DOT/FAA/AM-94/08

Office of Aviation Medicine
Washington, D.C. 20591

Scanning and Monitoring Performance: Effects of the Reinforcement Values of the Events Being Monitored

P.G. Rasmussen
A.M. Revels

Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City, Oklahoma 73125

April 1994

Final Report

U.S. Department
of Transportation

Federal Aviation
Administration

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-94/8	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Scanning and Monitoring Performance: Effects of the Reinforcement Values of the Events Being Monitored		5. Report Date April 1994
		6. Performing Organization Code
7. Author(s) P.G. Rasmussen and A.M. Revzin		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, DC 20591		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplemental Notes This work was performed under task AM-A-91-PHY-146		

16. Abstract

We formulated a hypothesis suggesting that operators could make scanning and monitoring errors if they tended to concentrate on a "high-value" display sub-area while ignoring "low-value" problems elsewhere on the display. Such "data" would have application to Air Traffic Control Specialist (ATCS) jobs. We tested the hypothesis in an experiment rewarding good performance in a laboratory task. Subjects monitored two visual display "work areas" with defined task difficulty. In the high-value work area, each error cost the subjects four or ten times as much as in the low-value work area. The data obtained suggest that differing task error penalties, or reinforcement values, can induce a greater than usual frequency of errors in some subjects. Rewarding good performance in two-work area tests without differing error penalties did not induce significant error rate differences, nor did such rewards significantly affect total task performance levels. This was true even in tests where such differential attention could benefit the subject's overall performance score, thereby increasing subject's performance bonus. However, about 15% of our subjects showed a marked tendency to concentrate their attention on a display sub-area having high-value events while periodically ignoring events elsewhere on the display. Such information may be useful in reducing the frequency of scanning errors by revising training protocols or personnel selection criteria.

17. Key Words Vision; Monitoring; Air Traffic Control; Visual displays; Reinforcement	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	10	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SCANNING AND MONITORING PERFORMANCE: EFFECTS OF THE REINFORCEMENT VALUES OF THE EVENTS BEING MONITORED

INTRODUCTION

Scanning and monitoring errors may increase if Air Traffic Control Specialists (ATCSs) focus or "lock" their attention onto a limited area of their control display to the exclusion of other relevant parts of the display. Such problems are seldom discussed in standard aviation references (1), but have received attention in the visual process and control literature (2). Because such errors can seriously compromise air safety, we initiated research to identify factors that could affect the occurrence of "locking" behavior and the concomitant error rate. The reviewed literature yielded no studies that appeared immediately relevant to the effects of target values, error costs, and rewards or penalties (reinforcements) on monitoring performance. Therefore, we decided to investigate whether target value or error costs are demonstrable factors in inducing locking. Such information may be useful in reducing the frequency of scanning errors by revising training protocols or personnel selection criteria.

Given a test where two work areas had similar task difficulty, but sharply different penalties for an error, we hypothesized that a reward for good performance would tend to cause locking on the task with the highest penalties.

METHODS

Equipment

A locally developed character recognition and scanning performance test system was used to generate the display as well as record and categorize subjects' (Ss) responses.

All programs were written in Borland's Turbo Pascal™ programming language and run on a standard IBM PC-AT™ micro-computer with an 8 MHz. clock, a standard EGA adapter and a 13" diagonal 640 by 350 pixel color monitor. A previous study (3) showed that the test produced results that were congruous with those found in other, more complex, tests of character recognition and scanning performance.

Task Description

The test required the Ss to visually monitor two 100 by 100 pixel "work areas" horizontally aligned either 3 or 12 deg. of arc apart, inner-edge to inner-edge along the midline of the computer display, as shown in Figure 1.

Each work area was filled with a changing random dot pattern, each dot being one pixel. As the dots were replaced, they slowly overwrote the whole work area. The dot replacement rate was 750 pixels per second per work area. At random intervals, the characters S, B, 0, 3, 5 and 8, which share some similar shape characteristics (4), were written somewhere in each work area within a 7 by 7 pixel array. The Ss were to indicate with simple keyboard inputs when, and in which work area, the target character "5" appeared before the next character was written to the same work area. When Ss made either type of error the system provided feedback to Ss by sounding a short beep or tone through the system speaker.

In Figure 1, the "3" near the center of the left work area was written 0.5 sec. before the frame was captured. It is still clearly legible. The top center of the same work area shows the remains of another "3", and the bottom right corner of the right work area shows the remains of a "5," both written 2.5 sec. before the screen was captured. Both of these characters are on the verge of becoming illegible.

At the intended viewing distance of 60 cm., each pixel subtended 2 min. of arc, the character 15 min. of arc, and each work area 3 deg. 20 min. of arc. Observed viewing distances varied between 40 and 70 cm. Preliminary attempts to fix the S's head position to provide a relatively constant viewing distance proved impractical, considering the length of the 80-minute test period.

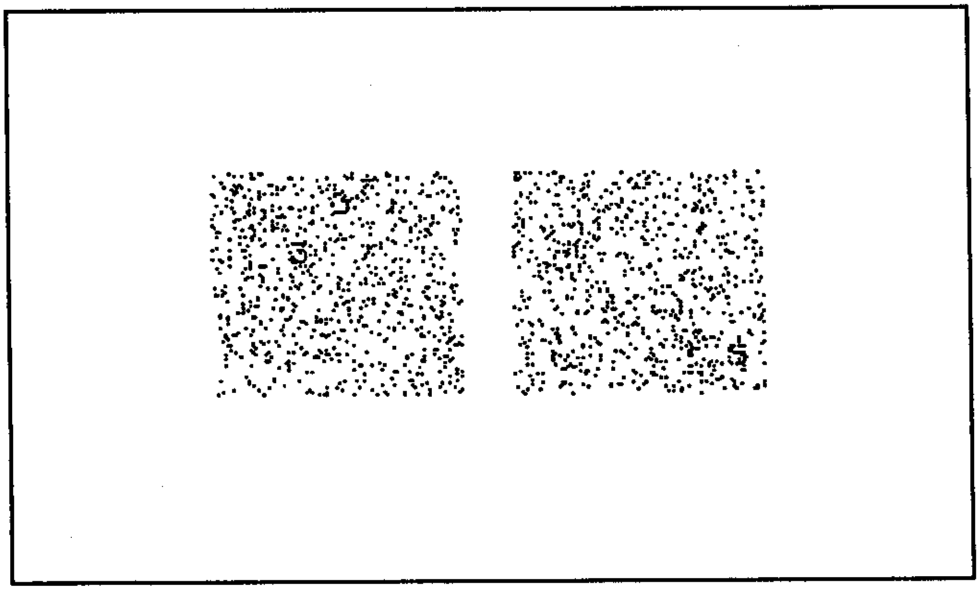

Figure 1

Example of screen display for 3° separation (shown in reverse contrast for clarity).

	First Segment	Second Segment	Third Segment	Fourth Segment
Group 1 8 Ss	Spacing 12 ° High Load/Value	Spacing 3 ° Low Load/Value	Spacing 12 ° Low Load/Value	Spacing 3 ° High Load/Value
Group 2 8 Ss	3 ° High	12 ° Low	3 ° Low	12 ° High
Group 3 8 Ss	3 ° Low	3 ° High	12 ° High	12 ° Low
Group 4 8Ss	12 ° Low	12 ° High	3 ° High	3 ° Low

Table 1. Experimental design and protocol.

Procedure

In this system, task difficulty, or "workload," is a function of the separation between work areas, and symbol presentation rates within each work area. Work area spacings and symbol presentation rates were varied according to a simple latin square design, as shown in Table 1. This enabled us to test for, and evaluate, learning or fatigue effects and any serial order interactions, none of which became evident.

Since the symbol presentations were controlled by a random number generator, the total number of symbols (including targets) and the number of target symbols presented in each test segment were not constant for all work areas and segments nor for all Ss. Based on the data for one experiment, the combined number of symbol presentations per segment for the two work areas ranged from 1005 to 1007, with individual work areas ranging from 500 to 518 symbols each. The number of target presentations was somewhat more variable, ranging from 70 to 134 per work area, or 239 to 254 per segment.

Therefore, to equalize data structure for all Ss, all detection and recognition error data were expressed as a percentage of the number of symbols presented in each test segment. We were primarily interested in the comparison of detection and recognition error frequencies for each S. Since the percentage data distributions were non-normal, all statistical analyses were run using the non-parametric Wilcoxon Matched Pairs Test, as implemented with the NCSS™ statistical program set.

Detection and recognition error classification was critical to these experiments. If a S locked onto any one work area, there should be a relative increase in detection errors (misses) in the other work area, since symbols occurring there would not be seen. However, recognition errors in both work areas would be similar since, once a symbol is seen, there should be no difference in the symbol recognition and response selection processes in the two work areas.

Errors were classified as *detection errors* when there was no response between a target symbol presentation and the display of the next symbol in any one work area. We assumed that the Ss had not detected that a target symbol was present. Errors were classified as *recognition errors* whenever a response was made to a non-target character. In this case, we assumed that the S had seen a symbol but had not accurately recognized it, or had made an error in recognition-response coupling. At the end of the test session, the program calculated a score, which was displayed for each S's information and for reward calculations.

Subjects

Thirty-two paid volunteer Ss were used in each experiment. The purpose and nature of the experiment was explained to them and they were advised that they could withdraw from participation at any time. They were tested to ensure that their correctable visual acuity was 20/25 or better. Their right/left eye dominance was also determined to be able to test for and evaluate any potential positional effects. Their ages ranged from 18 to 30 years; there were 67 men and 61 women.

Design of Experiments

Four sets of experiments were executed using the described test system:

Experiment 1. This tested the hypothesis that overall performance reinforcement would have little effect. The Ss were rewarded if their percent of correct responses to the total number of target presentations at the end of the 80-minute test period was above the group median, based on prior unrewarded test runs during an earlier study (3). We set identical task difficulties and error penalties of one point per error in the work areas. We used two groups of 16 Ss, one group composed of Ss having had some previous exposure to an identical display, the other of Ss who had no such prior experience.

Experiment 2. The second experiment tested the hypothesis that, to minimize penalties, the Ss would tend to lock onto the work area having the greatest task difficulty. It was identical to the first experiment, except that the work area task difficulties differed; the high-load work area had symbol and dot replacement rates twice that of the low-load work area. Though the high-load and low-load work areas were not specifically identified, preliminary tests demonstrated that the Ss determined (within seconds) which was which by the relative speed of the dot and character replacement rates.

Experiment 3. This tested our primary hypothesis that Ss would tend to lock onto the work area where errors were penalized the most in order to maximize their score (performance) and thus improve their chances of earning a reward (performance bonus). The protocol was identical to that in Experiment 1 above, except that:

(i) The performance bonus was split so that Ss scoring in the top quartile of a representative range of scores obtained from previous experimental sessions got a bonus equal to 50% of their guaranteed minimum earnings, while those in the next quartile received 25%;

(ii) Four points per error were deducted from the S's score in the clearly marked high-value work area, while one point per error was deducted in the low-value work area.

Experiment 4. This was identical to Experiment 3 except that 10 points per error were deducted in the high-value work area and 1 point per error in the low-value work area.

Upon completion of the vision testing, the S was seated at the display station and the general room lighting was dimmed. Pre-test instructions, including the possibility of a bonus for good performance, were read and explained to the S. The Ss were advised to pay constant attention to the display and ignore any distractions to improve their chance for earning a bonus. Based on the results from the earlier study (3), which showed the task to be unaffected by practice or previous experience, Ss were not given any task training or practice prior to the experimental task.

RESULTS AND DISCUSSION

General Findings

There was no significant correlation at the p = 0.10 level of confidence between any performance variable and S's age, sex, corrected visual acuity, or right/left eye dominance. There was also no significant correlation between an S's detection and recognition error performance. That is, for any S, a high or low error rate in the one did not necessarily mean a similar rate in the other. A summary of our findings is presented in Table 2.

Experiment 1

This experiment confirmed the test hypothesis. The reward (reinforcement) did not significantly affect performance when compared to results from our previous studies, where good performance was not rewarded. This conforms with a number of other studies

Experiment	Avg. Total % D/E * Rounded values	Avg. Total % R/E * Rounded values	Low D/E * Versus High D/E Significant @ p<0.10?	Low R/E * Versus High R/E Significant @ p<0.10?
1	9 ± 3	9 ± 2	No	No
2	19 ± 5	9 ± 3	No	No
3	22 ± 6	10 ± 3	Yes	No
4	21 ± 6	10 ± 4	Yes	No

Table 2. Summary of results.

* D/E = Detection Errors.
R/E = Recognition Errors.
Low/High D/E = Low/High Workload/Value Detection Error
Low/High R/E = Low/High Workload/Value Recognition Errors.

also indicating that human task performance may not be directly affected by a delayed reinforcement paradigm. In effect, the immediate problems of task performance override any awareness of rewards to be earned "later."

There were also no significant differences in error rates between the two work areas, nor were there significant differences in performance between the 16 experienced and 16 inexperienced Ss.

Experiment 2

The results contradicted the test hypothesis. Differing workloads had no effect on locking. There was no significant difference in either detection or recognition errors between the high- and low-workload work areas.

The results suggest, as in the first experiment, the relative ineffectiveness of a delayed reward paradigm for this protocol. That is, the task was difficult enough, or interesting enough to fully occupy the S's attention with the task, and they were not really aware of the connections between their actions and the promise of a reward at the end of the test. It is also possible that the Ss did not perceive the workload differential, a factor of about two, as significant enough to elicit the anticipated differential attention.

Experiment 3

The results partially confirmed the hypothesis that the Ss would tend to lock onto the high error value work area. The percentage of detection errors was

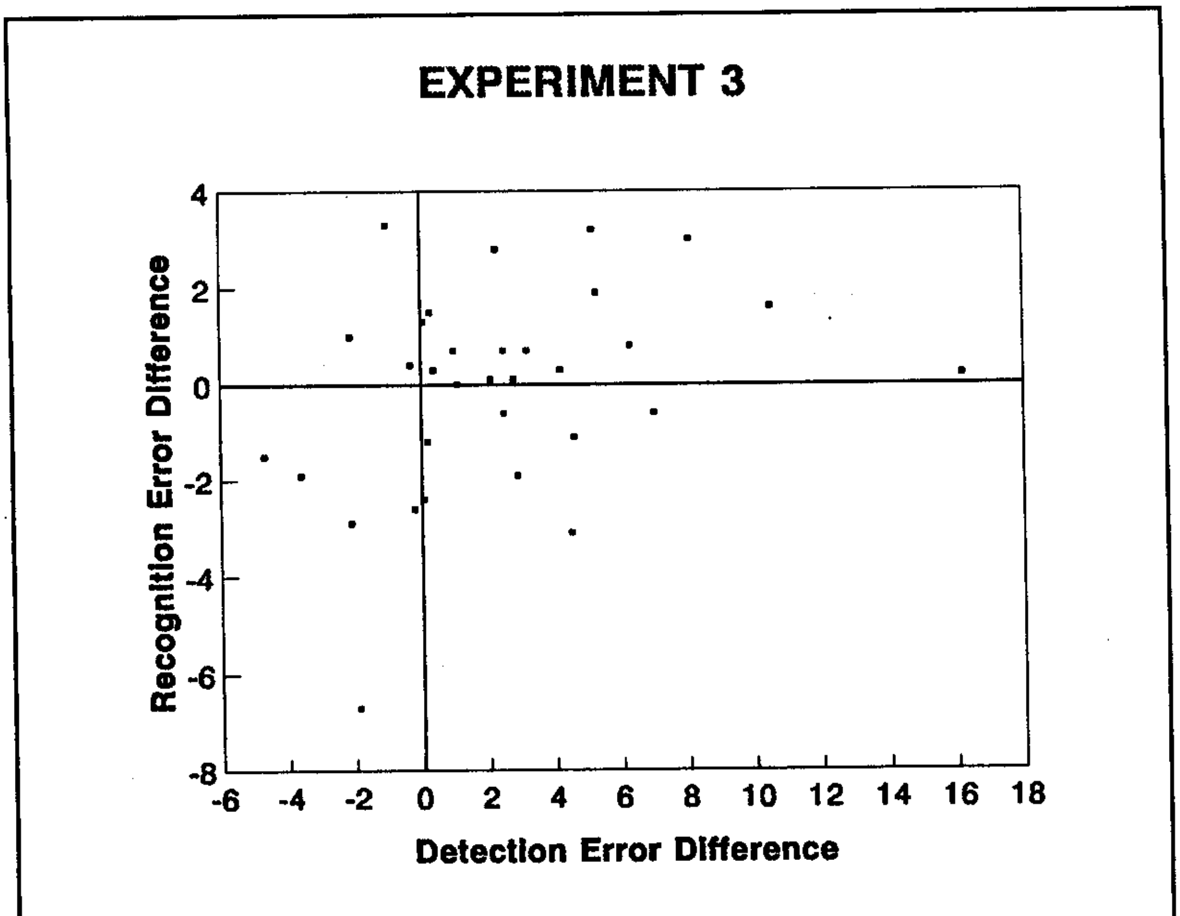

Figure 2

The Y axis was formed from the difference between an S's low- and high-value work area recognition error percentages. The X axis was similarly calculated, except it applies to detection error rates. Positive values mean that low-value work area errors were greater than those for the high-value work area.

statistically higher in the low-value work area ($p < 0.004$). The absence of significant differences in recognition errors between the two work areas remains to be explained. This was true for most Ss, but not all. This difference in individual scanning strategies deserves close attention in future research, as it may reflect basic differences in scanning ability.

In addition, the percentage of detection errors was roughly double the number of recognition errors in both work areas. This difference was also significant ($p < 0.001$), suggesting that the workload was sufficiently high to require the S's full attention for performance. That is, the time required to detect, recognize and respond to a target was long enough that some targets were not detected in the time window available, irrespective of work area. Therefore, boredom, or overall inattention, probably were not major factors in producing the performance differences seen.

We also studied the distribution of detection and recognition errors for each S. In Figure 2, we have plotted, on an XY graph, the differences between low- and high-value work area detection and recognition error percentages. The range of error differences, for most Ss and for both error types, is about ± 4. Nine Ss show a markedly high percentage of low-value work area detection errors. These differences among Ss may reflect differing score maximizing strategies, which are influenced and controlled by learning. However, the data may indicate that the increased tendency toward locking behavior reflected significant individual ability differences. If confirmed, this could provide a practical methodology for personnel selection in the air traffic control system.

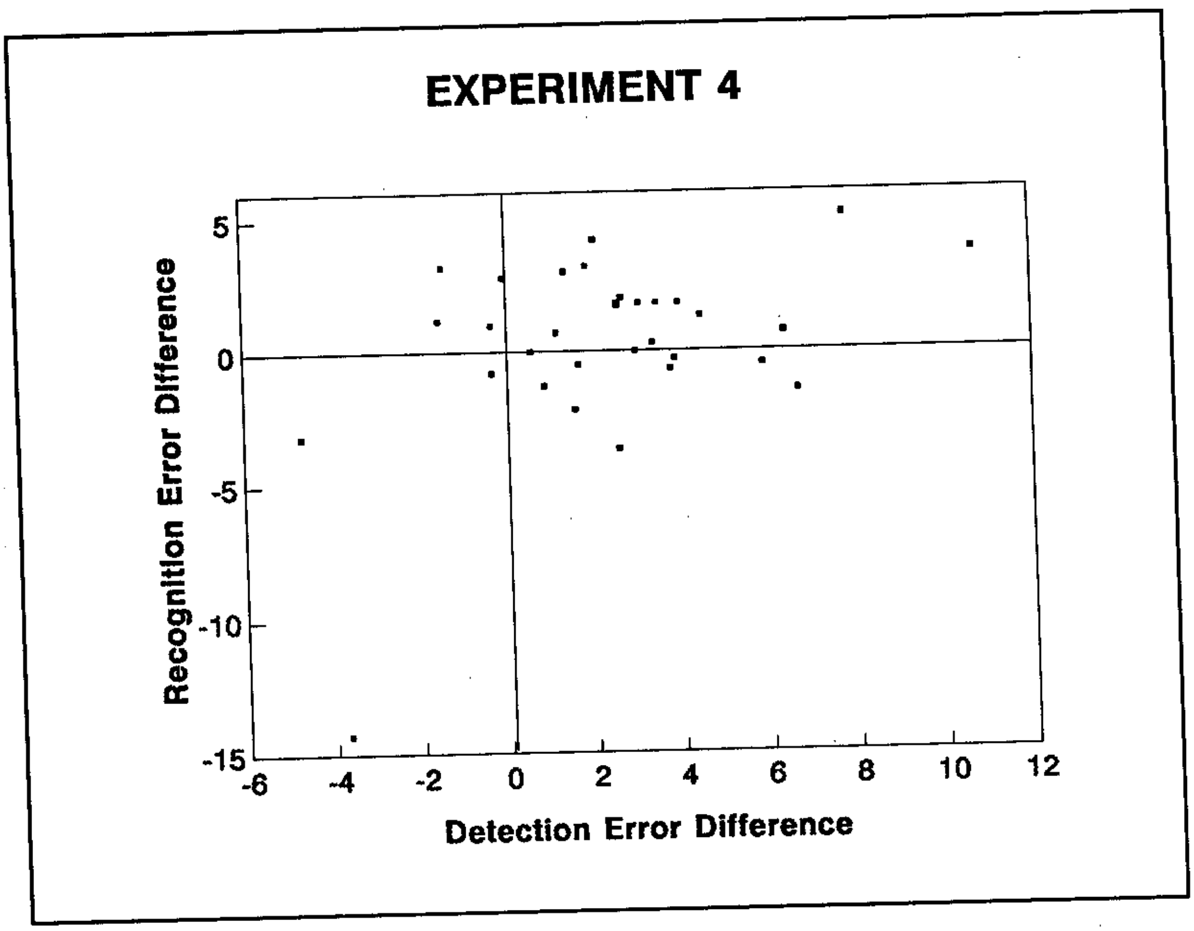

Figure 3

Results for Experiment 4 arranged as in Figure 2, but with modified axis scales.

Experiment 4

The results are graphed in Figure 3. The data confounded both predictions. Again, most Ss clustered in the ± 4 error range and there was a significant difference between high- and low- value work area detection error rates, but at a lower significance level ($p < 0.02$) than in Experiment 3. This was due to a decrease in the overall frequency of detection errors in the low-value work area ($p < 0.01$). Four Ss in experiment 4 showed a markedly high percentage of low-value detection errors, as against 9 in experiment 3. There were also no significant differences between Experiment 3 and Experiment 4 session average recognition or detection error rates, except for the shift caused by the fewer Ss having high detection error rates in the low-value work area.

The reason for the different results in Experiments 3 and 4 is more difficult to explain. Perhaps the Ss did not see much difference between a 4-point and 10-point error penalty. Even so, the reduction in the number of Ss with locking is still puzzling. One possible explanation is that the "motivation" of the Ss recruited for this test series differed from those recruited for Experiment 3. However, their overall scores and recognition error incidence were similar to the Experiment 3 data. Thus, their motivation and attention did not seem to differ from the other Ss. It may be that these results reflect slightly different scanning or working strategies adopted by the Ss in Experiments 3 and 4. For such relatively small groups, some group to group variability ought to be expected, even if each sample group's results seem normally distributed, as these were.

Since 9 Ss in Experiment 3, and 4 Ss in Experiment 4, did show an unusual tendency toward locking, there are clearly differences in scanning ability, or perhaps motivation, among our Ss. Present data do not permit an exact estimate of their prevalence. After allowances for possible population distribution variability, we estimate from our results that about 15% of the general population may exhibit this locking tendency. Thus, testing for scanning ability and any tendency towards locking could be a valuable addition to current personnel test and evaluation procedures.

There is also a general cautionary note in these results. Most studies of human scanning or monitoring performance use relatively few Ss. Indeed, many of the papers reviewed (e.g., 5, 6, 7) used fewer than 5 Ss. We do not believe that the results of such studies should have the general applicability claimed of them without extensive replication with adequate numbers of subjects.

CONCLUSIONS

About 15% of our Ss showed a marked tendency to concentrate on a display sub-area containing very high value events, while periodically ignoring events elsewhere on the display. This suggests that there may be significant individual differences in ability or strategies to effectively scan/monitor complex screen displays over long time periods. Tests for such differences could be useful for personnel selection and retention purposes in the Air Traffic Control System.

REFERENCES

1. Hopkin VD. Air Traffic Control. In: Wiener EL, Nagel DC. eds. Human factors in aviation. San Diego: Academic Press, 1988: 83-110.
2. Moray N, Rotenberg I. Fault management in process control: Eye movement and action. Ergonomics 1989; 32:1319-1342.
3. Revzin AM, Rasmussen PG. A new test of scanning and monitoring ability: Methods and initial results. Report DOT/FAA/AM-92/2. Washington, DC: Federal Aviation Administration, Office of Aviation Medicine, 1992.
4. Boles DB, Clifford JE. An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values. Behav Research Meth Instruments & Comput. 1989; 21:579-586.
5. Tulunay-Keesey U, VerHoeve JN. The role of eye movements in motion detection. Vision Res. 1987; 27:747-754.
6. Dawson, MR. Moving contexts do affect the perceived direction of apparent motion competition displays. Vision Res. 1987; 27:799-809.
7. Watamaniuk SNJ, Sekuler R, Williams DW. Direction perception in complex dynamic displays: The integration of direction information. Vision Res. 1989; 29:47-59.

☆U.S. GOVERNMENT PRINTING OFFICE: 1994 - 560-063/80123