditory condition ($M = -0.74$ and -0.65, respectively).

In the case of the Tense Arousal, Task-Related Cognitive Interference, and Task-Irrelevant Cognitive Interference scales, the scores increased significantly over time, $F(3, 120) > 3.59$, $p < .05$ in each case, and the temporal change was modified significantly by the sensory modality of signals, $F_{\text{Time} \times \text{Modality}}(3, 120) > 2.70$, $p < .05$ in each case. Inspection of Figure 5 reveals a similar pattern of changes over time with regard to the Tense Arousal scale and both types of cognitive interference. For these scales, the scores for the visual task increased monotonically over the course of the watch, whereas those for the auditory task followed an inverted-U function, first increasing and then returning to their initial level by the end of the watch.

In sum, the postvigil data reveal that both the sensory channel used for stimulus delivery and the work-time microstructure have a significant impact on the stress of sustained attention and that these effects are multidimensional in nature. Some elements of the DSSQ were not influenced by these factors in the fine-grained analysis. For other scales, only time or modality had a significant effect on feelings of task-induced stress, whereas for still other scales, signal modality was a moderator variable for the negative influence of time on watch. Note that for some scales (the Energetic Arousal, Motivation, Concentration, Hedonic Tone, Self-Focused Attention, and Self-Esteem scales), the previgil-postvigil changes observed at the coarse-grained level of analysis were not reflected in the temporal microstructure of the fine-grained analysis. Hence, for these scales, the effect of time on task occurred early in the watch and remained constant thereafter.

DISCUSSION

The purpose of this study was to explore the

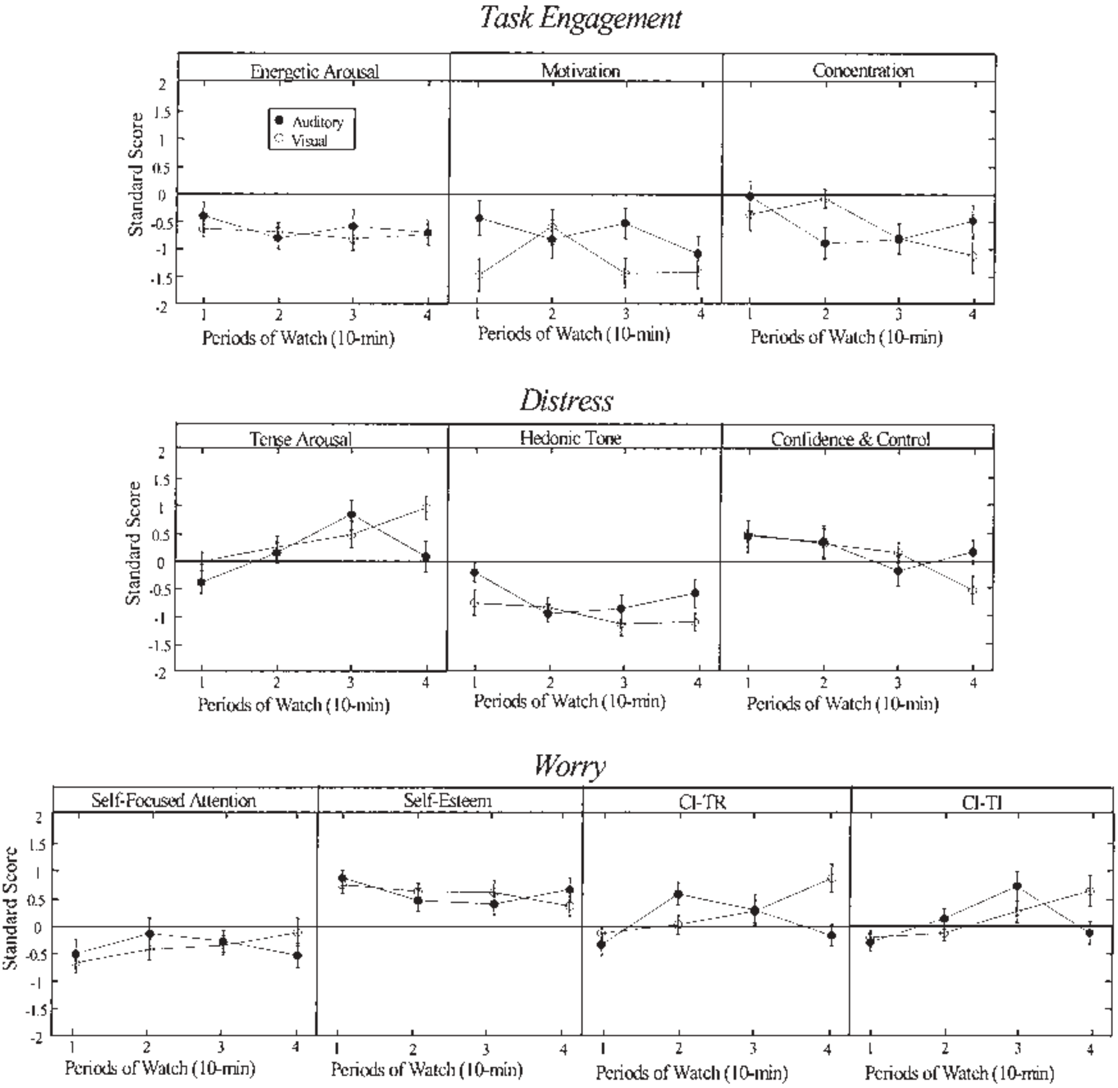

Figure 5. Mean postvigil DSSQ scores for the auditory and visual tasks as a function of periods of watch: fine-grained analysis. CI-TR = Task-Related Cognitive Interference; CI-TI = Task-Irrelevant Cognitive Interference. (Error bars are standard errors.)

effects of the sensory modality of signals and vigil duration on performance efficiency in a sustained attention task and on the dimensions of perceived mental workload and subjective stress that accompany vigilance performance. Essentially, the study was an effort to fill in gaps in the tapestry of research on these dimensions of sustained attention. Both modality and time have been studied extensively with regard to performance efficiency. However, whereas temporal factors have been examined with regard to workload, sensory effects have been ignored, and whereas sensory effects have been examined in regard to subjective stress, the possibility of systematic change over time has been ignored. In general, the results of this experiment provide additional support for the view that vigilance tasks present a major challenge to observers (Davies & Parasuraman, 1982; Parasuraman, Warm, & Dember, 1987; Warm et al., 1996), a challenge that is reflected in the complex roles that sensory and temporal factors play in performance efficiency, workload, and stress.

Performance and Workload

Consistent with prior vigilance studies, performance efficiency declined over the course of the watch and favored the auditory modality (Davies & Parasuraman, 1982; Warm & Jerison, 1984). These effects occurred even though the tasks were carefully equated for discrimination difficulty. Therefore these effects are likely to be rooted in factors other than differential discrimination per se, such as the postural constraint, discomfort, and restlessness that arise from observers' efforts to compensate for the loosely coupled nature of visual vigilance tasks (Galinsky et al., 1993; Hatfield & Loeb, 1968) and the eyestrain and tension that have been found to be associated with the use of VDTs (Galinsky et al.).

To be sure, it was the possibility that observers in the visual condition would have to work through the combination of those physical symptoms that led to the expectation that perceived mental workload would be elevated for visual as compared with auditory monitoring. That expectation, however, was not confirmed. As in previous vigilance studies (Warm et al., 1996), the overall scores on the NASA-TLX were substantial, but workload was unrelated to the sensory modality of signals. The absence of modality effects in the global workload scores represents a dissociation between workload and performance that is at variance with several prior experiments in which performance changes associated with psychophysical manipulations were closely mirrored by variations in perceived workload (Becker et al., 1995; Hitchcock et al., 1999; Warm et al., 1996).

Such a dissociation may be based on Yeh and Wickens' (1988) finding that perceived mental workload is driven to a considerable extent by the demands imposed on working memory. Given that the auditory and visual tasks both required difficulty-equated absolute judgments (or successive-type judgments; see Davies & Parasuraman, 1982) and that signal duration discrimination is based upon common mechanisms for both audition and vision (Eijkman & Vendrik, 1965; Loeb et al., 1966; Warm et al., 1975), the absence of modality effects on perceived workload may reflect the similar working memory demands of these tasks. A signal discrimination explanation of the modality component of the performance/workload dissociation implies that for the tasks employed herein, in which a common central processor and similar working memory demands were featured, the sensory channels used for stimulus presentation may not be differential factors in the workload of sustained attention. It remains to be determined, however, whether workload similarities in auditory and visual vigilance tasks extend to cases in which the perceptual dimensions involved are not homologous within channels, such as brightness and loudness (Eijkman & Vendrik, 1965).

As is typical in vigilance experiments (Davies & Parasuraman, 1982), the probability of signal detection in this study declined significantly over time. However, contrary to the earlier findings by Dember et al. (1993), the temporal decline in performance was not accompanied by a rise in global workload over the course of the watch. As in the case of the sensory modality of signals, this additional performance/workload dissociation may also be related to the type of discriminations involved. A visual task requiring spatial discriminations with a scanning imperative was employed in the earlier investigation by Dember et al., whereas temporal discriminations, which did not feature a scanning imperative, were required herein. It is possible that the differing demands imposed by environmental scanning could have led to the temporal changes noted in the early study as compared with this one.

Although global workload scores in this study remained stable over time, a marked temporal change was observed with the Frustration subscale of the NASA-TLX: Frustration increased linearly over time. Dember et al. (1993) also observed a significant rise in frustration with time on task. A possible explanation for this effect may lie in the resource model of vigilance performance. Parasuraman and his associates (Davies & Parasuraman, 1982; Parasuraman et al., 1987) have argued that because of the participant's need to make continuous signal/noise discriminations, the performance decrement reflects the depletion of information-processing resources that cannot be replenished in the time available. This argument has been supported by the finding that cerebral blood flow declines over time in a manner paralleling the performance decrement (Hitchcock et al., 2003; Mayleben et al., 1998; Schnittger, Johannes, Arnavaz, & Munte, 1997). In addition, Yeh and Wickens (1988) have suggested that the inability to summon resources leads to increased frustration. Hence the decrement in performance and the increment in frustration may originate from a common process: the decline of resource availability with time on watch.

Given such a temporally based change in the Frustration subscale, one might also wonder why there was no concomitant change in the global workload scores. Recall that the Frustration subscale is one of six independent subscales that contribute to global workload on the NASA-TLX. Thus its contribution to the temporal characteristics of the global score could have been vitiated by the stability of scores of the other five scales. Note that Dember et al. (1993) also reported that the Frustration subscale was the only one to show significant change over time. However, they observed a rise in global workload with time on task. Moreover, previous investigations have found that other workload manipulations in vigilance (e.g. event rate, signal saliency) influence the NASA-TLX subscales, demonstrating that these scales are sensitive to the effects of sustained attention (see Warm et al., 1996). Therefore it is unlikely that the absence of time effects for global workload in this study was attributable solely to the dilution of these effects by the other five scales.

Along with the finding of increased frustration over time, Frustration and Mental Demand emerged as the principal contributors to workload in the present study. The latter outcome is consistent with that of many previous investigations, indicating that there may be a typical workload profile or signature that reflects the particular demands imposed by vigilance tasks (Warm et al., 1996). According to Eggemeier and his associates (Eggemeier, Wilson, Kramer, & Damos, 1991; O'Donnell & Eggemeier, 1986), diagnosticity, or the ability of a measure to provide information about component factors in workload, is an important property of workload assessment techniques. Given the findings regarding the gain in frustration over time and the roles of frustration and mental demand as the principal contributors to workload, the results of this study join those of several other vigilance studies in testifying to the ability of the NASA-TLX to meet the diagnosticity criterion.

Stress

The results with the DSSQ confirmed earlier findings that vigilance tasks are stressful. They also indicated that the subjective stress of sustained attention is multidimensional in character and tied to both the sensory modality of signals and time on watch. The multidimensional nature of stress was revealed by findings at both the coarse-grained and fine-grained levels of analysis, showing that observers' pattern of responding to participation in the vigil was not uniform across the scales of the DSSQ. At the coarse-grained level of analysis, the state changes suggested that observers were less task engaged (e.g., reduced Energetic Arousal scores) and more distressed (e.g., lower Hedonic Tone scores) after the vigil than prior to its start. However, participation in the vigil had little impact on scales linked to worry, with the exception of the Self-Esteem scale, which increased after the vigil for observers in all conditions. These results are generally consistent with those of prior studies using the DSSQ to assess the stress of sustained attention (Grier et al., 2004; Helton et al., 1999; Matthews et al., 1998, 1999, 2001, 2002; Parsons et al., 2000; Temple et al., 2000).

The influence of the sensory modality of signals and time on watch on observers' stress reports was revealed through the fine-grained analysis of the DSSQ scales. As in the case of the coarse-grained analysis, the fine-grained analysis revealed that the reports of stress were not uniform across scales. Some, such as the Energetic Arousal, Concentration, Self-Focused Attention, and Self-Esteem scales, showed no effects for modality or time over the course of the watch. Others, such as the Motivation and Hedonic Tone scales, were sensitive to modality but not to time. Loss of motivation and happiness were greater for the visual than for the auditory task. Still other scales – the Confidence and Control, Tense Arousal, Task-Related Cognitive Interference, and Task-Irrelevant Cognitive Interference scales – showed systematic change with time on watch. Moreover, in the case of the last three scales, the temporal effect was modified by stimulus modality. Clearly, stress changes are not uniform in their temporal microstructure, a result that was masked in the many earlier studies of the stress of sustained attention that focused only on previgil-postvigil differences.

The diffuse pattern of changes on the DSSQ scales to the stress of sustained attention tasks reaffirms the notion that stress states require multidimensional description (Hockey, 1984). A key insight into task-induced stress is that operators actively regulate their handling of task demands in stressful environments by the use of different coping strategies (Lazarus & Folkman, 1984; Matthews, 2001; Matthews & Campbell, 1998; Stanton & Young, 2000). Included are task- or problem-focused coping, in which behaviors are directed toward constructive action to change the stressful situation; emotion-focused coping, which is designed to regulate the emotional consequences of a stressful situation; and avoidance coping, through which operators divert attention away from the task. Apparently all three coping strategies were operative in the efforts of the present observers to come to terms with the stress of sustained attention, given that a decrease in task engagement, such as that noted in this study, implies a loss of problem-focused coping (Matthews & Campbell), whereas an increase in distress implies use of emotion-focused and, to a lesser extent, avoidance coping styles (Matthews et al., 1999; Matthews et al., 2002).

The temporal microstructure of stress revealed by the fine-grained analysis may also be a consequence of changes in coping induced by task factors. Hence the temporal decline in confidence and control may reflect the inability of observers to generate task-focused strategies that successfully replenish their diminishing processing resources. As discussed earlier, the rise in frustration on the NASA-TLX may also result from this process.

The finding of modality-moderated changes over time in regard to the Tense Arousal scale and both cognitive interference scales may also reflect differences in the development of effective coping strategies. As Hancock and Warm (1989) and Matthews (2001) have pointed out, the implementation of coping strategies requires information-processing resources. The need to handle the disrupting physical symptoms uniquely associated with visual displays may serve to drain resources from the development of adaptive coping strategies in the visual modality. In contrast, observers in the auditory condition, who were presumably freer of these physical symptoms and attendant resource demands, were more successful in their efforts to cope with task stressors. However, it may have taken some time to formulate an effective coping strategy. The declines in distress (tense arousal) and worry (both types of cognitive interference) late in the vigil for those in the auditory group, which suggest a shift away from emotion-focused coping, indicate that by the end of the experimental session the vigilance task, although generally unpleasant and frustrating, was no longer personally threatening or disturbing to observers in the auditory condition (see Matthews et al., 1999, 2002).

Conclusions and Implications

In sum, the results of this study indicate that the workload of vigilance is substantial, and they underscore the importance of assessing perceived workload as well as performance when designing systems that include tasks requiring sustained attention. Along that line, this study provides further evidence for the utility of analyzing the subscales of the NASA-TLX as well as global workload scores and strengthens the view that multidimensional assessment of perceived workload is essential in order to capture the complexity of task effects on operators.

The findings of this study confirm that sustained attention is stressful and that sensory and temporal components of the task have differing effects on the various components of the subjective stress response. The modality effect affirms that assessment of stress state in operational settings picks up information in addition to that provided by the workload index, which was not influenced by display modality. Thus even short-duration tasks may provoke stress reactions that are potentially harmful to performance (see Matthews, 2001). However, in general, using the auditory channel reduces the stress of vigilance relative to a visual task, as far as the Distress scales (e.g., tense arousal) and Worry scales (e.g., cognitive interference) are concerned, provided that task duration is sufficient for recovery in the auditory channel late in the vigil.

The sensitivity of the stress response to psychophysical properties of the task implies that vigilance tasks may be intrinsically stressful, and so adverse reactions will generalize to real-world settings. At the same time, the dependence of response on coping strategy (Matthews & Campbell, 1998; Matthews et al., 2002) implies that the operator has some voluntary control over the stress response. Hence, in real-world settings, there are several possible techniques for alleviating stress associated with sustained attention. First, displays may be designed to reduce stress, taking into account the duration of the work period: Auditory presentation may be beneficial for longer task durations. Second, macroergonomic interventions that raise employee morale and interest in the task may be beneficial; Matthews et al. (2002) found that job perceptions correlated with job stress state during performance. Third, explicit training in coping strategies that maintain engagement, and that minimize distress and worry, may be useful.

ACKNOWLEDGMENTS

Completion of this work was facilitated by the Department of Defense Multidisciplinary University Research Initiative (MURI) program administered by the Army Research Office (Grant #DAAD19-01-1-0621). P. A. Hancock is the principal investigator on the grant. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. government. The authors wish to thank Dr. Sherry Tove, Dr. Elmar Schmeisser, Dr. Mike Drillings, and Dr. Paul Gade for providing administrative and technical direction for the grant.

REFERENCES

Becker, A. B., Warm, J. S., Dember, W. N., & Hancock, P. A. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *International Journal of Aviation Psychology, 5,* 49–62.

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance.* London: Academic.

Dember, W. N., & Warm, J. S. (1979). *Psychology of perception* (2nd ed.). New York: Holt, Rinehart, and Winston.

Dember, W. N., Warm, J. S., Nelson, W. T., Simons, K. G., Hancock, P. A., & Gluckman, J. P. (1993). The rate of gain of perceived workload in sustained attention. In *Proceedings of the Human Factors Society 37th Annual Meeting* (pp. 1388–1392). Santa Monica, CA: Human Factors and Ergonomics Society.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., & Damos, D. L. (1991). Workload assessment in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 207–216). London: Taylor & Francis.

Eijkman, E., & Vendrik, A. J. H. (1965). Can a sensory system be specified by its internal noise? *Journal of the Acoustical Society of America, 37,* 1102–1109.

Galinsky, T. L., Rosa, R. R., Warm, J. S., & Dember, W. N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors, 35,* 603–614.

Gescheider, G. A. (1997). *Psychophysics: The fundamentals* (3rd ed.). Mahwah, NJ: Erlbaum.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2004). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors, 45,* 349–359.

Hancock, P. A. (1998). The price of freedom. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1577–1578). Santa Monica, CA: Human Factors and Ergonomics Society.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hatfield, J. L., & Loeb, M. (1968). Sense mode and coupling in a vigilance task. *Perception & Psychophysics, 4,* 29–36.

Helton, W. S., Dember, W. N., Warm, J. S., & Matthews, G. (1999). Optimism-pessimism and false failure feedback: Effects on vigilance performance. *Current Psychology, 18,* 311–325.

Hitchcock, E. M., Dember, W. N., Warm, J. S., Moroney, B. W., & See, J. E. (1999). Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors, 41,* 365–372.

Hitchcock, E. M., Warm, J. S., Matthews, G. M., Dember, W. N., Shear, P. K., Tripp, L. D., et al. (2003). Automation cueing modulates cerebral blood flow and vigilance in a simulated air traffic control task. *Theoretical Issues in Ergonomics Science, 4,* 89–112.

Hockey, R. (1984). Varieties of attentional state: The effects of environment. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 449–483). New York: Academic.

Keppel, G. (1991). *Design and analysis: A researcher's handbook.* (3rd ed.) Englewood Cliffs, NJ: Prentice Hall.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer-Verlag.

Loeb, M., Behar, I., & Warm, J. S. (1966). Cross-modal correlations of the perceived durations of auditory and visual stimuli. *Psychonomic Science, 6,* 87.

Matthews, G. (2001). Levels of transaction: A cognitive science framework for operator stress. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 5–33). Mahwah, NJ: Erlbaum.

Matthews, G., & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. In *Proceedings of the Human Factors and Ergonomic Society 42nd Annual Meeting* (pp. 821–825). Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G., Campbell, S. E., Desmond, P. A., Huggins, J., Falconer, S., & Joyner, L. A. (1998). Assessment of task-induced state change: Stress, fatigue, and workload components. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance* (pp. 199–203). Mahwah, NJ: Erlbaum.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L., Huggins, J., Gilliland, K., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2,* 315–340.

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *Quarterly Journal of Experimental Psychology, 55A,* 659–686.

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). *Validation of a comprehensive stress state questionnaire: Towards a state "big three"?* In I. Mervielde, I. J. Deary, F. DeFruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg, Netherlands: Tilburg University Press.

Matthews, G., Warm, J. S., Dember, W. N., Mizoguchi, H., & Smith, A. P. (2001). The common cold impairs visual attention, psychomotor performance and task engagement. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1377–1381). Santa Monica, CA: Human factors and Ergonomics Society.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective.* (2nd ed.) Mahwah, NJ: Erlbaum.

Mayleben, D. W., Warm, J. S., Dember, W. N., Rosa, R. R., Shear, P. K., Temple, J. G., et al. (1998, March). *Cerebral bloodflow velocity during sustained attention.* Paper presented at the Third Annual Conference on Automation and Human Performance, Norfolk, VA.

Natsoulas, T. (1967). What are perceptual reports all about? *Psychological Bulletin, 67,* 249–272.

Nickerson, R. S. (1992). *Looking ahead: Human factors challenges in a changing world.* Hillsdale, NJ: Erlbaum.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human perception and performance: Vol. 2. Cognitive processes and performance* (pp. 42/1–42/49). New York: Wiley.

Parasuraman, R., Warm, J. S., & Dember, W. N. (1987). Vigilance: Taxonomy and utility. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), *Ergonomics and human factors: Recent research* (pp. 11–32). New York: Springer-Verlag.

Parsons, K. S., Warm, J. S., Matthews, G., Dember, W. N., Galinsky, T. L., & Hitchcock, E. M. (2000, November). *Changes in signal probability and vigilance performance.* Poster presented at the meeting of the Psychonomic Society, New Orleans, LA.

Scerbo, M. W. (1998). Fifty years of vigilance research: Where are we now? In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 1575–1576). Santa Monica, CA: Human Factors and Ergonomics Society.

Schnittger, C., Johannes, S., Arnavaz, A., & Munte, T. F. (1997). Relation of cerebral blood flow velocity and level of vigilance in humans. *NeuroReport, 8,* 1637–1639.

Sheridan, T. (1987). Supervisory control. In G. Salvendy, (Ed.), *Handbook of human factors* (pp. 1243–1268). New York: Wiley.

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science, 1,* 315–331.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors, 42,* 183–194.

Warm, J. S. (1993). Vigilance and target detection. In B. M. Huey & C. D. Wickens (Eds.), *Workload transition: Implications for individual and team performance* (pp. 139–170). Washington, DC: National Academy.

Warm, J. S., Dember, W. N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Hillsdale, NJ: Erlbaum.

Warm, J. S., & Jerison, H. J. (1984). The psychophysics of vigilance. In J. S. Warm (Ed.), *Sustained attention in human performance* (pp. 15–59). Chichester, UK: Wiley.

Warm, J. S., Stutz, R. M., & Vassolo, P. A. (1975). Intermodal transfer in temporal discrimination. *Perception & Psychophysics, 18,* 281–286.

Yeh, Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30,* 111–120.

James L. Szalma is a postdoctoral researcher at the University of Central Florida. He received his Ph.D. in experimental psychology/human factors from the University of Cincinnati in 1999.

Joel S. Warm is professor of psychology and director of graduate studies in psychology at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Alabama in 1966.

Gerald Matthews is professor of psychology at the University of Cincinnati. He received his Ph.D. from the University of Cambridge in 1984.

William N. Dember is dean emeritus and professor emeritus at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Michigan in 1955.

Ernest M. Weiler is professor of communication science/disorders in the College of Allied Health Sciences at the University of Cincinnati. He received his Ph.D. in experimental psychology from the University of Louisville in 1969.

Ashley Meier received her B.A. in psychology from the University of Cincinnati in 1999.

F. Thomas Eggemeier is professor of psychology and associate dean in the College of Arts and Sciences at the University of Dayton. He received his Ph.D. in experimental psychology from Ohio State University in 1971.

Date received: January 29, 2002
Date accepted: December 2, 2002