THE EFFECT OF INCREASED MONITORING LOAD ON VIGILANCE PERFORMANCE USING A SIMULATED RADAR DISPLAY

Richard I. Thackray
J. Powell Bailey
R. Mark Touchstone
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma

July 1977

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, D.C. 20591
THE EFFECT OF INCREASED MONITORING LOAD ON VIGILANCE PERFORMANCE USING A SIMULATED RADAR DISPLAY.

Richard I. Thackray, J. Powell/Bailey, and R. Mark/Touchstone

Office of Aviation Medicine
Federal Aviation Administration
800 Independence Avenue, S.W.
Washington, D.C. 20591

Work was performed under Task AM-C-77-PSY-61.

The present study examined the extent to which level of target density influences the ability to sustain attention for a complex monitoring task requiring only a detection response to stimulus change. The visual display was designed to approximate a futuristic, highly automated air traffic control radar display containing computer-generated alphanumeric symbols. Forty-eight male subjects, equally divided into three groups, were exposed to density levels of 4, 8, or 16 targets. Ten critical stimuli (signals) were randomly presented during each half-hour of the 2-hour session. Detection latency to the critical stimuli in the 16-target condition was significantly greater than latency to the 4- and 8-target conditions. There was no evidence of performance decrement in the two lower density conditions. The 16-target condition showed a significant progressive increase in mean detection latency, which was primarily the result of an increase in long latencies. The hypothesized decline in attention associated with this condition appeared to be independent of any major change in arousal level.

Key Words
Air Traffic Control Vigilance
Arousal
Attention
Automation
Performance

Distribution Statement
Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

Security Classification of this Report
Unclassified

Security Classification of this Page
Unclassified

No. of Pages
14

Price
22
THE EFFECT OF INCREASED MONITORING LOAD ON VIGILANCE PERFORMANCE USING A SIMULATED RADAR DISPLAY

I. Introduction.

Air traffic control systems are becoming increasingly automated. Assuming that this trend will continue as projected, the future radar controller may well find his primary role will be that of a system monitor rather than an active participant in traffic control. Under such conditions, controller intervention would be required only when certain types of malfunctions occur or system limits are exceeded that cannot be automatically detected and remedied by the computer (4). Since a malfunction or deviation from system limits not detected by the computer could be a quite infrequent occurrence, the controller would be required to maintain a high degree of sustained attention to a task in which he intervened only occasionally. Leaving aside the problems of boredom and job dissatisfaction that might result (12), the question of whether a controller can maintain the required level of sustained attention under these task conditions is a significant one.

Laboratory studies of prolonged performance on vigilance tasks typically employ near-threshold, short-duration signals and commonly reveal a progressive increase in errors, primarily of omission. The increase is generally quite rapid initially and tends to level off (approach asymptote) within the first 30 minutes of work. While this phenomenon, which has become known as the "decrement function," is based almost entirely on results obtained by using simple tasks, it is often assumed to apply to any vigilance task regardless of complexity. This assumption, however, has been subject to intermittent criticism over the past 10 to 15 years. Elliott (8), who was one of the first of the critics, noted that decrement of this form was never found in studies of closely simulated asdic or radar search. He states that while a slow, continuous degradation of performance may occasionally occur during 2-hour watches, more commonly no such "fatigue" effects are observed. Other critics have pointed out that modern operational monitoring tasks involving suprathreshold long-duration signals, extensive scanning of multiple stimuli, and complex decisional processes are so different from the simple vigilance task that to assume the existence of a comparable decrement function is unwarranted on the basis of our present knowledge (18, 22).

Much of the criticism directed toward assuming the existence of a decrement function in complex operational monitoring tasks stems from the series of studies by Adams and his coworkers at the University of
Illinois (1). The primary intent of these studies was to examine variables affecting complex monitoring performance, and most of the studies employed a task that simulated a semiautomated radar surveillance system. Alphanumeric symbols were rear projected onto a screen and the filmstrip was advanced every 4 seconds. In one of their experiments (2), subjects (2) were exposed to either 6 or 36 symbols; the required response was a button press whenever a particular symbol occurred. Critical stimuli were rarely missed in either of the two density conditions. While detection latency was significantly longer in the condition of greater visual load, performance decrement was the same in both conditions. The increase in latency, over a 3-hour period, amounted to approximately 1 second. The above authors interpret their results as supporting the general conclusion that any decrement found in operational monitoring tasks would not be of practical significance for most systems.

Another recent series of studies presents evidence that vigilance decrements do occur in complex monitoring tasks and that the magnitude of decrement may well be of practical importance (13). All of these studies employed a cathode-ray-tube (CRT) display in which Ss observed an 8x8 matrix of computer-generated alphanumeric symbols. The required response was a button press each time a symbol was randomly added to or removed from the display. In one of the studies, four stimulus densities—4, 8, 16, and 32—were compared under two levels of signal frequency (36 and 90 per hour). Although all conditions were accompanied by increases in mean detection latency over the 100-minute session, by far the most pronounced increase occurred under the maximum-density, low-signal-frequency condition, in which mean latency increased from 8 to 22 seconds. An analysis of this particular condition revealed that the increase in mean latency was primarily the result of a progressive increase, over successive time segments, in the duration of maximum or longest latencies. Minimum latencies showed no change during the test session. This pattern of change, which is really a reflection of increasingly skewed latency distributions, is a pattern that has been found to occur during prolonged performance in a variety of monotonous tasks (5, 10, 17, 24, 25). Although other explanations are possible, the most likely explanation for this pattern is that it is a reflection of increasingly frequent lapses of attention (3, 6, 13).

As noted earlier, whether radar controllers can sustain adequate levels of attention to advanced display systems that may require very infrequent intervention is an important question. The studies just presented suggest that the question has not yet been adequately resolved.

The increase in mean latency reported by Howell, Johnston, and Goldstein (13) for their high-density, low-signal-frequency condition was substantially larger than the 1-second increase reported by Adams, Stenson, and Humes (2). More striking, however, was the increase (from 24 to 46 seconds) in long detection latencies found by Howell et al. (13); if this increase reflects lapses of attention, it could have real operational significance.
The major intent of the present study was to provide further information relative to the magnitude of performance decrement associated with monitoring a complex visual display. More specifically, the study sought to determine whether performance decrement was a function of target density. Howell et al. (13) found evidence that it was, while Adams et al. (2) concluded that it was not.

In addition to performance, a number of physiological and subjective measures were also included in this study. These measures were some of the same ones employed in a previous study dealing with performance correlates of reported boredom and monotony (23) and were included for exploratory purposes. We hoped they might serve to suggest possible reasons for any performance difference found among the various target density groups.

II. Method.

Subjects. Forty-eight paid male university students were randomly assigned to three groups of equal size. The groups differed only in the number of targets—4, 8, or 16—to which they were exposed. Half the Ss in each group were tested in the morning and half in the afternoon. None had any prior experience with the task used.

Apparatus and Task Design. All task programming and recording of responses were accomplished by using a Digital Equipment Corporation (DEC) PDP 11/40 computer. The computer was interfaced with a VT-11 (DEC) 17-inch CRT, which served as the S's display. The CRT was located in a console designed to resemble an air traffic control radar unit. The stimuli (targets) consisted of small rectangular "blips" representing the locations of given aircraft. Adjacent to each target, and connected to it by means of a short diagonal line, was an alphanumeric data block. Data blocks comprised two rows of symbols: the top row, consisting of two letters and three numerals, identified the aircraft, while the bottom row of six numerals indicated its altitude and speed. The first three of these numerals gave altitude in hundreds of feet and the last three gave groundspeed. For a given target, the alphanumeric identifying the aircraft, its altitude and groundspeed, as well as its entry and exit points were randomly determined except for the following restrictions: (i) altitudes had to fall within the range of 180 to 600 (in hundreds of feet); (ii) groundspeeds had to fall within the range of 400 to 580 (in knots); (iii) the entry and exit points for a given target could not be separated by less than 30°; and (iv) all targets on the display had to have at least 2,000 feet of vertical separation.

At the beginning of an experimental session, the screen contained either 4, 8, or 16 targets, depending upon the condition to which the S was assigned. A simulated radar sweepline, the scale radius of which equaled 60 miles, made one complete clockwise revolution every 6 seconds. A target was updated as to location and any change in its data block moments after the sweepline passed the target's prior location. Targets normally moved in a
linear fashion unless a course change was programmed to avoid target overlaps. The overall impression was one of a pattern of targets moving in discrete jumps as the sweepline passed. This movement approximates very closely the way in which targets are updated in contemporary air traffic control radars with computer-generated alphanumeric displays. The critical stimulus or signal to which the S was instructed to respond consisted of 999 appearing in the altitude block. Ten such critical stimuli appeared in each half-hour period; five occurred in the first 15 minutes and five in the second. The mean intersignal interval was 3 minutes. Time of critical stimulus occurrence and the target in which it occurred were randomly determined with the restriction that two targets could not contain critical stimuli at the same time. The S's response to a critical stimulus consisted of pressing a button located on the console and then holding a light pen over the critical target. The light pen caused the altitude portion of the data block to revert to its previous value. If the S failed to detect a critical stimulus within 1 minute, the data block automatically reverted to its previous value.

The computer and other recording apparatus were located in an adjacent room from which the S was visible through a one-way mirror. Indirect lighting was used in the S's room, and the level of illumination at the

Figure 1. The simulated air traffic control console with a typical stimulus pattern. Only the lower left button was used.
display was 2 foot-candles. This level approximates that used in operational air traffic control environments. Figure 1 shows the S's console with a typical stimulus pattern displayed on the CRT.

A Beckman Type H Dynograph was used to record skin conductance, heart rate, body movement, blood pressure, and vertical eye movements and blinks. Except for eye movements, the transducers and electrode placements used for recording have been fully covered in a previous study (23) and will not be described here. The eye electrodes were Beckman miniature biopotential electrodes and were attached directly above and below the right eye. The outputs were AC-coupled to the Dynograph by using a 10-second time constant. In addition to recording the physiological measures, the Dynograph recorded on one channel the output of a photocell recessed in the console shelf, which was used to monitor the S's hand placement. Marker channels on the recorder also signaled the onset of a critical stimulus and the occurrence of the required button press. Outputs from a cardiotachometer and a pulse integrator for body movement were led to the digital inputs of the computer for subsequent analysis.

**Procedure.** On arrival the S was taken to the testing room, given orientation instructions, and then instrumented for physiological recording. Following this a nine-point subjective rating scale was administered that dealt with present feelings of attentiveness, fatigue, strain, boredom, and irritation.

The task instructions emphasized the necessity of pressing the button immediately on detection of a critical stimulus. The S was told that the critical stimulus represented some form of malfunction not detected by the computerized radar system. Following the taped instructions, the S was given a 3-minute practice period containing three critical stimuli. After the practice period three blood pressure recordings were obtained and the 2-hour test period was initiated. The S's watch was taken from him before the test period began.

After the 2-hour test period, three measures of blood pressure were again taken and the S completed a second form of the subjective rating scale. This form was identical to the first except that the S was asked to rate each item, plus one additional item dealing with task monotony, on the basis of how the S felt near the end of the test period just completed.

**Measurement of the Performance and Physiological Data.** Performance data were computer processed and the following measures obtained for each 30-minute time period:

(i) Mean response latencies to critical stimuli correctly identified.
(ii) Number of button presses without a critical stimulus.
(iii) Number of critical stimuli missed.
(iv) Number of light pen hits to a critical stimulus without a preceding button press.

The computer program described in a previous study (25) was used to obtain the mean and standard deviation of heart rate for each successive 5-minute period. These were then averaged to give values for the four 30-minute periods. A separate computer program summed the number of pulses from the body movement integrator for each 30-minute period. Conductance levels for each half-hour period were averages based on measures obtained at the beginning and end of the 30-minute periods. The systolic and diastolic blood pressure recordings taken at the beginning and end of the task period were measured according to the procedure described by Thackray, Bailey, and Touchstone (23). Vertical eye movement-eyeblink recordings were visually scanned for evidence of any eye closures during the intervals following each critical stimulus.

III. Results.

Mean detection latencies for the three target density conditions are shown in Figure 2. Analysis of variance applied to these data yielded a significant (p < .05 throughout) main effect for density, F(2/45) = 37.56, but no significant main effect for the half-hour periods, F(3/135) = 2.26.

![Figure 2. Mean detection latencies for the three target density conditions.](image-url)
The density-by-periods interaction, however, was significant, F(6/135) = 2.39. Mean detection latencies for the 4-, 8-, and 16-target conditions were 2.5, 3.7, and 7.6 seconds respectively. A Newman-Keuls test of differences between these means revealed the 16-target condition to differ significantly from both the 4- and 8-target conditions, but there was no significant difference between these latter two density conditions. Tests on the simple effects of the interaction indicated that only the 16-target condition increased significantly in mean detection latency over the half-hour periods, (F3/135) = 6.55.

Since the 16-target condition was the only one to show evidence of significant decrement, it was selected for detailed examination. Each individual's longest and shortest latencies in each of the half-hour time periods were obtained and the means computed. These values are shown in Figure 3. Trend analyses revealed a significant linear component for longest latencies, F(1/45) = 4.64, but not for shortest latencies, F(1/45) = 3.73. This finding suggests that the increase in mean response latency for the 16-target condition shown in Figure 2 was apparently the result of a progressive increase in long response latencies rather than a general increase in all latencies.

Errors of omission, commission, and procedure were virtually nonexistent over all conditions. Two button presses to critical stimuli were made

![Figure 3. Mean minimum and maximum detection latencies for the 16-target condition.](attachment:image.png)
without subsequent light pen confirmation, two light pen confirmations were made without preceding button presses, and one critical stimulus was not detected within the 1-minute period.

Recordings obtained from the eye electrodes were visually examined for any evidence of apparent eye closures during the periods between each critical stimulus occurrence and the detection response (button press). Two Ss in the 4-target, four in the 8-target, and six in the 16-target condition showed evidence of eye closures during these periods. The general pattern was that of a single eye closure lasting an average of 2.4 seconds. Of the 34 detection latencies accompanied by eye closures, 27 were found to exceed the mean latency to critical stimuli occurring within the same half-hour measurement period. Of these 27, 9 were long enough to have been identified by the computer program as "maximum latencies"; 5 of these 9 occurred under the 16-target condition. Not surprisingly, eye closures tended to increase during the task session. Eleven percent of all stimulus periods containing closures occurred during the first half-hour period, 23 percent during the second, 37 percent during the third, and 28 percent during the fourth.

Analyses of variance conducted on the physiological data revealed significant changes across periods (or, in the case of blood pressure, between pretest and posttest measures) for all variables except diastolic blood pressure. Heart rate variability and body movement increased over the measurement periods, while heart rate, conductance level, and systolic blood pressure decreased. There were no significant differences between conditions and no significant interaction effects. Since no between-group or interaction effects were significant, the obtained physiological data are not presented.

Analyses of variance were also applied to the subjective rating scale variables. Significant differences between measurement periods were obtained for boredom, irritation, fatigue, strain, and attentiveness. Attentiveness decreased, while all the rest increased. As with the physiological data, there were no significant differences between density conditions and no significant interactions. A separate one-way analysis of variance of the monotony data likewise revealed no significant difference between conditions. Statements on the scales corresponding to the mean ratings for the above variables at the completion of the task period suggested that the Ss were only slightly bored, were mildly annoyed, felt more tired than usual, felt more tense than usual, and felt less stimulated than usual.

*It was relatively easy to visually discern the differences between the recordings of eyeblinks and vertical eye movements. An eye closure was assumed if a pen deflection occurred that was equal to or greater than the average blink amplitude and lasted for 1 second or longer. There was no way of determining, however, whether the eyes were completely or only partly closed during these times. Consequently, all references to eye closures should be taken to mean that there was evidence that the eyes were at least partly closed.
experienced little strain, and considered themselves reasonably attentive. The task itself was judged to be only slightly monotonous. As with the physiological data, obtained values are not presented because of a lack of significant between-group and interaction effects.

IV. Discussion.

Although mean detection latency of the eight-target condition (3.7 seconds) exceeded that of the four-target condition (2.5 seconds), the difference was not found to be significant. Nor was there any evidence of performance decrement over the time period examined.

As noted earlier, several studies of complex monitoring have employed long-duration signals with similar density levels and, although they differ in a number of respects from the present study, it is of interest to compare findings. Both Thackray et al. (23) and Adams et al. (2) used filmstrip simulation of contemporary radar displays with target densities of eight and six respectively. Mean detection latency was 2.9 seconds in the former study and 2.4 seconds (estimated) in the latter. A direct comparison of the findings of the study reported by Howell et al. (13) is more difficult, however, since their monitoring task incorporated an additional short term memory requirement. However, if one uses their low frequency (36 signals/hour) "add" condition, which most nearly corresponds to the conditions of the above studies, mean detection latencies for stimulus densities of four and eight were 3.9 and 4.9 seconds (estimated) respectively. Only Adams et al. (2) obtained any evidence of performance decrement at these density levels, and the increase in latency obtained in their study only amounted to approximately 1 second over a 3-hour period.

Combining the above data across studies gives an overall mean detection latency of 3.4 seconds (SD = 0.95 second) for densities of four to eight stimuli. Thus, from a practical standpoint, it would appear that detection of simple stimulus change, while monitoring four to eight targets, is quite rapid, is relatively uniform over a variety of task conditions, and shows virtually no evidence of the performance decrement associated with simple vigilance tasks.

With a stimulus density of 16 targets, however, the present study revealed a progressive, significant increase in mean detection latency across the 2-hour period. This increase was found to be primarily the result of an increase in long detection latencies, which, as indicated earlier, appear to reflect lapses of attention.

Why evidence of declining attention in both the present study and the previously described study of Howell et al. (13) should be confined to high stimulus density conditions is an interesting question. One possibility involves the change in ratio of signals to observing responses as density is increased. Howell et al. (13) found that the decrement associated
with high densities could be abolished if signal frequency was approximately doubled. As these authors note, their findings are quite consistent with Jerison and Pickett's (14) decision theory, which emphasizes the role of reinforcement in determining the observer's "decision" about whether to observe or be attentive to stimulus events. According to Jerison and Pickett (14), a high proportion of signals to observing responses results in a high rate of reinforcement, and little or no decrement is obtained. If the proportion is low, relatively few observing responses are reinforced and the S becomes increasingly inattentive. In the context of a complex monitoring task, this could imply either (i) that scanning behavior would be reduced under conditions in which there is a low proportion of signals to observing responses or (ii) that it would be maintained, but Ss might fail during scanning to "see" the critical changes or signals.

Presumably, this theory could also be used to explain the results of the present study. However, it does not appear to provide any explanation for the results obtained by Adams et al. (2). It will be recalled that these authors employed target densities of 6 and 36 stimuli while keeping the signal rate constant at 12 per hour. On the basis of the above theory, one would expect a sizeable decrement under the greater density condition, yet the minimal decrement obtained was the same under both conditions. No ready explanation is apparent for the difference in findings of their study relative to the findings of the present study and those of Howell et al. (13).

Data obtained from the eye electrode recordings suggest that eye closures tended to occur with greater frequency in the 16-target condition and to increase in number during the course of the session. Further, long detection latencies appeared to be more frequently associated with eye closures in the 16-target condition than in either of the lower density conditions. Failure to respond to the critical stimuli because the S's eyes were closed, however, was probably not a principal reason for the long detection times. (It should be recalled that the average duration of eye closure during a critical stimulus-response period was only 2.4 seconds, while average maximum detection latencies for the 16-target condition ranged from approximately 17 to 23 seconds.) Rather the eye closure data suggest a reduction in alertness that was probably greater in the 16-target condition than in the other conditions.

While Jerison and Pickett's (14) theory may provide the most satisfactory explanation for the apparent decline in alertness associated with the greater density condition, one puzzling aspect remains. With the possible exception of frequency of eye closures, which is suggestive at best, the 16-target condition was found not to differ from the lower density conditions on any of the other physiological and subjective measures of alertness, arousal, or attention that were employed. Heart rate variability, for example, has been found to vary inversely with behavioral measures of attention in numerous studies (7, 9, 15, 16, 19, 21, 23, 25). Yet this measure increased equally for all conditions. If alertness and attention did decline more in the greater density condition, one can only conclude that these physiological and subjective measures were too insensitive to detect the difference.
Haider (11) has recently outlined a concept of a hierarchical system of activation in which selective attention to a task may begin to show behavioral evidence of change or decline with little or no change in arousal or activation level. Only as selective attention continues to decline (habituate) and show more pronounced oscillations are phasic and, finally, tonic activation reactions involved. Since these reactions involve different levels of the central nervous system, their manifestations would differ not only in degree, but in kind. Extending Haider's (11) formulations, the many factors known to affect vigilance performance (e.g., signal frequency, intersignal interval, signal duration, task duration, rest pauses, knowledge of results, personality variables, fatigue) would presumably affect the course or rate of decline in selective attention and, hence, the extent and type of involvement of the activating system.

It is proposed that the decrement associated with the 16-target condition involved a decline in attention that was apparently independent of any major change in arousal, although some changes in the EEG might have been noted had this measure been employed (20). Further research is needed to better understand the nature of the variables that contribute to and are associated with performance decrement in complex monitoring tasks.

V. Conclusions.

The results of the present study are consistent with those of other studies of complex monitoring in which low target densities (four to eight targets) and critical signal rates of 12 to 36 per hour have been employed. Latency of detection of simple stimulus change was found to be quite uniform across the studies reviewed (Mn = 3.4 seconds; SD = 0.96 second) with little or no evidence of performance decrement. This finding is rather striking when one considers that the various studies differed on a number of potentially relevant variables; e.g., ratio of signals to stimuli or targets, type of display employed, length of prior $S$ training, task duration. It would suggest that scanning a typical radar display for the purpose of detecting relatively infrequent but readily observable and identifiable critical changes can be performed quite rapidly and efficiently when the visual load falls within this density range.

There is less agreement among studies with regard to higher target densities. While a general increase in detection latency is to be expected with increases in target density, evidence of performance decrement associated with higher densities has not been obtained universally. The present study found a significant increase in mean detection latency across the 2-hour period that appeared to be due primarily to an increase in long detection latencies. Long latencies increased from a mean of 17 seconds in the first half-hour to 23 seconds in the last. It was hypothesized that this increase reflected a progressive increase in lapses of attention during the session. However, as noted, not all previous studies have found a similar decline in attention associated with high target densities. Consequently, it
would be premature to speculate extensively as to the implications of the present findings for future radar tasks. The obtained data suggest the possibility that observers monitoring highly automated systems under conditions of high target density, with probably far less intervention than employed in this study, may find it difficult to maintain the necessary high level of sustained attention required for rapid, consistent detection of target changes. Additional research to confirm or extend the present findings is needed before further speculations or recommendations would be warranted.
References


