[redacted]

FOREWORD

This report was prepared in the Human Engineering Division of the Behavioral Sciences Laboratory, 6570th Aerospace Medical Research Laboratories, Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio. The work was performed jointly under Program 665A, "Reconnaissance/Strike," and Project 7184, "Human Performance in Advanced Systems," Task 718404, "Advanced Systems Human Engineering Design Criteria." The authors are grateful for the helpful comments of Mr. Charles Bates, Jr., Chief, Performance Requirements Branch. Thanks are also due to Mr. L. L. Griffin for his administrative support and to Mr. Don F. McKechnie for his assistance in constructing the display console used. This study was begun in September and completed in December 1963.

ABSTRACT

Operator performance in detecting and identifying airfields, bridges, dams, and railroad yards from side-looking radar imagery was tested. The imagery was presented to four college students on a 14 by 14 inch screen at speeds representing Mach 0.9, 1.6, 2.3 and 3.1. An average of 65% of the targets were identified. Percentage of false targets varied significantly between subjects and ranged from 42 to 64%. A slight but consistent trend toward more detections at slower speeds was not found to be statistically significant. The number of false positive responses was not significantly affected by image speed. No decrement in performance could be found as a result of extended (3 hours maximum) trial duration.

PUBLICATION REVIEW

This technical documentary report is approved.

Walter F. Grether
WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory

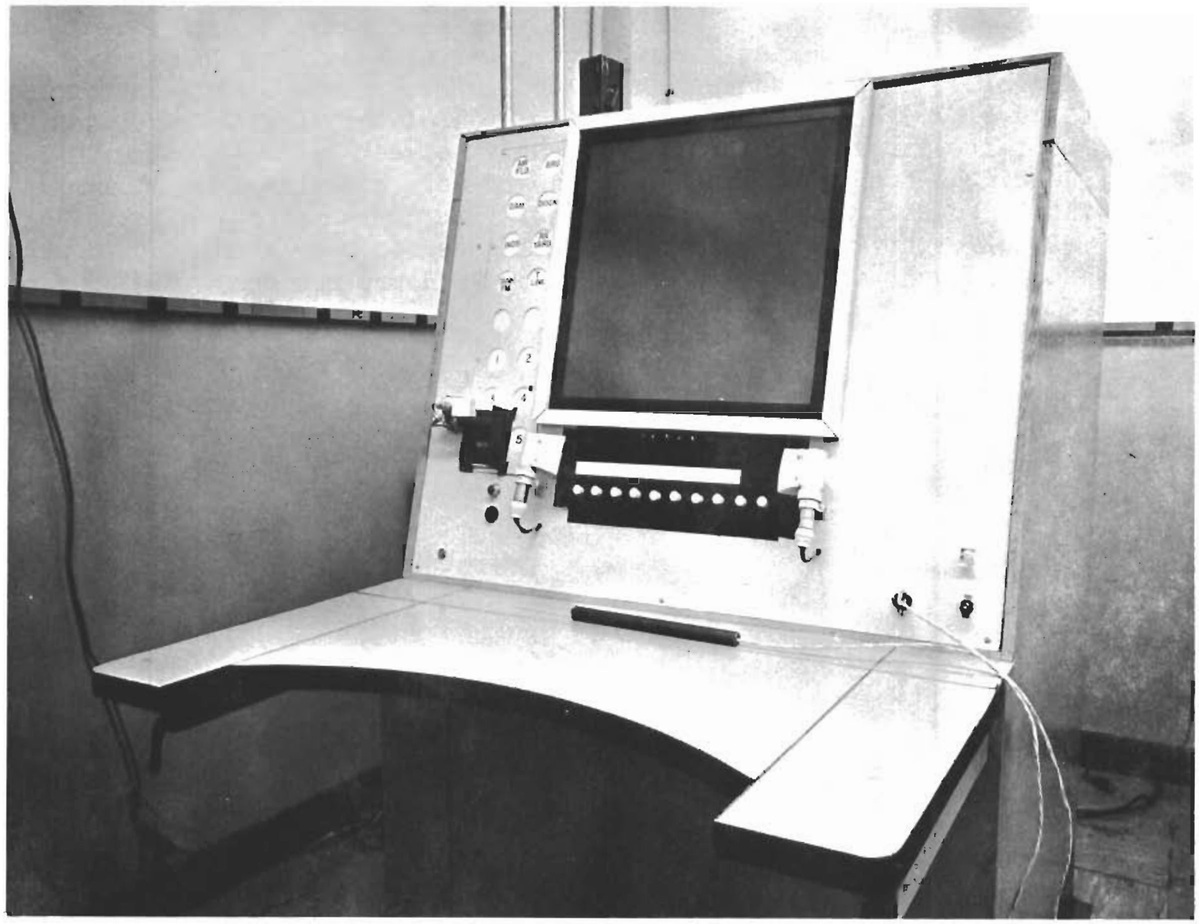

Figure 1. Display Console Used for Testing Subjects

THE EFFECT OF SIMULATED AIRCRAFT SPEED ON DETECTING AND IDENTIFYING TARGETS FROM SIDE-LOOKING-RADAR IMAGERY

H. C. Self

Fen Rhodes

I. INTRODUCTION

Airborne rapid reconnaissance and strike reconnaissance systems will require one or more human operators to locate and identify ground targets. These systems will require the operator to perform his tasks with great rapidity. Some work has been done to determine human capabilities in assessing images during extended viewing of static frames of side-looking radar imagery, but the results cannot easily be generalized to the moving imagery case.

The purpose of this investigation was to determine some of the effects of image motion upon performance in detecting and identifying different targets. Moving imagery representing continuous coverage by side-looking mapping radar of a strip of terrain 32 nautical miles wide was displayed to operators at four simulated aircraft speeds ranging from Mach 0.9 to Mach 3.1. Because unmodified radar imagery was used as stimulus material, the types and number of targets were limited to those located in a single strip of terrain coverage. Target density and relative target difficulty likewise could not be manipulated. These conditions have operated to restrict the generality of performance data obtained, and we recommend that future investigations of this type employ modified imagery in which these variables have been controlled, ie, by techniques that permit insertion and deletion of targets.

II. EQUIPMENT

Equipment for measuring subjects' performance consisted of a display console and a single-frame data camera. On the front of the console were a 14 by 14 inch rear-projection screen, a group of indicator lights, and a set of pushbuttons for activating these lights (fig. 1). The indicator lights were labeled with four target names and five confidence levels so that a subject's responses could be recorded, along with the image on the screen, by the data camera. The camera was located to the left of the subject and several feet behind him so that it had unobstructed view of the console panel.

Inside the console was a condenser-type optical projector with a film-drive mechanism that allowed precise control of the rate of image motion on the screen. The condensers and an enlarging lens stopped down to _f_/14 gave an image that was adequately bright and preserved the resolution, contrast, and dynamic range of the radar film. The magnification of 4.4 X was more than sufficient to insure that the visual acuity of subjects was not a limiting factor in perception of image detail.

III. EXPERIMENTAL PROCEDURE

The side-looking-radar film used represented two strips of terrain 32 nautical miles wide and totaling 1,870 nautical miles in length. These were displayed to subjects as one continuous strip. The exact terrain coverage, divided into 21 equal sections, is shown on the map in figure 2. The imagery was projected on the display at a scale of 1:164,000 and was of good resolution in terms of standards currently applied to side-looking radar. Typical brightness ratios were 18:1 for land to water and 4.2:1 for the center of a city to the surrounding countryside.

There were 130 targets (ie, airfields, bridges, dams, and railroad yards) that could be discerned in the imagery. Targets that were not directly visible in the imagery and could be presumed present only from background cues were not counted. Relative frequency of occurrence of different targets at each terrain section along with the flight path is given in table 1.

The average size in millimeters for each type of target and the size range appear in table 2. Maximum viewing distance was 41 cm (16 in.), but subjects were allowed to move in as close as about 15 cm (6 in.) whenever they felt the need. In the absence of more precise information, the smaller distance would have to be assumed as appropriate in every case for purposes of determining the visual angles subtended by targets.

Subjects were four male college students with normal visual acuity as indicated by test charts. Training began with a review of aerial photographs to familiarize subjects with the appearance of targets as seen from the air. They then studied a brief technical introduction to radar and side-looking radar, after which they examined selected examples of side-looking radar imagery representing different types of targets. This introduction was followed by approximately 10 hours of finding targets on enlarged segments of side-looking radar film. Subjects evaluated their performance with accompanying transparent overlays containing target information. During this period they also studied material describing pertinent cues of the various targets and associated terrain features. At the end of training each subject had at least 4 hours of practice locating and identifying targets in moving imagery displayed on the test console.

A Latin-square arrangement of subjects, trials, and display speeds was used in taking performance measures, as shown in table 3. Selection of the square and randomization of rows and columns were in accordance with the procedure described by Edwards (ref. 1). The stimulus strip of radar film, covering terrain not seen during training, was presented at all speeds to each of the four subjects. These speeds, expressed both as image motion rates and as simulated aircraft speeds, are shown together with trial durations in table 4. Subjects were given no more than one trial per day.

At the start of each trial the subject was told approximately how long it would last. No attempt was made to conceal from subjects the fact that the same strip of film was used for all trials. The subject's task was to locate and identify airfields, bridges, dams, and railroad yards as soon as possible after they moved into view on the screen. He did not know, however, the total number of targets present.

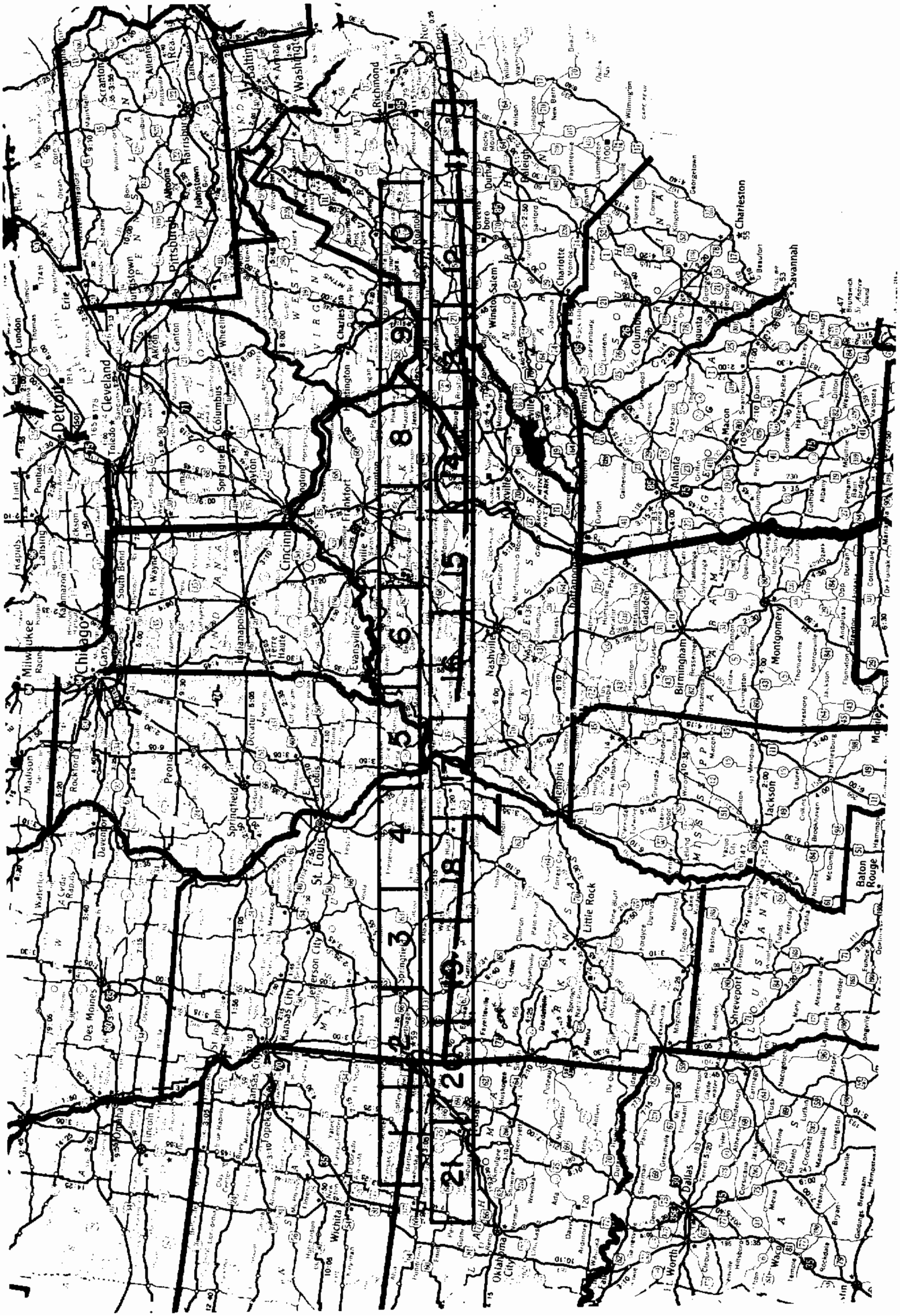

Figure 2. Map Showing Terrain Covered by Radar Imagery

TABLE 1

DISTRIBUTION OF TARGETS BY TERRAIN SECTION

Terrain Section	Number of Targets			
	Airfields	Bridges	Dams	Railroad Yards
1	3			1
2	2	3		2
3	1	1	1	
4				
5	1	8	1	
6		6	3	
7		2		1
8				
9		4		
10	2	5	3	1
11	2	2	3	
12		9	1	
13				
14		3		1
15	1	6	1	
16		3	1	
17		6		
18			1	1
19		10		
20		10	1	
21		16	1	
TOTAL	12	94	17	7

TABLE 2

MEAN SIZE AND RANGE FOR EACH TYPE OF TARGET

	Airfields	Bridges	Dams	Railroad Yards
Mean Size (mm)	10.1	3.7	4.7	10.9
Smallest (mm)	7.2	1.1	1.4	5.4
Largest (mm)	14.7	12.9	9.7	16.9

TABLE 3

LATIN SQUARE USED FOR PRESENTATION
OF DISPLAY SPEEDS

Subject	Trial			
	1	2	3	4
1	1.6†	2.3	3.1	0.9
2	0.9	3.1	1.6	2.3
3	2.3	1.6	0.9	3.1
4	3.1	0.9	2.3	1.6

† Speed in simulated Mach.

TABLE 4

IMAGE MOTION RATES, SIMULATED SPEEDS, AND TRIAL DURATIONS FOR EXPERIMENTAL CONDITIONS

Image Motion Rate		Simulated Aircraft Speed			Trial Duration (Min.)
Cm/Min	In/Min	Km/Hr	Knots	Mach †	
11.2	4.4	1100	593	0.9	180
20.0	7.9	1980	1070	1.6	100
29.0	11.4	2850	1540	2.3	69
37.8	14.9	3740	2020	3.1	53

† Computed using Mach 1.0 = 660 knots at sea level.

When the subject found what he believed to be a target, he pointed to it with an illuminated stylus and indicated both his judgment as to the target name and his level of confidence (expressed on a five-point scale), using the pushbuttons beneath the screen. The event was automatically recorded on 35-mm film by the data camera.

The following performance measures were obtained:

1. Number of targets detected.

A detection was said to occur when the subject pointed with the stylus to one of the 130 targets. Of the targets detected only 12% were not correctly identified and most (97%) of these errors represented confusion between bridges and dams. In view of this low rate of identification errors, "detections" and "detections correctly identified" were not considered different enough to warrant separate treatment in the analysis.

2. Number of false positives and percentage of false positives, based on total responses.

A false-positive response occurred when the subject reported a target where there was none present.

3. Distance across screen that targets had traveled before being detected.

This was measured using an 11-category, equal-interval scale that covered the 14-inch screen.

4. Subjective confidence level associated with each detection, ie, how sure the subject was his response was correct.

There were five levels, number one representing high confidence and five low confidence.

Of the total number of responses for all 16 trials 118 (3.8%) could not be scored because of difficulties with the data camera. These unscorable responses were divided proportionally between detections and false positives separately for each trial in order to obtain accurate estimates of detection frequency under each condition.

IV. RESULTS

The percentage, averaged over all subjects, of targets detected at each simulated aircraft speed is shown graphically in figure 3. The percentage of targets detected by each subject individually is shown in figure 4. The overall average of targets detected was 65%.

Detections were divided into two groups for separate analyses: (1) bridges plus dams and (2) airfields plus railroad yards. The small number of targets in each category precluded further subdivision. The logic for this particular grouping was that both bridges and dams had relatively more distinctive contextual cues, such as streams and lakes, than either airfields or railroad yards. It was assumed that the latter targets would, therefore, be more sensitive to the effects of image speed. This hypothesis of differential speed sensitivity was not, however, borne out by the results. The percentage of bridges and dams detected and the percentage of airfields and railroad yards detected at each speed are also shown in figure 3. The difference between the average percentages (69% and 44%) was significant at the .05 level.

Analyses of variance for number of targets detected at the different speeds by the four subjects are shown in tables 5, 6, and 7. The $\underline{F}$ ratios for image speed were not significant.

The percentages of false positive responses were also analyzed (table 8). These are plotted in figure 5 for each subject at the four speeds used. The only component of variance found to be significant was that for subjects. (In this experimental design subject variance is completely confounded with variance due to treatment orders. However, order of presenting image speeds was judged to have little influence upon a subject's overall performance, and the "subject-order" effect is therefore referred to simply as the "subject" effect.) The proportion of total responses representing false positives was not significantly related to image speed.

Table 9 compares the average number of false positive responses for each type of target with the number of targets present and the number of detections. Chi-square

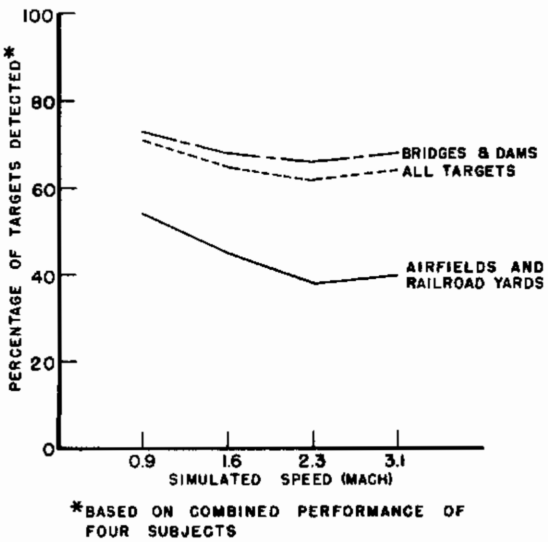

Figure 3. Percentage of Targets Detected at Four Simulated Aircraft Speeds

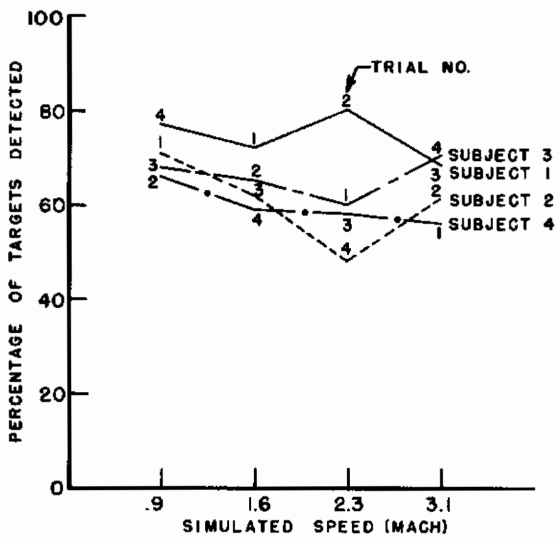

Figure 4. Percentage of all Targets Detected by Each Subject

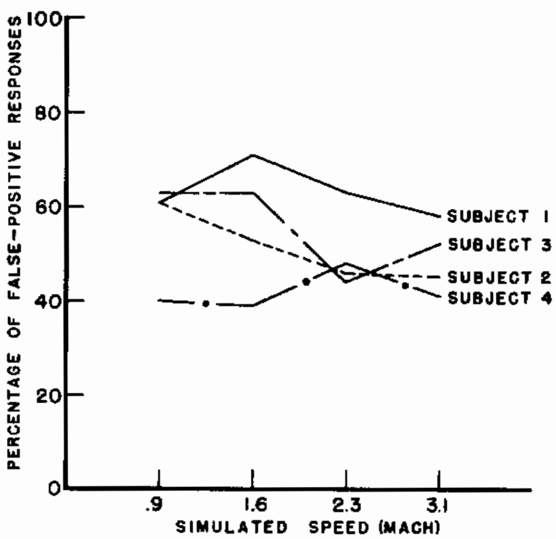

Figure 5. Percentage of False-Positive Responses for Four Subjects

TABLE 5

ANALYSIS OF VARIANCE FOR NUMBER OF DETECTIONS
(ALL TARGETS)

Source of Variation	Sum of Squares	df	Mean Square	F
Image speeds	300.69	3	100.23	1.57
Subjects	896.19	3	298.73	4.69
Trials	80.69	3	26.90	.42
Error	381.87	6	63.64	
Total	1659.44	15		

TABLE 6

ANALYSIS OF VARIANCE FOR NUMBER OF DETECTIONS
(BRIDGES AND DAMS)

Source of Variation	Sum of Squares	df	Mean Square	F
Image speeds	166.69	3	55.56	1.10
Subjects	877.68	3	292.56	5.81*
Trial	100.69	3	33.56	.67
Error	302.38	6	50.40	
Total	1447.44	15		

*Significant at .05 level.

TABLE 7

ANALYSIS OF VARIANCE FOR NUMBER OF DETECTIONS
(AIRFIELDS AND RAILROAD YARDS)

Source of Variation	Sum of Squares	*df*	Mean Square	*F*
Image speeds	22.25	3	7.42	1.35
Subjects	34.25	3	11.42	2.08
Trials	6.25	3	2.08	.38
Error	33.00	6	5.50	
Total	95.75	15		

TABLE 8

ANALYSIS OF VARIANCE FOR PERCENTAGE OF
FALSE POSITIVE RESPONSES†

Source of Variation	Sum of Squares	*df*	Mean Square	*F*
Image speeds	63.23	3	21.08	1.27
Subjects	319.00	3	106.33	6.39*
Trials	27.36	3	9.12	.55
Error	99.81	6	16.64	
Total	509.40	15		

†Arcsin transformation applied to original data.
*Significant at .05 level.

TABLE 9

NUMBER OF TARGETS, DETECTIONS, AND FALSE POSITIVES BY TARGET TYPE

	Airfields	Bridges	Dams	Railroad Yards
Targets present	12	94	17	7
Detections †	6.0	59.8	16.5	2.3
False positives †	10.5	64.8	18.3	12.5

† Frequencies are averages based on performance of four subjects

goodness-of-fit tests were performed to determine whether or not the observed distribution of false-positive frequencies could be said to match theoretical frequencies derived from (1) the distribution of targets in the imagery, (2) the distribution of detections for the four target types, and (3) an equal-probability distribution assuming equal false-positive frequencies in all target categories. In every case the observed distribution of false positives was found to be significantly different ($p < .01$) from the computed theoretical one. A goodness-of-fit test was also performed comparing the distribution of detections with that of targets with a finding of no significant difference ($p > .20$) between the two. A chi-square test for independence between type of response (detection or false positive) and target type resulted in acceptance ($p > .10$) of the hypothesis of no relationship between the two variables.

Analyses of variance for the average distance targets had traveled across the screen at the time of detection are reproduced in tables 10, 11, and 12. These results are plotted in figure 6. Both the variance due to image speeds and that attributable to differences among subjects were significant for bridges plus dams and for all targets combined. Neither of these effects was significant for airfields and railroad yards, perhaps because of the relatively small number of detections for these targets.

Figures 7, 8, 9, and 10 show for each simulated speed the number of detections versus the number of false positive responses plotted as a function of subjects' expressed confidence level. At every speed there was found to be a significant relationship ($p < .01$) between type of response and expressed confidence level as tested by chi square.

To find out whether or not there was any decrement to performance associated with extended trial duration, subjects' responses near the beginning of the 180-minute trial (the slowest one) were compared to those near the end. The number of

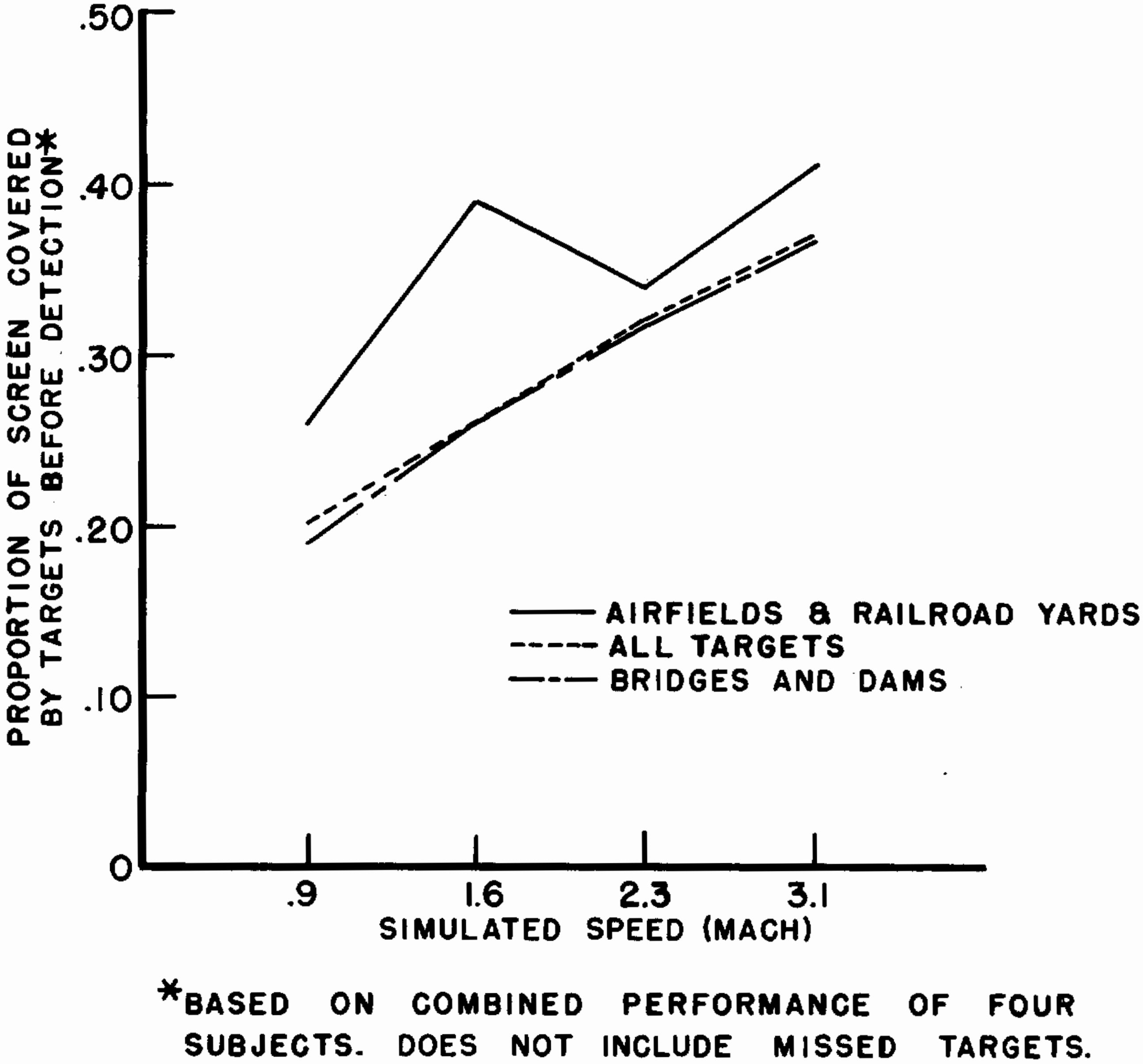

Figure 6. Average Distance Traveled by Targets Across Screen Before Detection

TABLE 10

ANALYSIS OF VARIANCE FOR DISTANCE TRAVELED BY TARGETS ACROSS SCREEN BEFORE DETECTION (ALL TARGETS)

Source of Variation	Sum of Squares	df	Mean Square	F
Image speeds	8.21	3	2.74	10.14**
Subjects	4.45	3	1.48	5.49*
Trials	2.02	3	.67	2.49
Error	1.62	6	.27	
Total	16.30	15		

*Significant at .05 level.
**Significant at .01 level.

TABLE 11

ANALYSIS OF VARIANCE FOR DISTANCE TRAVELED BY TARGETS ACROSS SCREEN BEFORE DETECTION (BRIDGES AND DAMS)

Source of Variation	Sum of Squares	df	Mean Square	F
Image speeds	8.48	3	2.83	20.44**
Subjects	4.94	3	1.65	11.91**
Trials	1.13	3	.38	2.72
Error	.83	6	.14	
Total	15.38	15		

**Significant at .01 level.

TABLE 12

ANALYSIS OF VARIANCE FOR DISTANCE TRAVELED BY TARGETS ACROSS SCREEN BEFORE DETECTION (AIRFIELDS AND RAILROAD YARDS)

Source of Variation	Sum of Squares	*df*	Mean Square	*F*
Image speeds	6.70	3	2.23	1.57
Subjects	7.96	3	2.65	1.86
Trials	2.03	3	.68	.47
Error	8.56	6	1.43	
Total	25.25	15		

detections that occurred during a 45-minute period beginning 9 minutes after the start of the trial and another period of equal duration terminating 9 minutes before the end of the trial were used as a basis for this comparison. There were 29 targets in the first segment of film and 34 in the second. The distributions of targets by type were roughly equivalent, with bridges predominating in both groups.

The average percentage of targets detected in the 45-minute segment at the beginning of the trial was 55%; the percentage of targets detected at the end was 86%. This difference was significant beyond the .01 level, supporting a conclusion of no fatigue effect for the three-hour viewing period. However, a confounding factor was the finding that targets at the end of the run were easier than those at the beginning. This was tested by comparing for the beginning and end segments the number of targets detected by none, one, two, three, and four subjects. The result showed a significant difference ($p < .01$) in difficulty distributions ($\chi^2 = 14.89$, $df = 4$). Any conclusion regarding the effect of extended viewing based on the present results must therefore be regarded as highly tentative.

V. DISCUSSION AND CONCLUSIONS

On the plot of detections vs speed there was a slight but consistent decrease in detections with increase in image speed. However, analysis of variance demonstrated that there was no statistically significant speed effect on either detections or false positives. The result of the statistical analysis for detections makes good sense when certain facts are considered. First, the screen size and the 1:164,000 scale at which terrain was displayed permitted a target to remain on the screen for nearly one minute, even at the fastest display speed. This fact and the finding that targets traveled on the average less than halfway down the screen before being detected support a conclusion that more than enough search time was available to subjects under all speed conditions.

Second, the 65% average of targets detected indicated that the level of target difficulty was in the easy to moderately hard range. Although the stimulus material represented real targets on real terrain, no systematic control was exercised over the range or representativeness of target difficulties. It is uncertain whether or not the addition of harder-to-locate targets would change the present results with respect to display speed, ie, whether speed would become an influential determiner of performance for more difficult targets, but it could reasonably be expected that such would be the case.

The significant difference ($\underline{z}$ = 2.08, $\underline{p} < .05$) between the proportion of bridges and dams detected versus that for airfields and railroad yards, which is illustrated in figure 2, would seem to indicate a difference in target difficulty, with the bridge-dam category being the easier. However, this conclusion is not borne out by the results of the chi-square test in which detections were considered by individual target type and compared with theoretical frequencies derived under the assumption that equal proportions of the airfields, bridges, dams, and railroad yards were detected by subjects. The obtained value for chi square indicated a nonsignificant difference ($\underline{p} > .20$) between the observed and theoretical detection frequencies for the four types of targets. The authors' tendency is to discount the barely significant difference obtained when targets were combined into airfield-railroad yard and bridge-dam categories as an artifact of the grouping procedure itself and to accept the chi-square results implying no differences in target difficulty.

The finding that screen distance traveled by targets before detection increased as a function of display speed indicated that response time or search time did not keep pace with the increases in image speed, although subjects' response times did decrease appreciably as speed increased. It would appear from figure 6 that a considerable increase in image speed beyond the maximum investigated could be tolerated without seriously affecting measures of performance other than response time. It would likewise be possible to narrow the screen if faster speeds or larger image scales were not required.

The significant variability in the percentage of false-positive responses with respect to subjects, ranging from 42 to 64% for different individuals, was attributed to failure to establish a common standard of judgment for reporting the presence of targets. The high false-positive rates for all subjects, coupled with the lack of significant variation from subject to subject in number of targets detected, pointed to a conclusion that every subject's criterion or cutoff point, although different from that of other subjects, was sufficiently low to include most of the real targets plus a large number of false positives or false alarms.

The situation is analogous to that described by Tanner and Swets (ref. 2) in their theory of visual detection, where the observer must distinguish between signals arising from a distribution of noise alone and those arising from a distribution of signal plus noise. Ease of discrimination of true signals (targets) from noise (nontargets) depends upon the amount of overlap between the two distributions, determined by the variances of the distributions and the distance between their means (ie, how much the signal overrides the noise or how different, on the average, targets are from nontargets). The criterion established by each observer represents an individual, subjective standard or cutoff point above which a stimulus will be accepted as a true signal and below which it will be rejected as pure noise.

The relationship found to exist between type of response (detection versus false positive) and subjects' expressed confidence that the response was correct (see fig. 7, 8, 9, and 10) underlines the importance of the subjective criterion as a determiner of overall performance quality. Moving a subject's signal acceptance criterion up from confidence level five to level three, for example, would improve his total performance by greatly reducing the frequency of false positive responses with little sacrifice in the number of detections.

It should be possible to establish a more favorable balance between number of false positives versus actual target detections than was observed in the present study, at the same time keeping the proportion of targets missed to an acceptable level. Additional experimentation would be required to establish the extent to which this could be accomplished, and would involve the manipulation of observers' cutoff points by pretrial training and differential reward.

VI. SUMMARY

This study was conducted to examine the performance of human operators in detecting and identifying targets contained in a strip of side-looking radar imagery displayed at four simulated aircraft speeds ranging from Mach 0.9 to Mach 3.1. Subjects were four male college students who had been trained to find targets from side-looking radar coverage of terrain. Imagery was rear-projected onto a 14 by 14 inch translucent display screen at a scale of 1:164,000. The direction of image motion was vertical from bottom to top of the screen. A 4 by 4 Latin square of subjects by trials by speeds was used in the experiment. Trial duration varied from 53 minutes at the fastest image speed (37.8 cm or 14.9 in. per min.) to 3 hours at the slowest (11.2 cm or 4.4 in. per min.). Subjects were required to point out and identify airfields, bridges, dams, and railroad yards as quickly as possible as they moved into view on the screen. The average percentage of all targets detected was 65%. Analysis of results indicated that the number of targets detected did not vary significantly across the range of speeds tested. Furthermore the percentage of false-positive responses did not vary as a function of image speed. Percentage of false positives varied significantly among subjects, however, and ranged from 42 to 64%. Also, the relative number of false-positive responses compared with detections increased significantly at all speeds as subjects' expressed level of confidence decreased. The distance targets traveled across the screen before being detected was found to be significantly related to speed of image motion. At the slowest speed, targets that were detected traveled an average of 20% of the screen width, while at the fastest speed this distance increased to 37% of total screen width. Performance of subjects on the longest trial was analyzed to ascertain whether there was any decrement associated with

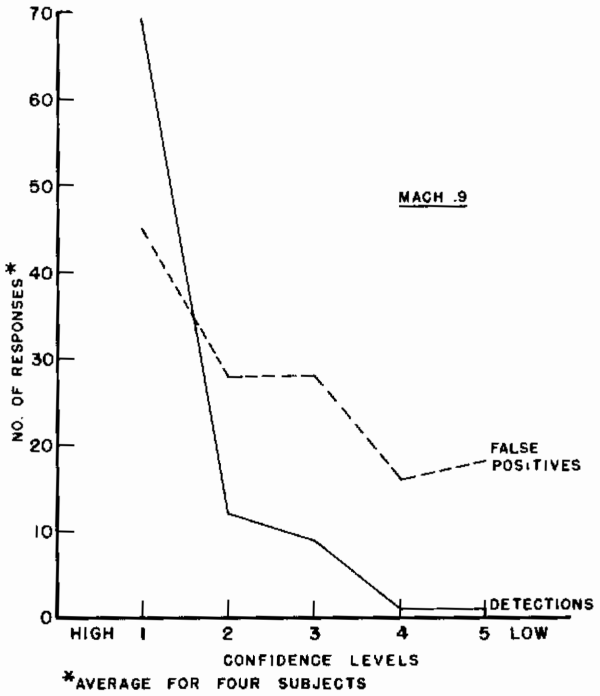

Figure 7. Detections vs. False Positives as a Function of Confidence Level at Simulated Mach 0.9

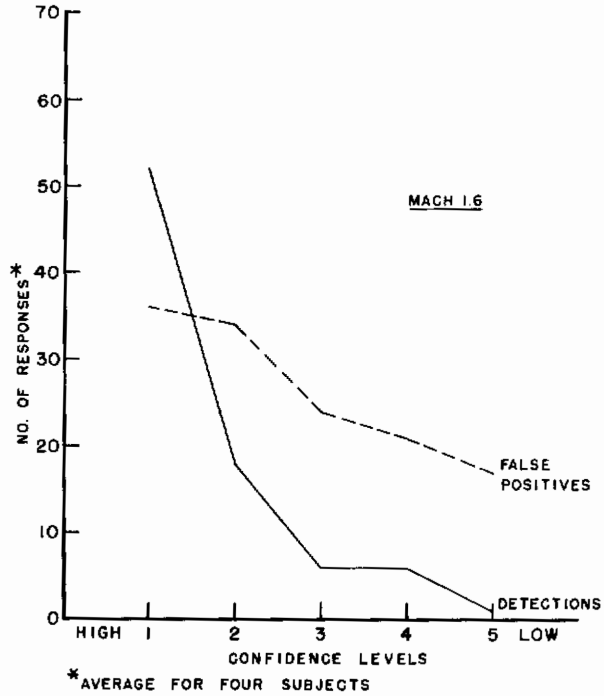

Figure 8. Detections vs. False Positives as a Function of Confidence Level at Simulated Mach 1.6

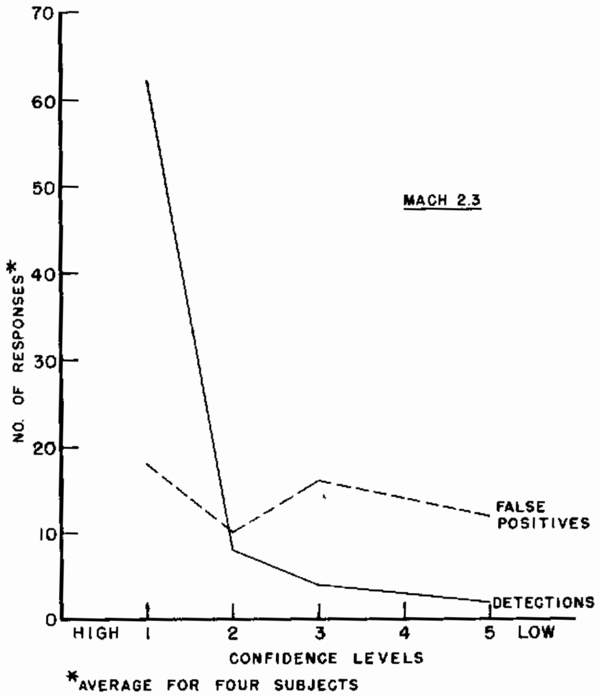

Figure 9. Detections vs. False Positives as a Function of Confidence Level at Simulated Mach 2.3

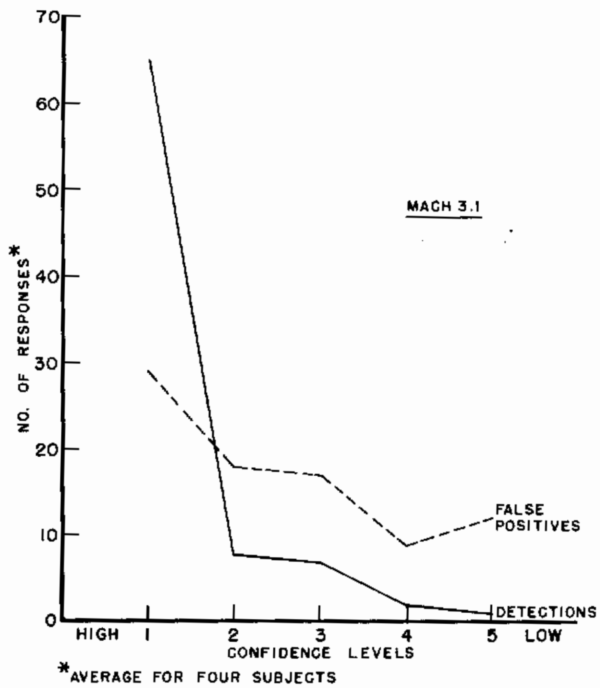

Figure 10. Detection vs. False Positives as a Function of Confidence Level at Simulated Mach 3.1

extended viewing time. No decrement was found (in fact the percentage of targets detected was actually higher near the end of the trial), but the results were considered inconclusive because the targets at the end were determined to be less difficult than those at the beginning.

REFERENCES

1. Edwards, A. L. Experimental design in psychological research. New York: Holt, 1960.

2. Tanner, W. P., Jr., and Swets, J. A. A decision-making theory of visual detection. Psychol. Rev., 61, 401-409, 1954.