Technical Research Report 1129

IMAGE INTERPRETATION TASK--STATUS REPORT,

30 JUNE 1962

U. S. ARMY
PERSONNEL
RESEARCH
OFFICE

US ARMY PERSONNEL RESEARCH OFFICE

An activity of the Chief, Research and Development

Charles S. Gersoni
Colonel, GS
Commanding

J. E. Uhlaner
Director, Research Laboratories

Hubert E. Brogden
Chief Scientist

Initial distribution of this report has been made by the US Army Personnel Research Office. Additional copies may be obtained as follows. Department of Defense agencies and their contractors should address requests for additional copies to: Commander, Armed Services Technical Information Agency, ATTN: TIPDR, Arlington Hall Station, Arlington 12, Virginia. Other agencies and persons may obtain information on cost of microfilm or photostatic copies from: Office of Technical Services, Department of Commerce, Washington 25, D. C. ATTN: Technical Reports Section. Copies may also be obtained on loan from local depository libraries. A list of these libraries may be obtained from: Documents Expediting Project, Library of Congress, Washington 25, D. C. The mailing address of USAPRO is: US Army Personnel Research Office, Washington 25, D. C.

AD Div 23/1, 28/4

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by Robert Sadacca, Harold Martinek, and Alfred I. Schwartz. October 1962. Rept on Image Interpretation Task. 33 p. incl. tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129) (DA Project 0J95-60-001) Unclassified Report

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by Robert Sadacca, Harold Martinek, and Alfred I. Schwartz. October 1962. Rept on Image Interpretation Task. 33 p. incl. tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129) (DA Project 0J95-60-001) Unclassified Report

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by Robert Sadacca, Harold Martinek, and Alfred I. Schwartz. October 1962. Rept on Image Interpretation Task. 33 p. incl. tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129) (DA Project 0J95-60-001) Unclassified Report

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by Robert Sadacca, Harold Martinek, and Alfred I. Schwartz. October 1962. Rept on Image Interpretation Task. 33 p. incl. tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129) (DA Project 0J95-60-001) Unclassified Report

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

USAPRO Technical Research Report 1129

IMAGE INTERPRETATION TASK, ~~STATUS REPORT,~~
~~30 JUNE 1962~~

Robert Sadacca, Harold Martinek,
and Alfred I. Schwartz

~~Submitted by~~ Joseph Zeidner
Chief, Support Systems Research Laboratory

Approved by

J. E. Uhlaner
Director, Research Laboratories

Hubert E. Brogden
Chief Scientist

Charles S. Gersoni
Colonel, GS
Commanding

Army Project Number
OJ95-60-001

Image Interpretation

October 1962

USAPRO Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form. Status reports in the blue cover depart from the usual format in that they do not necessarily contain separate management and technical supplements.

ACKNOWLEDGMENTS

Appreciation is expressed to personnel of the U. S. Army Intelligence Material Developments Agency, the U. S. Army Intelligence Combat Developments Agency, the U. S. Army Photo Interpretation Center, the U. S. Army Intelligence School, and to the staff of the U. S. Army Intelligence Center for assistance in the formulation of problems and the development of test materials. Of special assistance in the conduct of this research program have been Lt Col Eugene H. Manley, Lt Col Henry F. Hauser, Lt Col Matthew J. Toia, Lt Col Eugene E. Fels, and Lt Col Ralph M. Luman, in the providing of valuable advice and guidance.

Appreciation is also expressed to the Image Interpretation Section of the 525th Military Intelligence Group for assistance in image selection, historical research, and development of performance measures, and to the Office of the Chief of Military History, the World War II Records Division of the National Archives, and the USAF Aeronautical Chart and Information Center for special assistance and use of their facilities by the research personnel.

Appreciation is specifically expressed to Captain John E. Ranes, USAPRO, for his invaluable technical assistance and efforts as the principal liaison officer between the research staff and the sponsoring and monitoring agencies of the intelligence community and of the Department of Defense.

BRIEF

IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962

Requirement:

The Office of the Assistant Chief of Staff for Intelligence, DA, the US Army Materiel Command, and the U. S. Army Combat Developments Command have a requirement for research into the human factors underlying imagery interpretation. The primary objective is to maximize the information extracted from imagery in terms of accuracy, completeness, and speed.

Procedure:

Based on an extensive survey of the Army's image interpretation facilities and procedures, and of completed and ongoing research, a research program was formulated encompassing the following subtasks:

a. Identification of basic factors in image interpretation

b. Development of selection techniques for image interpreter personnel

c. Utilization measures under conditions of emergency demands

d. Effective group approaches to accomplish major image interpreter missions

Performance measures reproducing operational image interpretation situations and yielding indexes of accuracy, completeness, and speed of information extraction were developed for use in experimentation. In each subtask, exploratory studies have delineated specific problems on information extraction and indicated directions of experimentation likely to be profitable.

Utilization of Findings:

The following initial results have implications for improvement of image interpreter performance:

1. Definite improvement over individual performance, in terms of accuracy and in number of correct identifications, was achieved when two or three interpreters worked together as a team.

2. Information from other intelligence sources concerning the presence of objects in an area to be searched can act to suggest to image interpreters the presence of objects not actually present in the imagery. Where accurate reporting is essential, an interpreter should probably work without benefit of information available from other sources, at least for an initial period.

3. Time does not necessarily work to advantage in image interpretation. Later interpretations in a given viewing period tend to contain a higher proportion of erroneous information than interpretations made earlier.

4. Interpreters tend to express greater confidence in identifications later proved correct than in those which prove wrong. If systematic advantage can be taken of expressed confidence, the amount of erroneous information may be substantially reduced.

5. Preliminary experimentation indicates that interpreter performance changes depending upon whether accuracy or completeness is emphasized in instructions given interpreters. Providing interpreters with more precise knowledge of the intelligence requirement in terms of accuracy vs completeness could result in a more appropriate output.

6. Interpreters who checked their own reports against established ground truth made a reduced number of wrong identifications in searching similar photos on a later assignment. Knowledge of how well they have done may be effective in reducing error, especially if interpreters can see exactly what kinds of error they are making. Also, allowing interpreters to take a second look at imagery after a time lapse may increase accuracy of interpretation.

7. Interpreters receiving the same training show wide individual differences in performance. These differences can be predicted by aptitude measures such as tests of spatial visualization and mathematical reasoning. An effective battery for improving the selection of interpreter trainees is being recommended.

IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962

CONTENTS

TABLES

FIGURES

IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962

BACKGROUND AND BASIC OBJECTIVES

Survey of Human Factor Requirements in Image Interpretation

In response to a research requirement of the Department of the Army initiated by the Assistant Chief of Staff for Intelligence, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by members of the Support Systems Research Laboratory, U. S. Army Personnel Research Office (APRO). The objective of the survey was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long range needs. Over 20 military and civilian organizations were contacted by the research personnel of the laboratory.

A statement of the military requirements in image interpretation and the formulation of the APRO research program in image interpretation were direct outgrowths of this survey. The military requirement is the communication to action officers of timely, relevant, complete, and accurate intelligence information that can be derived from aerial imagery. Until recently, the Army's image interpreter was confronted only with the problem of interpreting relatively large-scale, conventional black-and-white photographs. Today he is also required to interpret small-scale and degraded photographs, as well as radar, infra-red, and TV imagery. All the military services are currently increasing both their capacity to obtain imagery and their capacity to process imagery at a rapid rate. This increased capacity and variety of image sources have inevitably increased the quality range of images. But good or poor, the imagery obtained through these media is ultimately placed before the image interpreter who is asked to extract information useful for important military decisions.

The Research Program Formulated

Existing empirical knowledge about the basic psychological factors operating when image interpreters interpret reconnaissance imagery is severely limited. Previous research has tended to concentrate on improving the nature of the material presented to the image interpreter; that is, on improving the quality of obtained imagery. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources in order to cope with the ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The research program of the Image Interpretation Task was divided into four[1] areas or subtasks:

---

[1] A fifth subtask area, Identification of Basic Factors in Real Time Interpretation, is being investigated by the IMAGE SYSTEMS Task within the Support Systems Research Laboratory.

Identification of Basic Factors in Image Interpretation
Development of Selection Techniques for Image Interpreter Personnel
Utilization Measures Under Conditions of Emergency Demands
Effective Group Approaches to Accomplish Major Image Interpreter Missions

Details of this program were described in a report, Survey of Psychological Factors in Image Interpretation (Zeidner and Birnbaum, 1959), which was submitted for critical evaluation to a number of interested military intelligence and surveillance agencies. On the basis of the comments received (Mendelson, 1959), priorities were established within the Task and experiments were initiated in each subtask area.

The present report briefly describes the research objectives and summarizes the experiments conducted to date within each subtask area. Where a previous publication describes in greater detail results of a particular experiment, the publication is referenced and only the general conclusions and recommendations growing out of the experimental work are summarized.

Performance Measures Developed for Experimental Use

In order to conduct the many studies needed to answer the formulated research questions, a large proportion of the Task effort has been devoted to the development and tryout of a series of performance measures. These performance measures have been used to gauge the effect of the various experimental conditions on the accuracy, completeness and speed of image interpretation. The measures reproduce, as far as possible, typical real-life situations which confront the image interpreter. Most of the imagery was obtained on actual tactical and strategic air reconnaissance missions during wartime and shows enemy-held positions. Imagery collected during maneuvers was used when sufficiently realistic. The objects the image interpreter is directed to look for reflect the intelligence requirements of the military situation at the time of the air reconnaissance mission. As in an operational situation, the interpreter is provided with maps of the relevant areas, sortie plot overlays, and general and specific information about the location and disposition of friendly and enemy forces. Also as in the operational situation, the image interpreter is allowed to examine the imagery presented to him in any way he desires and is allowed to use any instruments in his kit.

In determining the content of the photographs--the ground truth--different approaches were used for tactical and for strategic performance measures. For the tactical measures, historical records of World War II and the Korean War were studied and related to the photography. In the case of imagery collected on maneuvers, reports, sketches, and photos showing ground truth were employed. Using the knowledge obtained from these records, several experienced image interpreters of the 525th Military Intelligence Group at Fort Meade, Maryland, interpreted the photographs. Their interpretations were checked by experienced interpreters assigned to the Army Personnel Research Office. A third check was made by instructors at the U. S. Army Intelligence School. The few objects on which the three groups could not agree are scored as neither right nor wrong. In a few instances no background information was available and ground truth was determined solely by the consensus of experienced photo interpreters.

For the strategic measures, ground truth was determined by experienced interpreters assigned to APRO, using maps, the bombing encyclopedia, the National Intelligence Surveys, and other sources. Identifications were checked by instructors of the Intelligence School. Any object or installation actually existing but not clearly recognizable by the experts is scored as neither right nor wrong. Additional performance measures are being developed as needed, using the procedures described.

Experimental Subjects

All the experimental results described in the present report are based on performance measure scores made by officers and enlisted men in the image interpretation course at the Intelligence School, Fort Holabird, Maryland. Through coordination with USCONARC, ACSI, and the Intelligence Center and School, a continuing arragement was made to conduct experiments with student interpreters during a three-day period immediately prior to their completion of the interpretation course. Thus, interpreters in the samples had had no on-the-job experience. From the performance of the graduating interpreters used in these experiments, it was not possible to determine directly the contribution that image interpreters make to intelligence estimates. However, by comparing techniques and procedures under controlled experimental conditions, useful indications are provided as to how the operational effectiveness of image interpreters may be improved. More recently, experienced interpreters from operational units have served as experimental subjects. Data from these subjects have not yet been analyzed.

IDENTIFICATION OF BASIC FACTORS IN IMAGE INTERPRETATION

The ultimate objective of this subtask is to improve the image interpreter's productivity through the optimal use of his talents, skills, and abilities. Research has tended to deal with three main aspects of image interpretation: imagery input, interpreter operations and techniques, and intelligence output. In the input stage, concern has been with the stimulus dimensions of the imagery presented to the interpreter. In the area of interpreter operations and techniques, effort is directed toward the man himself, and how he goes about his job of interpreting. Interest is in the instructions and background information he receives, the search and viewing techniques he employs, and the references, equipment, and other aids he uses to accomplish his job. In the output stage, concern is with the performance fractions themselves, that is, with the accuracy, timeliness, and completeness of the interpreter's report. The functional interrelationships of criterion fractions have been explored in such studies as that relating output rate and accuracy over time.

Stereoscopic Viewing. In this study (Zeidner, Sadacca, and Schwartz, 1961), stereo photos were provided to one of two matched groups and non-stereo photos to the other. The task of the interpreters was to free search the imagery and locate and identify critical enemy targets. Analysis of the data revealed no significant or practical differences in the performance of the two groups in terms of the accuracy of the information extracted or the confidence the interpreters had in their identifications. The number of objects reported during initial viewing periods, however, tended to be higher under non-stereo conditions, perhaps because interpreters did not spend as much time aligning the photos. Findings suggest that stereo viewing is not necessarily always distinctly advantageous for the detection and identification of enemy objects and that the need for stereo capability should be demonstrated before new display equipment is developed for use of military image interpreters. It was recommended that further research be conducted in order to examine systematically stereo vs non-stereo interpretation across functions other than free search, taking into account such photo factors as quality, content, and format. Until such additional research is accomplished, any decision to eliminate stereo cover should be made only after careful consideration. The possible uses of aerial cover can not always be predicted. Future research may determine a definite requirement for stereo in other interpreter functions such as terrain and trafficability studies.

Vertical and Oblique Viewing. The value to image interpreters of examining both vertical and oblique views of a target area rather than views of either type alone was explored (Sadacca, Ranes, and Schwartz, 1961). The task required of five matched groups of interpreters was varied so as to provide a basis for comparing performance when either vertical or oblique photos were used alone and when the two types of photo were used in combination. No significant differences were found in number of correct identifications made when interpreters had both types of photo or either type alone. More misidentifications were made when a second type of photo was provided after an initial viewing period of a single type. Changing the order of presentation of the imagery did not appear to affect performance. Thus, results of this rather limited exploratory study--the degree of slant or departure from verticality of the oblique photos was slight--indicated that having both vertical and oblique photos of a target area does not necessarily make for improved interpreter performance. As in the study of stereo effects, there is need for more comprehensive studies in which such factors as the scale, quality, obliquity, and content of photos are systematically varied.

Image Quality. The need has long been recognized for a systematic study of the relationships of such physical characteristics of imagery as image scale, contrast, granularity, and sharpness to the overall usefulness of imagery for the extraction of intelligence information. To date, many investigations have been conducted in this area, but none has solved the problem completely. Research in this area has suffered from the limited range of the physical variables studied, the artificiality of the imagery employed, the lack of experimental controls utilized, and inadequate performance criteria. In recognition of the need for a definitive study in this area, a research program was recently initiated to determine the optimal photo quality combinations for image information extraction. Prime consideration has been given to making the research findings operationally meaningful in terms of enabling both interpreters and laboratory technicians to select better imagery rapidly.

Some exploratory work has been accomplished on the consistency and predictability with which interpreters make overall quality judgments of prints. Twenty-four prints were prepared from each of three negatives. The three negatives, taken by air reconnaissance missions over parts of East Asia, showed a variety of strategic and tactical objects. The 24 prints were generated by varying the scale (4 levels), contrast (3 levels), and sharpness (2 levels) of the reproduced imagery (4 x 3 x 2 = 24). The three variables were introduced systematically and identically into each of the three sets of 24 prints. Intercorrelations could then be computed among quality ratings given prints of the three scenes. The mean intercorrelation across the three scenes of the average rankings given the 24 prints was .92. The mean multiple correlation coefficient for predicting the average print ranking from the scale, contrast, and sharpness variables was .91, with sharpness and scale carrying most of the prediction. Eighteen image interpreters ranked the 3 sets of prints in order of usefulness for the extraction of intelligence information. The average split-half reliability of the judged ranks was .98.

These data indicate that prints can be ranked reliably on usefulness and photo quality; that such overall rankings are highly predictable from the image variables; and that prints similarly treated but from different negatives will tend to be ranked equivalently. If subsequent experimentation indicates that overall rankings correlate significantly with performance and that judgments of image variables such as contrast and sharpness can also be validly and reliably made, considerable progress will have been achieved in indicating profitable areas for future image quality research.

Studies of Interpreter Operations and Techniques

Interpreters' Confidence in their Reports. One of the first studies conducted in the area of interpreter operations and techniques was to determine whether the degree of confidence interpreters expressed in their identifications was related to the accuracy of their identifications. In the operational situation, an interpreter normally indicates the degree of confidence or certitude he has in an identification by one of the following qualifiers: "Positive", "Probable", or "Possible". The same procedure was followed in the performance measures. Table 1 shows the mean confidence level for right and wrong identifications in five performance measures. The mean confidence level for identifications later proved to be right was consistently higher than for identifications later proved to be wrong. This finding has practical implications. If systematic advantage can be taken of accompanying expressions of confidence, the amount of erroneous information produced by interpreters may be substantially reduced.

Table 1

MEAN CONFIDENCE ASSIGNED TO RIGHT AND WRONG IDENTIFICATIONS
ON FIVE PERFORMANCE MEASURES
(Positive = 1; Probable = 2; Possible = 3)

Performance Measures	Mean Confidence Right	Mean Confidence Wrong
Tactical - 1	1.32	1.99*
Tactical - 2	1.54	1.86*
Tactical - 3	1.39	1.81*
Strategic - 1	1.31	1.70*
Strategic - 2	1.23	1.94*

*Significantly different ($P < .01$)

As a first step in this direction, the effect of increasing the customary 3-point confidence scale used by interpreters to a 5-point scale was explored. Two groups of interpreters matched on the basis of final course grades and composite aptitude test scores were established. Group A (N = 58) used the regular 3-point scale. Group B (N = 57) used a 5-point scale with the following confidence levels: Highly Positive, Positive, Highly Probable, Probable, and Possible. Two performance measures, one involving tactical, the other strategic imagery, were administered under identical conditions to both groups of interpreters. Table 2 shows the results obtained. Note that with the 3-point scale, accuracy fell off sharply for identifications interpreters felt were only probable--from 32% to 13% in the tactical photos and from 71% to 45% in the strategic photos. The 5-point scale seemed to achieve a more even distribution of accuracy percentages.

However, it was soon discovered that there was an inherent weakness in the 5-point scale, and for that matter, in any scale whose points are defined with words like "probable" and "possible": The words mean different things to different interpreters. Poll of a small sample of experienced and inexperienced interpreters revealed that wide differences in subjective probability were attached to the different words by interpreters. For example, when one experienced interpreter described an identification as "possible", he expected to be correct about 85% of the time. Another experienced interpreter used the same term when he expected to be correct about 20% of the time.

Table 2

PERCENT ACCURACY OF IDENTIFICATIONS MADE BY INTERPRETERS USING A 3-POINT AND A 5-POINT SCALE OF CONFIDENCE IN THEIR REPORTS

Confidence Scale Values	Tactical Measure Group A 3-pt scale	Tactical Measure Group B 5-pt scale	Strategic Measure Group A 3-pt scale	Strategic Measure Group B 5-pt scale
Highly Positive		37		81
Positive	32	20	71	66
Highly Probable		11		46
Probable	13	11	45	38
Possible	07	05	36	36
Overall Accuracy	18	15	58	61

Subsequently, a numerical scale was introduced into the experimental studies. Interpreters were asked to estimate directly the probability that their identifications were correct by placing a 90 next to those identifications which they felt would be correct 90% of the time, an 80 next to those identifications they felt would be correct 80% of the time, and so forth. Experience with this new method of indicating degree of confidence has been promising. Interpreters readily understand it, and it allows specification of a wide range in confidence values. Difficulties still remain, however. Many interpreters fail to discriminate accurately their internal feelings of confidence and report the same confidence level for all their identifications. Also, interpreters for the most part over-estimate their probability of being correct. Perhaps additional instruction or training in confidence specification would be desirable. Experiments along these lines will be conducted shortly. It seems, however, that a numerical scale offers a convenient, easy, and unambiguous way of specifying degree of confidence felt. Such a scale will probably be recommended for use by operational interpreters if its usefulness is borne out in analysis of currently available data.

Impact of Additional Intelligence Information. Two closely related studies were undertaken to determine whether interpreter performance is influenced by intelligence information regarding enemy presence or activity in the area covered by the imagery to be searched. Among interpreters furnished additional information suggesting the presence of enemy objects, relatively more were consistently above the median in number of correct identifications than among those not having such information. However, interpreters having the additional information also reported more objects where actually none appeared. When the additional information was accompanied by statements about the reliability of the information or its source, such as "confirmed" or "unreliable", there was no evidence that the suggestive impact of the intelligence information was affected. Even when the source was characterized as "completely unreliable" and the intelligence information as "improbable", the impact of the information was not diminished (Sadacca, Castelnovo, and Ranes, 1961). Nor was there any evidence of significant differences in performance in response to suggestive information between officers and enlisted men or between interpreters with above average school grades and interpreters with low grades. Average confidence levels were also apparently not affected by the presence or absence of the additional information.

In view of these findings, a recommendation was made that in those instances where accurate reporting is essential, one interpreter, at least, should work without knowledge of available additional intelligence information, if only for an initial period.

There was also evidence in these studies that the very fact that interpreters are asked to search for specific objects may suggest the presence of the objects. It was therefore recommended that the feasibility and value of employing standard checklists for operational reporting should be considered. Further research to determine techniques by which intelligence information can be imparted to interpreters without undue suggestion is needed.

Studies of errors in identification. A number of experimental studies have been conducted to find promising techniques to reduce errors in interpretation. Errors made by interpreters were found to subdivide into two types. Interpreters may misidentify militarily important objects--what is really a bunker, for example, may be identified as a tank; or a cement plant may be identified as a flour mill. This "misidentification" type of error is relatively common in reports for strategic imagery, where the objects tend to be relatively large and well defined. Interpreters may also make what can be termed "inventive" errors, that is, they identify insignificant features of the photographs--tonal differences in a road, for instance--as significant objects. The tendency to make "inventive" errors is more common in the interpretation of tactical photographs, in which the targets are relatively small and dispersed and in which there are many variations in shape and tone. In this regard, it must be remembered that aerial imagery is often of very small scale and photo quality is not always the best.

One of the first error studies conducted explored the effect of feedback--the effect on current performance of giving the interpreters some idea of how well they had previously done. Three groups of matched image interpreters were administered the same tactical performance measure (based on photos made on a flight over enemy-held territory in East Asia) under standard conditions. Later, interpreters in the first group, Group A, were told their scores--that is, the number of correct and erroreous identifications they had made, their accuracy level, and the number of objects (vehicles and revetments) they had failed to identify.

Interpreters in Group B were allowed to score their own reports so that they could see for themselves the errors they had made. Interpreters in Group C were given no information about their prior performance. All three groups were then given a second performance measure using photos taken in the same flight line.

No statistically significant differences were found among the three groups in mean number of correct identifications made or in average accuracy levels. However, the groups did differ significantly in mean number of wrong identifications. Fewest errors were made by the men who scored their own reports (Group B in Table 3). Those who were given general feedback made the most errors. The result suggests that knowledge of how well they are doing may be effective in reducing the number of errors made by interpreters, especially if feedback is made in such a way that interpreters can see exactly what kind of errors they are making.

Table 3

MEAN ERROR SCORES OBTAINED BY GROUPS DIFFERING WITH RESPECT TO KNOWLEDGE OF SCORES ON A PRIOR PERFORMANCE MEASURE
(N = 30 in each group)

Group A (Interpreters were told their scores)	Group B (Interpreters scored own papers)	Group C (No knowledge of performance)	Significance (Analysis of Variance)
16.57	9.70	11.60	$P < .01$

Several other techniques were studied in an attempt to reduce errors. In one experiment, interpreters were told to list for each identification they made the principal reason they felt the object was what they had identified it to be. In another experiment, interpreters were asked to state the actual length and width of each vehicle they identified. In a third experiment, interpreters were asked to identify the objects of interest to a finer level of detail than ordinarily demanded in the performance measure--to specify a trench as a "trench, communication, zigzag w/weapons emplacements", rather than simply as a "trench". The three techniques had in common the rationale of forcing the interpreter to pay close attention to the objects he was examining. Unfortunately, there was no evidence that interpreters working under these experimental conditions made fewer errors than interpreters in control groups who did not have the added requirements.

One technique which did show promise, however, was to have interpreters re-examine imagery they had interpreted once before. In one experiment, where the same tactical performance measure was readministered two and one-half days after the initial interpretation, accuracy level for objects which interpreters reported both times they examined the imagery was significantly higher than for objects reported only once (53% vs 34%). The result, although not supported in the case of a strategic measure administered under the same conditions, suggests that allowing interpreters to take a second look at imagery after a time lapse may prove to be useful in eliminating some erroneous identifications.

Present plans call for a continuation of experimental studies in error reduction. It is planned to conduct a detailed analysis of the types of error interpreters make with the aim of building up an "error" key which would list frequently made errors so that interpreters can learn to avoid them. The effect of providing short review periods in which interpreters examine annotated photos of the objects of interest will be studied. The annotated objects will be of similar scale and will appear under similar conditions as the objects in the test imagery to be given shortly after the review period. Variations in procedures for allowing interpreters to re-examine imagery will be explored, as will also varying methods of feedback.

Studies of Interpreter Output

Output Rate and Accuracy. To determine whether the rate and accuracy of information extraction varies systematically over time, the cumulative number of identifications made after each five-minute period of viewing time was computed for right, wrong, and total number of identifications for several performance measures. These values, when plotted, resulted in fairly uniform curves. Figure 1 shows the curves for a typical example--a tactical measure administered with a 30-minute time limit. Notice that the average number of new identifications scored right tapered off much sooner than did identifications that were wrong. Although the interpreters were still responding at the end of the 30-minute time period, the later information contained a higher proportion of wrong identifications than did information produced earlier in the viewing period.

The effect is seen more clearly in Figure 2 which shows accuracy plotted as a function of time for the same performance measure. The average accuracy of interpreters fell off the longer they examined the imagery. Evidently, time does not necessarily work to the advantage of the image interpreter. After examining imagery for a while, interpreters may begin to respond to doubtful cues, cues that are on the threshold of resolution.

Additional analysis of existing output data to further determine the effect of viewing time on performance is planned. A critical problem of optimization or "trade-off" is involved. If an interpreter's time is cut short so that he will not be induced to make doubtful identifications,

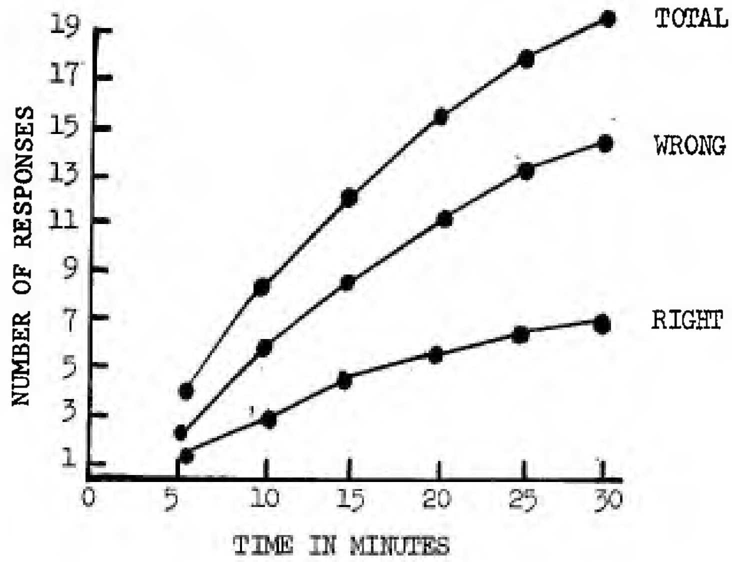


Figure 1. Representative Cumulative Curves for Right, Wrong and Total Responses (For a tactical image interpreter performance measure)

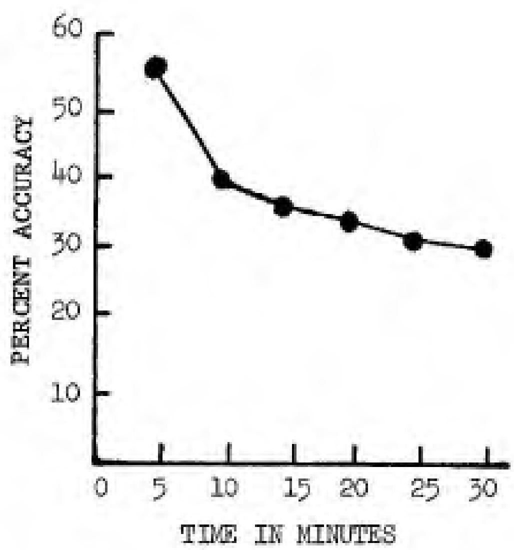


Figure 2. Representative Cumulative Accuracy Curve for data in Figure 1

some correct identifications will probably be lost. It is hoped through the planned additional analysis--and, if necessary, through further experimentation--to develop new techniques for establishing optimal viewing time.

Factor Analysis of Output Variables. To achieve a better understanding of the basic relationships underlying performance, the scores of interpreters on 12 performance measures were intercorrelated and factor analyzed. In six of the measures, objects in the imagery were mostly of a tactical nature; in the other six measures, objects were of a strategic type. The task of the interpreters in nine of the measures was to detect and identify objects of military significance through free search of the photos involved. Both right and wrong scores were used. In three strategic measures, the task of the interpreter was to identify specific annotated areas on the photos. For these three measures, only the number of correct identifications was used in the analysis. The final course grades of the 105 officers who comprised the sample were also included in the analysis.

Three factors were clearly identified: Factor I, with positive loadings attained by all the right scores, and Factor II, with positive loadings attained by all the wrong scores, clearly represent "right" and "wrong" factors. Evidently there is no relationship between the right scores and wrong scores attained by the subjects. This conclusion is borne out by an average correlation of only -.05 between the nine pairs of right and wrong scores obtained from the same performance measure. Findings indicated that although interpreters may tend to perform consistently across a variety of image content areas in terms of number of right or wrong identifications made, there may be very little relationship between the two aspects of their performance.

Some differentiation between tactical and strategic wrong scores was suggested by the apparent make-up of Factor III. All strategic wrong score factor loadings were negative; most tactical wrong score loadings were positive. If the finding is corroborated in further studies, a possible basis for the differential assignment of interpreters into content areas may be established.

DEVELOPMENT OF SELECTION TECHNIQUES FOR IMAGE INTERPRETER PERSONNEL

The second broad research area or subtask is concerned with the development of instruments for use in selecting image interpreters. The major objective was to develop means of identifying, prior to actual training and assignment, personnel well qualified for image interpretation work. A related objective was to determine what aptitudes, personality factors, and background experiences are most associated with high interpreter performance. As a first step in conducting the research, an analysis was conducted of current interpreter job requirements as well as likely future requirements. Based on these requirements, an eight-hour experimental test battery was developed, in which were included tests measuring spatial and perceptual abilities, reasoning, memory, and vision as well as interest and personality factors (Table 4). Performance measures were used in conjunction with course grades as criteria against which to validate the experimental predictors.

Table 4

LIST OF EXPERIMENTAL IMAGE INTERPRETER SELECTION TESTS

Spatial and Perceptual Tests	Visual Tests
Image Orientation	Stereopsis
Perceiving Relations	Phoria, far vision, vertical
Spatial Orientation, Part I	Phoria, near vision, vertical
Spatial Orientation, Part II	Phoria, far vision, horizontal
Camouflaged Outlines	Phoria, near vision, horizontal
Mutilated Words	Acuity, far vision, best eye
Estimation of Length	Acuity, near vision, best eye
Object Completion	
Speed of Identification	**Army Classification Battery Tests**
Reproduction of Visual Designs	
Figures	Verbal (Vocabulary)
Aircraft Orientation	Arithmetic Reasoning
Hidden Faces	Pattern Analysis
	Mechanical Aptitude
Reasoning Tests	Army Clerical Speed
Mathematical Reasoning	
False Premises	**Miscellaneous Tests**
Reasons	
Figure Analogies	Mechanical Principles
Practical Situations	Reading and Vocabulary
Figure Classification	Image Interpreter Self-Description Blank
	Personal Data Record
Memory Tests	Things Round
Sentence Completion	Image Interpreter Information
Meaningful Memory--Numbers	

For approximately 18 months beginning early in 1960, the battery--with the exception of the vision tests--was administered under controlled conditions to all students entering the basic interpretation course at the Intelligence School. The vision and performance measures were administered immediately after training was completed, just prior to graduation. Complete data were collected for 120 officers (including 105 interpreters used in the factor analysis study previously described) and 65 enlisted men.

Data on the officer and enlisted samples were analyzed separately. Army Classification Battery test scores were included for the enlisted sample in order to assess the predictive value of using tests in the experimental battery in addition to already existing ACB scores. Standard test selection procedures (Wherry-Doolittle) were performed on the data. Both final course grades and performance test scores were used as criteria in the test selection process. Strategic right and tactical right scores on the performance measures were computed. Wrong scores were not used as criteria in the test selection procedure because of the general lack of relationship found between predictor scores and wrong scores.

Battery Recommended for Selection of Officer Interpreter Trainees

Table 5 presents validity coefficients of the tests selected for use with officers. The substantial correlation between the battery composite score and the criteria indicates that considerable improvement in the quality of interpreter personnel could be achieved through the use of the recommended selection battery. With an expected cross-validity coefficient of approximately .60 between the selection battery score and course grades, the number of relatively unsatisfactory students in the image interpretation course could be cut to less than one-third the present number if the course input could be selected from a large enough pool of potential assignees (that is, if the selection ratio was about 2.6 potential assignees to every 1 selected).

Table 5

RECOMMENDED SELECTION BATTERY AND VALIDITY COEFFICIENTS OF INDIVIDUAL TESTS IN A SAMPLE OF 120 OFFICER INTERPRETER TRAINEES

Tests	Raw Score Wts.	Administration Time (Min.)	Validity Coefficients: Course Grades	Tactical Rights	Strategic Rights
Spatial Orientation, Part II	2	30	.52	.54	.56
Mathematical Reasoning	2	26	.56	.37	.38
Aircraft Orientation	1	18	.38	.48	.36
Image Interpretation Information	1	36	.56	.52	.58
Battery Composite[a]		110	.68	.63	.63

[a] A shorter composite battery consisting of the Spatial Orientation and Information Tests, weighted equally, may be used if time does not permit administration of the full battery.

Selection of Enlisted Interpreter Trainees

In analyzing the data from the enlisted sample, emphasis was placed on determining whether prediction attained with the present selection measure (the General Technical (GT) Aptitude Area of the Army Classification Battery) could be enhanced sufficiently by any of the tests of the experimental battery to justify their operational use. When corrected for restriction in range, GT was found to correlate .76 with the final course grades of the enlisted men and approximately .60 with the number right performance criteria. Although validity could perhaps be improved by the addition of the Spatial Orientation, Part II and the Image Interpretation Information tests, the expected improvement was slight and probably would not justify the cost of adding these tests to the present method of selection. If a higher quality of enlisted trainees is necessary, the cutting score (presently 100) of the General Technical Aptitude Area should simply be increased.

Vision Measures in Selection

Vision measures of stereopsis, acuity, and phoria failed to correlate to a large degree with course grades or performance measure scores. Apparently, the task of the image interpreter is a highly cognitive one. If a man can pass the basic visual standards now used in selecting students, his ability to make the complex decisions and judgments necessary in image interpretation will depend mostly on his mental capacities, interests, and experience.

Work in the area of selection research will continue. Analysis of background and experimental data collected from students as well as from interpreters in the field will begin shortly. Data is currently being collected for cross validation of the selection battery. The best time and place to administer the operational selection battery must be determined. Here the most important problems involve finding a substantial pool of officers from which to conveniently select interpreter trainees and assuring that the selection process does not drain off too many high aptitude trainees and thus prevent other vital personnel requirements of the Intelligence community from being met. In addition, the possibility of achieving effective differential classification and assignment within the interpreter field will be explored--special aptitudes may be needed to extract information from multi-sensor displays or to detect targets in rapid screening or in near-real-time interpretation.

UTILIZATION MEASURES UNDER CONDITIONS OF EMERGENCY DEMANDS

This subtask deals with the mobilization and utilization of interpreter talent under emergency conditions. In any future combat setting, aerial surveillance is expected to provide invaluable up-to-date information concerning enemy activity. The volume of imagery necessary to maintain adequate coverage of dispersed, mobile enemy forces is large.

In examining this imagery, interpreters may be forced to work as long as their energies can hold up. Critical questions concerning the human factors in interpreter output under these stress conditions are raised: What is the point beyond which the image interpreter can not produce useful information? How can fatigue be best delayed through spaced rest periods? How rapidly can an interpreter recover lost efficiency? Work has recently been initiated in this vital area. Experiments have been designed to compare performance of three groups of interpreters: interpreters given rest periods, interpreters with no rest periods working throughout on the same imagery, and interpreters with no rest periods but given different kinds of imagery to interpret. Basic work decrement curves will be established by administering performance measures without rest to matched groups of interpreters.

In conducting this research, every effort will be made to motivate the experimental subjects. In any research on work-rest cycles, the problem of isolating motivational effects from physical fatigue is crucial. Success of the research will depend on how well differentiation can be made between lack of interest or motivation and real physical inability to continue working adequately.

Another vital research area in this subtask deals with the flexibility or "set" of interpreters under a variety of demand conditions. For example, an interpreter may be told: "Find a missile site in this batch of imagery within 10 minutes." What effect does this kind of instruction have on the interpreter's output? Does he alter his search techniques to one more commensurate with time allowed? Or can an interpreter adjust his normal performance when he is told: "Our missile stock is low. We must be certain before we expend any more of our missiles that a target is really there. Report only enemy missile sites which you are absolutely positive about."

Problems involved in adjusting performance set are related to those discussed earlier in connection with interpreters' feelings of certitude or confidence. Also involved is the problem of optimization raised in discussing output rate and accuracy. If the military situation demands that all targets be identified, interpreters may begin to respond to doubtful cues and achieve completeness at the expense of accuracy. On the other hand, a military demand for high accuracy may induce interpreters to ignore doubtful images and to search only for clearly discernible targets.

In one experiment just completed, results indicated that interpreter performance will change depending upon whether accuracy or completeness is emphasized in initial instructions. If subsequent experimentation can delineate the range and variability of the mental sets with which interpreters approach their tasks, valuable guidance can be given to commanders as to how their intelligence requirements can be better met through more precise instructions to interpreters. If a commander, for example, can specify the relative importance of accuracy vs completeness, the interpreter may well be able to adjust his set and adopt techniques which will enable him to perform close to optimally within the given military situation.

EFFECTIVE GROUP APPROACHES TO ACCOMPLISH MAJOR IMAGE INTERPRETER MISSIONS

This subtask area, unlike the preceding areas which are concerned with individual abilities and techniques, deals with effective group approaches to image interpreter missions. Interest is focussed here on the tactical image interpreter unit and its performance. The objectives of the experimental studies are to determine the optimal working size of image interpreter units, to determine efficient combinations and distributions of skills and skill levels within the unit, and, most important, to determine the standing operating procedures and sequences of unit operations necessary for the timely accomplishment of the unit's missions.

In preliminary experimentation with various group approaches, it soon became apparent that teamwork procedures might provide a powerful technique for reducing the number of erroneous identifications made by interpreters and at the same time increase the number of correct identifications. Team procedures thus seem to offer a way out of the accuracy-completeness dilemma discussed earlier in regard to interpreter confidence and viewing time. In our experimentation we did not assume that merely by working together a group of image interpreters would automatically produce more accurate and complete identifications. We arrived at a fundamental basis for determining how the efforts of a group of interpreters could best be pooled--the concept of correlation of performance among individuals. This way of looking at team procedures appears to hold promise of real improvement in performance.

The correlation concept was applied to three different methods of obtaining intelligence information from interpreter teams. In Method I, the image interpreters worked and reported independently. The performance of two interpreters--to take the simplest case--was scored for right identifications, wrong identifications, and omits (targets appearing in the imagery but not reported). Table 6 shows schematically the correlation between performances of the two individuals and how each element of performance--rights, wrongs, omits--enters into the final product.

If there were perfect agreement between the two interpreters, every right identification made by one would also have been made by the other. The same would obtain for wrong identifications. Another way of stating this principle: when correlation between the performance of two interpreters is perfect, or equal to unity, there are no entries in the off-diagonal cells of Table 6. Perfect agreement, however, is obviously not what is desired when interpreter efforts are pooled. For if the interpreters agreed completely to begin with, no gain would result from pooling their efforts. Method I can be expected to produce the maximum number of correct identifications to be obtained from the two interpreters, as shown from the computation of total scores indicated in Table 6. However, the information resulting from this method also includes all the wrong identifications made by both interpreters.

In Method II, individuals comprising the teams also worked and reported independently, but their identifications were pooled in a different way. Only identifications upon which the two interpreters agreed were considered as having been made by the team. All identifications made by only one member of the team were disregarded. The formulas for total scores under Method II (Table 6) indicate that total number of right identifications, as well as the total number of wrong identifications, would almost certainly be less than for Method I. Whether accuracy of interpretation with Method II is in general higher than with Method I depends upon whether image interpreters tend to agree more on right identifications than on wrong identifications. Since it is reasonable to suppose that interpreters will tend to agree more in reporting targets that are actually present than in reporting targets which are not present, proportionally more right identifications than wrong identifications would be expected to survive the agreement criterion established for Method II.

In Method III, interpreters worked in teams instead of independently. Each member of a team was given complete sets of photographs, but only one report sheet was given to a team. The team was instructed to record only those identifications which both members of the team agreed upon (at least two members in the case of three-man teams). Team members could discuss the identifications freely, and go about cooperating in any way they desired.

It was expected that as a result of their discussion interpreters would agree on some targets on which they previously disagreed. Table 7 shows the expected effect of this tendency on the elements involved in image interpreter performance correlation. We see a general movement from the off-diagonal cells to the diagonal cells and, as indicated by the heaviness of the arrows, relatively more agreement on right responses than on wrong responses. Consider, for example, targets which one interpreter correctly identified but which the other interpreter omitted. When interpreters compare identifications, their attention is on the particular targets being compared, and chances are that the correct information will be reported. We would expect considerably more agreement upon right responses for Method III than for Method II; but, because some right identifications will not be agreed upon, the total number of rights from the cooperative method will be less than for Method I, which pooled all right responses of team members.

With respect to wrong responses, one would expect an even greater tendency for wrong identifications made by only one interpreter to drop out, as the other interpreter would probably be more resistant to agreeing to misidentifications. Considering that interpreters are more confident of identifications that are right than they are of their wrong identifications, there is a good chance that the interpreter making a wrong identification would not argue as strongly for it as for his right responses. It was to be expected, then, that wrongs would be fewer in number than would be obtained by Method I and that, although the rights would also be fewer, overall accuracy would be improved--although not as much as with Method II, where there was no opportunity for teammates to induce each other into agreeing to misidentifications.

Table 6

PATTERN OF RESPONSES FOR TWO INTERPRETERS
WORKING INDEPENDENTLY
(Methods I and II)

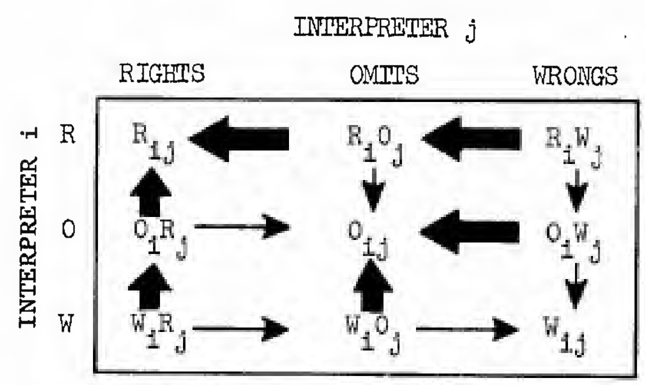
INTERPRETER i \ INTERPRETER j	RIGHTS	OMITS	WRONGS	Σ
R	R_{ij}	R_iO_j	R_iW_j	R_i
O	O_iR_j	O_{ij}	O_iW_j	O_i
W	W_iR_j	W_iO_j	W_{ij}	W_i
Σ	R_j	O_j	W_j	

TOTAL SCORES

$R_T = R_i + R_j - R_{ij}$	(Method I)
$W_T = W_i + W_j - W_{ij}$	(Method I)
$R_T = R_{ij}$	(Method II)
$W_T = W_{ij}$	(Method II)

Table 7

PATTERN OF RESPONSES FOR TWO INTERPRETERS
WORKING COOPERATIVELY
(Method III)



Legend

R_{ij} = targets both interpreters got right

W_{ij} = targets both interpreters got wrong

O_{ij} = targets both interpreters omitted

R_iO_j = targets Interpreter i got right but Interpreter j omitted

R_iW_j = targets Interpreter i got right but Interpreter j got wrong

Other elements in matrix similarly defined

How closely empirical data corresponds to expected performance is shown in Table 8 which summarizes some results of the experiment. When all identifications made by two or three interpreters working independently were pooled (Method I), a much larger number of correct identifications was obtained (mean of approximately 15 as against 8 for individual interpreters). However, by so pooling all identifications, many wrong identifications were retained. When, instead of pooling all identifications, only those identifications were retained which the independent interpreters agreed upon (Method II), a marked increase in overall accuracy resulted. In the three performance measures used here, accuracy level rose to 33%. However, many fewer correct identifications were retained--only about four, on the average. Notice the enormous drop in number of wrong responses when only the agreed-upon identifications were considered. Apparently, interpreters do not tend to agree on wrong identifications--erroneous responses are idiosyncratic, that is, peculiar to the individual. If one interpreter calls a tonal difference on the road a vehicle, chances are that another interpreter will not agree with him, although the second interpreter may call another tonal difference a vehicle. The cooperative method, in which interpreters in a team discussed their identifications freely, appeared to offer a reasonable compromise between the extremes of the methods just outlined. Definite improvement over individual interpretation in terms of rights, wrongs, and accuracy scores was achieved by the cooperating teams (see Table 8).

Table 8

TOTAL MEAN PERFORMANCE SCORES FOR TEAMS WORKING UNDER DIFFERENT METHODS AND FOR INDIVIDUALS WORKING ALONE

Team Methods*	Right	Wrong	% Accuracy
Independent: All response (N = 15)	15.1	112.9	14
Independent: Agreed responses (N = 15)	3.9	7.1	33
Cooperative (N = 30)	10.1	40.6	22
Individuals Alone (N = 36)	8.1	50.8	16

*Mean Right, Wrong, and Accuracy scores between team methods are significantly different ($P < .01$).

Subsequent experimentation provided evidence reinforcing the promise of team procedures. For this reason, comprehensive experimentation was undertaken from which more definitive conclusions concerning teamwork methods could be drawn. Data collection for this experiment has recently been completed. In all, 10 team methods were tried, involving variations in team size, procedures for examining prints, and amount of communication among team members. Each team accomplished 15 performance measures over a 2 1/2-day period. The performance measures varied in administrative time from 20 minutes to 2 hours. For the first time, data were collected on interpreters in operational units in the field in addition to data from the Intelligence School. Altogether, 66 operational interpreters and 110 graduating interpreters were used in the experimental sample.

The experimentation should not only yield important comparisons among various team methods and compositions, but should also provide much useful information for determining adequate confidence scales, efficient viewing periods, and performance differentials between operational and graduating interpreters. It must be emphasized, however, that even upon completion of the analysis for this experiment, much research in team methods will still be needed. The whole problem of team vs individual viewing of multi-sensor imagery is one of many that remain to be investigated. As priorities and manpower permit, additional studies in this vital area will be conducted.

EVALUATION OF CURRENT STATUS AND FUTURE PLANS

Almost every study undertaken has revealed new problem areas. Image interpretation is an extremely complex process. Impinging upon an interpreter's performance are factors related to his background, training, aptitudes, interests, and personality as well as his momentary state of motivation and fatigue. Factors related to the amount, quality, and content of the input imagery, the time available for viewing the imagery, the kinds of instruction and target information the interpreter has received, and the techniques and equipment he uses in examining the imagery also play major roles in determining performance. Performance itself can be measured in a variety of ways--number of correct identifications, number of errors, percent accuracy, percent completeness, rate or speed of information extraction. And, of course, the task of the interpreter can vary--trafficability and functional analysis, detection, identification, location, and mensuration. Too, interpreters are highly variable in regard to both intra- and inter-individual differences: On one performance measure, errors ranged from zero to over fifty, and completeness from zero to 100%.

Definitive answers to human factor problems of this complexity come slowly. However, early in 1963, an operational selection battery will be recommended. Also in 1963, recommendation of effective teamwork procedures should be practicable. Toward the end of 1964, suggestions may be made for reducing the error rate of interpreters. Also, results of research on image quality research may be available. If the requisite imagery can be obtained during 1963, effective multi-sensor viewing techniques may be identified by the end of 1965.

In addition to work on these immediate problems, effort is being devoted to the development of a theoretical model encompassing the many diverse factors that impinge on interpreter performance. Current developments in signal detection and statistical decision theories show promise of being applicable to image interpretation processes. Out of such theoretical development it is hoped to arrive eventually at equations with constants specific to the individual and occasion which will describe an interpreter's performance over a variety of job situations. Effort will be made to develop equations which simultaneously take into account the constants of interpreter performance, quality of imagery viewed, number of interpreters agreeing on an identification, and the other objects in the area that have been identified either by interpreters or other intelligence sources. The resultant statements of probability that a given identification is correct would be invaluable to intelligence officers in evaluating the output of individual interpreters as well as the output of the reconnaissance system as a whole.

Another research area with perhaps more immediate payoff value would be the development of performance feedback techniques for use on computer facilities installed in operational tactical image interpretation vans or facilities. If, as seems likely, computers will be employed in the mobile vans of tomorrow, they could be an excellent tool for maintaining interpreter proficiency during peacetime. The keys to performance measures could be placed on the van's computer memory drum in very much the same way as keys are now placed in our computer when we score interpreter responses. The interpreter could obtain immediate feedback as well as summary reports by target type as to the accuracy and completeness of his identifications. The feedback and error experiments now being conducted would provide background information as to how best to maintain and enhance interpreter performance. Additional experimentation with specific computer feedback mechanisms would of course be necessary. The personal equations and constants discussed above could, incidentally, be determined by the computer directly from the performance indices obtained on the practice exercises.

If a computer were available, group experiments could be conducted in which the identifications made by one team member could be fed directly into a computer, compared with those of his teammates, and a composite report prepared. The computer could also direct one teammate's attention to particular frames or image areas for confirmation of another teammate's identifications. Computers and tape transports could also be used as rapid access files to store all available information about particular geographic areas. Experiments utilizing various display techniques and devices could determine what background information is most needed by interpreters and how, when, and where the information could best be presented. To conduct such studies in a timely and controlled fashion it would be highly desirable to be able to simulate operational conditions in a laboratory setting. The Support Systems Research Laboratory is currently exploring the possibility of obtaining a computer-display facility for experimental purposes. Plans would then be modified to make maximum use of the research opportunity afforded by such a facility.

Innovations will continue to affect the role of the human interpreter in aerial reconnaissance systems. Platforms, sensor devices, and image processing and display equipment are continuously changing. Military requirements are changing; interpreter training and techniques are changing. The nature and direction of the research effort must change also if its findings are to have practical value. Conclusions can not be expected to hold over all possible present and future situations. However, through continuous studies of significant aspects of performance, some useful guidelines can be provided to make the human being more effective in extracting and processing intelligence information derived from imagery.

REFERENCES

Selected Publications of the US Army Personnel Research Office dealing with IMAGE INTERPRETATION Research

Birnbaum, A. H. Effect of selected photo characteristics on detection and identification. Paper presented at the Symposium on Photo Interpretation Commission VII, International Society for Photogrammetry. Delft, Netherlands. 3 September 1962.

Birnbaum, A. H. An evaluation of photo display modes for extraction of intelligence information. Paper presented at the Army Human Factors Engineering Conference. Fort Benning, Georgia. 17 October 1962.

Birnbaum, A. H. Human Factors Research in image systems--Status report, 30 June 1962. Technical Research Note 122. June 1962.

Mendelson, M. A. Comments of interested agencies on planned image interpretation research. Research Memorandum 59-6. July 1959.

Sadacca, R. The accuracy and completeness of individual and team photo information extraction. Paper presented at the Sixth Annual Human Factors Engineering Conference. Fort Belvoir, Virginia. 3-6 October 1960.

Sadacca, R. New techniques in image interpretation systems. Paper presented at the 7th Annual Army Human Factors Engineering Conference. University of Michigan. Ann Arbor, Michigan. 3-6 October 1961.

Sadacca, R. Techniques for optimizing image interpreter performance. Paper presented at the Army Science Conference. West Point, New York. June 1962.

Sadacca, R. Choosing among objective performance criteria. Paper presented at the American Psychological Association Convention. St. Louis, Missouri. August - September 1962.

Sadacca, R., Castelnovo, A., and Ranes, J. Human factors studies in image interpretation: The impact of intelligence information furnished interpreters. Technical Research Note 117. June 1961.

Sadacca, R., Ranes, J. E., and Schwartz, A. I. Human factors studies in image interpretation: Vertical and Oblique Photos. Technical Research Note 120. December 1961.

Zeidner, J. Human factors considerations in the performance of image interpreters. Paper presented at the Tri-Service Military Conference. Pensacola, Florida. 25-27 May 1960.

Zeidner, J. Some human factors studies in the Army's command control information systems 1970. Paper presented at the 10th Military Operations Research Symposium, Santa Monica, California. 16 October 1962.

Zeidner, J. and Birnbaum, A. H. Survey of psychological studies in image interpretation. Research Study 59-1. February 1959.

Zeidner, J., Sadacca, R., and Schwartz, A. I. Human factors studies in tactical image interpretation: The value of stereoscopic viewing. Technical Research Note 114. May 1961.

AD Div 23/1, 28/4

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by
Robert Sadacca, Harold Martinek, and Alfred I. Schwartz.
October 1962. Rept on Image Interpretation Task. 33 p. incl.
tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129)
(DA Project 0J95-60-001) Unclassified Report

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by
Robert Sadacca, Harold Martinek, and Alfred I. Schwartz.
October 1962. Rept on Image Interpretation Task. 33 p. incl.
tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129)
(DA Project 0J95-60-001) Unclassified Report

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by
Robert Sadacca, Harold Martinek, and Alfred I. Schwartz.
October 1962. Rept on Image Interpretation Task. 33 p. incl.
tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129)
(DA Project 0J95-60-001) Unclassified Report

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.

AD Div 23/1, 28/4

U. S. Army Personnel Research Office, OCRD, DA
IMAGE INTERPRETATION TASK--STATUS REPORT, 30 JUNE 1962 by
Robert Sadacca, Harold Martinek, and Alfred I. Schwartz.
October 1962. Rept on Image Interpretation Task. 33 p. incl.
tables, figures, 14 Ref. (USAPRO Technical Research Report No. 1129)
(DA Project 0J95-60-001) Unclassified Report

UNCLASSIFIED
Human Resources Research
--Personnel Utilization

In response to a research requirement of the Department of the Army, an extensive exploratory survey of human factors problems in image interpretation was undertaken in 1958 by the Support Systems Research Laboratory (APRO). The primary objective was to define problems which were amenable to research attack and which showed promise of meeting the Army's immediate and long-range needs. The APRO research effort was formulated to answer two broad questions: (1) What are the skills, abilities, and techniques necessary to extract intelligence information from conventional and newer types of imagery? (2) How can the Army best utilize its available human resources to cope with ever increasing variety of image types and at the same time maintain required standards of speed, accuracy, and completeness? The present report briefly describes the research objectives and summarizes the experiments conducted to date within each of four subtask areas encompassed by the research program of the Image Interpretation Task.