MEMORANDUM
RM-6158/1-PR
FEBRUARY 1970

TARGET DETECTION THROUGH VISUAL RECOGNITION: A QUANTITATIVE MODEL

H. H. Bailey

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

36

MEMORANDUM

RM-6158/1-PR

FEBRUARY 1970

TARGET DETECTION THROUGH VISUAL RECOGNITION: A QUANTITATIVE MODEL

H. H. Bailey

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the United States Air Force.

DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

The RAND Corporation

1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

PREFACE

This study is a part of Rand's continuing effort to support weapon systems analyses and performance predictions with detailed understanding of all aspects of a problem. For example, many operations (both military and nonmilitary) depend critically on a human observer's ability to search for and find a desired object or "target" amid background clutter within a limited time. Equations are developed in this Memorandum which permit the calculation of recognition probabilities as a function of the observed or displayed target contrast and size (angular subtense), the number of resolution cells across the minimum dimension of a target, the required search area and available search time, the false-target density or some other measure of scene congestion, and the signal-to-noise ratio.

The results should be helpful to both designers and users of all systems in which visual observation plays a significant role. In addition, the model can be used to formulate realistic display requirements for those systems in which a sensor is interposed between the observer and the real world.

SUMMARY

This Memorandum presents a model for describing analytically the capabilities and limitations of a human observer in the task of looking for and finding known or expected fixed objects. The description takes the form of six algebraic equations which together enable the user to estimate recognition probabilities as a function of the many parameters required to describe a specific situation. The model is tailored to the case of an airborne observer looking at terrain with or without optical aids or electro-optical sensors, but with prior knowledge of the approximate appearance of an object. In Air Force applications, it estimates the probability that a pilot or observer will be able to say, "There is the target!"

The model is structured according to three distinguishable psychophysical processes: deliberate search over a fairly well-defined area, detection of contrasts (a subconscious retino-neural process), and recognition of shapes outlined by the contrast contours (a conscious decision based on comparison with memory). In addition, when the observer is viewing a displayed image of a scene, noise is usually present which degrades his performance of these three steps. The probability that the three steps are completed successfully, multiplied by a noise degradation factor, gives the probability of target recognition.

A search term expresses the probability of looking in the right direction for the target as a function of the desired search rate (with the area normalized to the target area) and a measure of scene congestion or false-target density. A contrast term expresses the probability of spot detection as a function of the ratio of actual to threshold contrast. The latter is determined by the angular subtense of the target or its image at the eye. A resolution or shape-recognition term expresses the probability of recognition as a function of the number of resolution cells--be they equipment-limited or set by the observer's eye--contained within the shortest dimension of the target. A final term gives the degradation in recognition probability caused by image noise, expressed as a function of the signal-to-noise ratio.

In a narrow sense, the only values that need to be supplied by the user of this model are the apparent size and contrast of the target as seen by the observer, the desired search rate, and the congestion of the scene (defined as the average number of fixation points or false targets in an area 100 times the area of the target). In practice, particularly when artificial (e.g., electro-optical) sensors are used, additional information must be given about the displayed contrast, scale, resolution, and noise.

In view of the paucity and inconsistency of available experimental evidence, the accuracy of most inputs to the model (i.e., contrast, number of resolution cells, etc.) is expected to be no better than 20 to 30 percent; estimates of the congestion factor, another input, may well be in error by a factor of two or so in either direction. Hence the real utility of the model is in setting bounds to what should be expected of observers in real situations.

When applied reiteratively to successive designs in a systems context, the model serves to define--albeit loosely at present--the requirements that a human observer places on any system which he must operate.

ACKNOWLEDGMENT

The author is happy to express his appreciation to L. G. Mundie of The Rand Corporation for several very helpful discussions, particularly on the method of presenting this model. However, the basic concepts and the final formulation are the responsibility of the author alone.

CONTENTS

SYMBOLS

A_g = glimpse aperture

A_s = area to be searched

a_T = area of target

C = observed contrast (apparent or displayed)

C_o = intrinsic (zero-range) contrast

C_T = threshold contrast

f = spatial frequency

G = congestion factor, number of fixation centers per 100 a_T

K = a constant, set equal to 2.3

k = number of target areas (a_T) in an average glimpse aperture (A_g), i.e., A_g/a_T

k_o = nominal value of k, set equal to 100

M = average number of fixation centers per target area

N_r = number of resolution cells in shortest target dimension

P_1 = probability that an observer looks in the direction of the target with his foveal vision (see p. 3)

P_2 = probability that a target viewed foveally for one glimpse period is detected (see p. 3)

P_3 = probability that a detected target is recognized (see p. 3)

P_R = probability of target recognition

t = time

u = dummy integration variable

α = angular subtense of target or image at the eye

η = overall degradation factor arising from noise in the image viewed by an observer

I. INTRODUCTION

In many operations, success depends on a human observer's finding quickly a certain object in a scene or in some image of that scene. In Air Force operations, armed reconnaissance and many kinds of strike missions depend critically on the timely identification of a target (or its image) by an airborne observer. Whether he is observing directly with his unaided vision, using optical aids, or viewing the display produced by an intervening sensor (e.g., television, radar, or any other imaging transducer), the same capabilities for visual search, discrimination, and recognition are involved. No matter how complex or sophisticated the sensor in front of him or the computer and other mechanisms behind him (e.g., for measuring coordinates or rates or for aiming weapons), the most crucial--and least understood--step in the whole operation is his conscious decision that "There is the target!" The purpose of this Memorandum is to propose a model that describes analytically the performance of a human observer in such a task as a function of a number of well-defined and measurable parameters.

While no subjective act can be analyzed completely, the kind of situation described in the above paragraph is one in which the usually cited sources of variability and unpredictability in human behavior are minimized. By contrast, the task of monitoring an empty scene or display--waiting for something to happen--would be extremely difficult to model because the observer is so quickly subject to boredom and to "wandering" of an otherwise unoccupied mind. But the present case requires active search in a structured field for a known (or briefed) specific object, or perhaps for any of a class of familiar objects, such as trucks on a road. In either case, the task is carried out for a fairly short period of time under conditions of very strong motivation. Under such circumstances, the variability in individual performance and the difficulty in specifying that performance may well be less than the variability between scenes and the difficulty in quantitatively describing the content and the degree of congestion in typical pieces of terrain. In this Memorandum formulas are proposed which, when the inputs available to the observer are completely known, permit an estimation of the probability of his recognizing a target as a function of these inputs.

The proposed analytical expressions, which constitute the model of the observer developed in this Memorandum, can provide valuable assistance not only to designers of display equipment, but also to designers, purchasers, and users of complete systems. However, the limitations of this model should also be recognized. First, it does not attempt to provide a mechanistic analogue of an observer. All that is required of it--and all it provides--is an estimation of recognition probabilities. Second, as has been indicated, the user of the model must provide estimates of the pertinent observable properties of a scene, or its displayed image, as follows: In direct viewing, the size and apparent contrast of the target, the required search rate, and the congestion of the scene are all that are needed; when an intermediate display is used, the displayed target size (scale) and contrast, system resolution, and signal-to-noise ratio (S/N) must also be given. The model describes only the observer, but by so doing provides an essential portion of the analysis that must be employed in evaluating any manned system.

It should not be inferred from the foregoing statements that this is the first or only such model. Many existing models have been utilized in formulating the present one. It differs from others, however, in its conceptual approach at some important points, and it reflects a conscious effort to structure the model according to distinguishable psychophysical processes. It is these conceptual differences, including the selection of pertinent variables, which may justify the presentation of yet another model of the human observer.

II. THE MODEL

The performance of a human observer is often a very complicated function of many interacting variables. In order to simplify this difficult situation and yet stay reasonably close to reality, we consider explicitly the task described in the introduction: the finding of known and fixed objects in a complex field in a short time. This process, even when so restricted, is still complex, but it can be considered to consist of the following three distinct steps: deliberate search over a fairly well-defined area, detection of contrasts (a subconscious retino-neural process), and recognition of shapes outlined by the contrast contours (a conscious decision based on comparison with memory). In addition, when the observer is viewing a displayed image of a scene, noise is usually present which degrades his performance of all three of these steps.

On the basis of assorted experimental data, four formulas can be devised: three for the probabilities of completing each of the three steps separately, and one for a noise degradation factor. It is postulated that the overall target recognition probability can be expressed by the product of these four terms. Accordingly, we establish the following definitions:

1. P_1 is the probability that an observer, searching an area that is known to contain a target, looks for a specified glimpse time (viz., 1/3 sec) in the direction of the target with his foveal vision. P_1 is a function of the ratio of an acceptable search rate to that demanded in a given situation; the loosely defined concept of foveal vision is replaced by that of an effective glimpse aperture.

2. P_2 is the probability that if a target is viewed foveally for one glimpse period it will, in the absence of noise, be detected. P_2 is determined by psychophysical limits operating on the observed or displayed target size and contrast.

3. P_3 is the probability that if a target is detected it will be recognized (again during a single glimpse and in the absence of noise). Recognition is usually (but not necessarily) accomplished on the basis of intrinsic shape without reliance on context.

4. η is an overall degradation factor arising from any noise in the image that is viewed by the observer.

We then write for P_R, the probability of target recognition,

$$P_R = P_1 \times P_2 \times P_3 \times \eta \qquad (1)$$

Inasmuch as (1) the first three steps described above are independent events, (2) P_2 and P_3 (as defined) each represent a conditional probability under the one preceding it, and (3) η is an overall degradation factor, the product formulation of Eq. (1) is obvious and rigorously correct. This is so despite the fact that the individual terms are not strictly independent in the sense that they may be functions of some of the same variables (contrast and S/N, for example). This and certain other subtle interactions are discussed briefly in Section III of this Memorandum. In the following subsections the nature of each of the four terms is examined in some detail, and a specific analytical expression is developed for each one.

THE SEARCH TERM

The first term, P_1, describes the search limitations; the primary concern is structured search. By contrast, in free search, large objects (such as clearings in woods) or objects with outstanding contrast are usually spotted first by peripheral vision and are then examined more carefully. In such cases, a "visual lobe" theory[1] of detection is appropriate in which successive looks in random directions are postulated and off-axis detections are significant. Indeed, such a model[2] was used effectively in the analysis of some classified visual reconnaissance tests[3] in which most of the targets were highly visible once found. That is not the kind of situation treated here,* nor are moving targets to be considered. Motion cues are recognized to be quite important--in fact, often overriding--but are not included.

*Visual lobe theory was developed for completely unstructured search, such as horizon search at sea or search of the sky in daylight; its application to search of terrain, even under the conditions mentioned, is therefore somewhat suspect.

When searching from the air for a terrestrial object whose location is known only approximately, it becomes both possible and necessary to utilize foveal vision and to search fairly systematically. The maximum acuity of foveal vision is a necessity, since there is always a need to find targets at the earliest possible moment during approach, and at long range either the apparent size or the available contrast or both may be marginal; few military targets really stand out. Foveal vision is also usually feasible, since only a limited area needs to be covered. The required area may be as much as the whole of an electro-optical display, but more commonly it is an area set by navigation errors and target location uncertainties, centered on a predicted or expected target location. Even under these conditions, however, search rates are extremely variable and almost intractable for the fundamental reason that pieces of terrain (not to mention possible targets) differ widely and almost defy quantification. Nevertheless, some bounds can be set.

It is well known that the eye moves in discrete steps, ordinarily with about three stops, called fixations, per second.(4) (Actually an observer occasionally takes longer to examine certain points, but this does not affect very much the average search rates described below.) Our approach, therefore, is to postulate that an experienced observer searches by moving an apparent aperture (essentially his foveal vision) in some fairly regular pattern over the area of interest, and furthermore that he adjusts his average interfixation distance, and hence the effective size of his scanning aperture and his overall search rate, in accordance with his a priori information on the size and contrast of the target or its image. Intuitively, one recognizes that an observer will scan the floor around him differently if he is looking for a pencil or an ant. Stated more formally, the observer estimates how far off his visual axis he will still have an adequate probability of detecting the expected image, and he automatically adjusts his search rate accordingly. A key concept, therefore, is the size of the effective scanning aperture--here called a glimpse aperture, A_g. This is a quantity that commonly ranges from 10 to 100 times the area of the target, a_T, but can sometimes vary between 1 and 1000 times a_T.

The reason for this huge spread is not just the observer's inability to predict the nature of the image or his own detection probabilities. It lies in a second important factor--the structure, complexity, or "congestion" of the surrounding scene. The search for an ant mentioned above will also be quite different depending on whether the floor is covered with a nearly featureless lineleum or a textured and patterned rug. However, this "congestion" cannot be described solely by the two-dimensional spatial-frequency content in a scene. What really matters is the density of contrast points--the natural fixation centers for the eye--or other "confusion objects" that are present in the scene. The writer once experienced a striking example of many such false targets (natural decoys, as it were) while flying over the notorieus Coso Range in California. This region contains scattered trees and bushes which appear very dark against the background of sandy soil or dried grass, as do the vehicles and "bridges" which were placed in the area as "targets." Almost every tree had to be examined to see whether or not it had straight sides before the true targets could be found. Indeed, tests there have produced some of the lowest target acquisition probabilities ever measured.[5]

The kind of adaptive search rate described here, in which the observer automatically reacts to both the character of the scene and the (anticipated) nature of the target imbedded in that scene, has been advocated informally by this writer for several years. The only independent reference to such a concept found in the literature is by Williams.[6] He talks about target "conspicuity," which is measured by the rate at which a particular target can be successfully searched for in a particular field, and he points out that the commonly observed lack of dependence of target acquisition on display scale factor (within limits, and assuming no change in information content on the display) is another manifestation of observer adaptation. Other experimenters, of whom Richardson[7] is an important example, recognize the strong dependence of search performance on "target class."

A heuristic derivation of an expression for P_1 follows, along with an indication of the supporting experimental evidence. If an area A_s is to be searched, the number of glimpses (each of area $A_g = ka_T$)

required to cover the area is A_s/A_g. The number of glimpses that are available in t sec, at 1/3 sec per glimpse, is 3t. With perfectly systematic search, the probability of "looking at" the target (i.e., including it within a glimpse aperture) would be just the ratio of the available glimpses to the total number required, or $3t/(A_s/A_g)$. This would give P_1 the form of a linear ramp function with time. Real search is probably something between perfectly systematic and purely random, so that P_1 should have a form that lies between the ramp and an exponential rise. We conservatively adopt the latter and postulate

$$P_1 = 1 - e^{-K \times 3t/(A_s/ka_T)}$$

where K is a constant and k is a parameter related to scene congestion.*

The exponential form proposed for the dependence of P_1 on t was predicted by Williams(6) and was found by Boynton and Bush.(8) The dependence on kt (or t/M) found, though quoted somewhat differently, by Boynton et al.(9) and by Nygaard, Slocum et al.,(10) and still differently by Stathacopoulos et al.,(11) can be closely approximated by the identical exponential function. The evaluation of the coefficient K is accomplished as follows. The previous equation can be interpreted in terms of search rates as well as total numbers of glimpses. In that case the exponent is merely K times the ratio of an "acceptable" or successful search rate, ka_T per 1/3 sec, to the required rate A_s/t. If "acceptable" is defined as yielding a value of 0.9 for P_1, then k must be selected from measured data for which $P_1 = 0.9$ and at the same time K must be set so that, when the real rate is equal to this acceptable rate, $P_1 = 1 - e^{-K} = 0.9$. Therefore $K = \log_e 10 = 2.3$. (If some other definition were adopted for "acceptable," K and k would

*Alternatively and by completely parallel reasoning, in a scene (like the Coso Range mentioned above) in which the average density of confusing objects or fixation centers is M per target area, the number of glimpses required to cover the area A_s is given by MA_s/a_T. This leads to an identical expression for P_1 if M is set equal to 1/k. The effect of extra fixation points is therefore to increase the number of glimpses per unit area (or to decrease the average interfixation jump distance) and hence to reduce the effective glimpse aperture and the areal search rate that can be achieved.

change reciprocally, maintaining a constant product.) The search rates measured in Boynton's experiments,[8,9] when normalized to the number of target areas per glimpse time at a 0.9 probability of success,[*] yield a value of $k \approx 200$. Simon's data,[12] on the other hand, dealing with real imagery of very congested scenes (e.g., metropolitan Los Angeles) yield a value in the neighborhood of $k = 10$. As might be expected, this kind of spread in observed search rates is not uncommon. Bennett's data[13] lead to an average value of 135 for k, while this writer in an old unpublished experiment found $k = 40$. Since we think that Boynton's artificial scenes may be unrealistically low in clutter, we conclude, from the foregoing and a wide range of similar data, that values of k for real scenes typically fall between 10 and 100, but that values well outside that range are also possible. For convenience, we write k_o/G for k, where k_o is a nominal value of k for which we adopt the figure 100, and G is a "congestion factor" equal to unity in the nominal case but taking on various positive values, usually between 1 and 10, for other scenes. Accordingly, we propose the following expression for P_1:

$$P_1 = 1 - e^{-[(700/G)(a_T/A_s)t]} \qquad (2)$$

Since by definition $G = k_o/k = 100/k$ ($\approx$ 100M), it can be visualized as the average number of fixation centers per nominal glimpse aperture of 100 a_T, and this indeed constitutes a valid physical definition for G. In practice, however, it may be little more than a measure of relative congestion. Values of G less than 1 are possible, as has already been implied, but these should be invoked by the user sparingly and only for relatively open scenes--those naturally containing regions of uniform brightness (e.g., lakes or empty fields) that can be jumped over quickly, or artificially so by virtue of moving-target indicating (MTI) radar or multispectral cueing. Values greater than 10 are also

[*] The experiments cited in this paragraph were all essentially search-limited; i.e., the targets were easily recognized once they were actually looked at (fixated upon). In the terms of the present model, the conditional probabilities P_2 and P_3 were high, approaching unity. Hence "successful search" can be translated into "looking in the direction of the target" as required for P_1.

possible, but they are not very common either, except in the sense that the effective search rate would be quite low whenever significant decision times are invelved, as when examining truly confusing objects or decoys.

It may be noted that the exponent in Eq. (2) is simply seven times the reciprocal of the rate at which fixation points must be examined in order to cover the search area in the time allowed. This interpretation is theoretically sound and is intriguing in its simplicity. However, it is probably not very helpful in practice (at least with the present state of our knowledge) because of the difficulties in predicting which points in a scene will prove to be fixation centers. By providing the user with a nominal glimpse aperture (and search rate), Eq. (2) demands of him only that he estimate deviations from that nominal--by selecting for G a number that, in most cases, lies between 1 and 10.

Speaking realistically, even an experienced observer who can judge the relative congestion of a given scene with respect to others may have difficulty in estimating the value of G better than to within about a factor of two, but this is still much better than having no bounds whatsoever. In fact, it permits one to draw such general but important conclusions as these: Broad area search from high-speed aircraft is rather futile, while road recce or other one-dimensional search may, on the other hand, be quite feasible up to speeds of a few hundred knots.*

THE CONTRAST TERM

The second term, P_2, has to do with the basic process of contrast detection by the human visual system. Blackwell's[14] classical experiments provide the fundamental data here, yielding curves of threshold contrast (50-percent detection probability) versus size of circular discs under various levels of ambient illumination. These are commonly called "demand" contrast functions. However, there is a good deal of

*Consider, for example, linear search at 10 truck lengths/glimpse, which corresponds to 600 ft/sec or 350 kn permissible speed; however, by the same argument, a two-dimensional search for a tank over a swath width of as little as 1000 ft would be limited to a speed of 70 kn.

evidence that the best (i.e., lowest) threshold values obtained by Blackwell must be adjusted upward substantially for application to the practical situations discussed in this Memorandum. A good critique of the pertinent experiments on this subject is that by Davies.[15] Following him in part, we take an average of the data for exposures of 1/3 sec obtained by Blackwell and McCready[16] and by Taylor[17] as the most relevant starting point (consistent with the search model developed in the previous section), assume photopic vision with 30 to 100 fL average scene brightness,* and then apply the following corrections. A factor of 2.4 in contrast is suggested by Blackwell[18] for the difference between free-choice situations and the more easily controlled but less realistic forced-choice experiments, while he also suggests a factor of about 1.5 to allow for uncertainties in position or time of target appearance. Similarly, Vos et al.[19] found that an

*An aside on the effects at other light levels may be of some interest at this point. The luminance level chosen above is intended to cover ordinary daylight seeing and also the (photopic) viewing of bright electro-optical displays. With more light, the curve of Fig. 1 on p. 12 shifts downward and to the left, but only slightly. As the available light decreases, however, the curve moves sharply to the right and up by a factor that is roughly the square root of the factor by which the luminance changes. This performance "loss" can be recovered by electronic gain, as in image intensifiers, up to the point that the electronic gain is merely amplifying "empty" photon noise. At this point the performance is limited, not by the eye, but by the information contained in the arriving photon stream. This new limit is somewhat different in shape, being approximately hyperbolic in resolution and contrast (linear on Fig. 1, with a slope of -1), and of course it depends on the luminance level and on several properties of the intensifier hardware. For example, following Richards (Ref. 20) in a slight refinement over the original Rose formula (Ref. 21),

$$\alpha C \geq 3440 \frac{2k}{D} \sqrt{\frac{(2 - C)e}{\tau S t B}}$$

in which k is the effective S/N (~ 5 (see p. 16)), D^2 is the area of the collecting aperture, e is the electronic charge, τ is the transmission of the optics, S is the photocathode sensitivity (A/lm), t is the integration time, and B is the scene luminance (lm/sr/unit area) of the brighter of two patches just resolved at apparent contrast C.

overall factor varying between 2.5 and 3.5 in contrast is required to reconcile certain of Blackwell's data with theirs taken under somewhat more realistic conditions (but still in a laboratory using uniform backgrounds). In a flight environment there are still further degradations, primarily two: the direct blurring effects of vibration and the inability of an observer to accommodate simultaneously to all intensity levels when viewing a real scene that probably contains at least 20 dB of dynamic range. Davies argues, rather vaguely, that another 60-percent degradation (a factor of 1.6) is little enough to allow for these and other effects, and we agree. It is proposed, therefore, that the shape of the "demand" curve of threshold contrast C_T versus angular subtense α in minutes of arc (min) be taken from the average of the two best-known sources of 1/3-sec data, and that this curve be adjusted upward by a factor of about 5.5 in contrast—or that 0.75 be added to log contrast. The resulting curve is plotted as a dashed line in Fig. 1.

In addition to the evidence that has been cited by Taylor[22] for various "field factors" of the sort just described, there are some meager flight test data by Heap,[23] reported more fully by Davies,[15] the results of which are plotted in Fig. 1. It may also be observed that clinical optometrists use gray scale prints consisting of 20 1-dB steps and assert[24] that this is all that can be seen in a "mixed field." This is not exactly "hard data," but simply corroborative evidence from another field concerning the coarseness of contrast discrimination in practical situations.

Since, in the absence of bright lights or specular glint, target contrasts* greater than unity are rarely observed through the real atmosphere,[25] and even less frequently on military targets, the dashed curve in Fig. 1 can be approximated by the hyperbola

$$(\log C_T + 2)(\log \alpha + 0.5) = 1 \qquad (3)$$

*Contrast is defined here as the absolute value of the difference between target and background luminances divided by the background luminance.

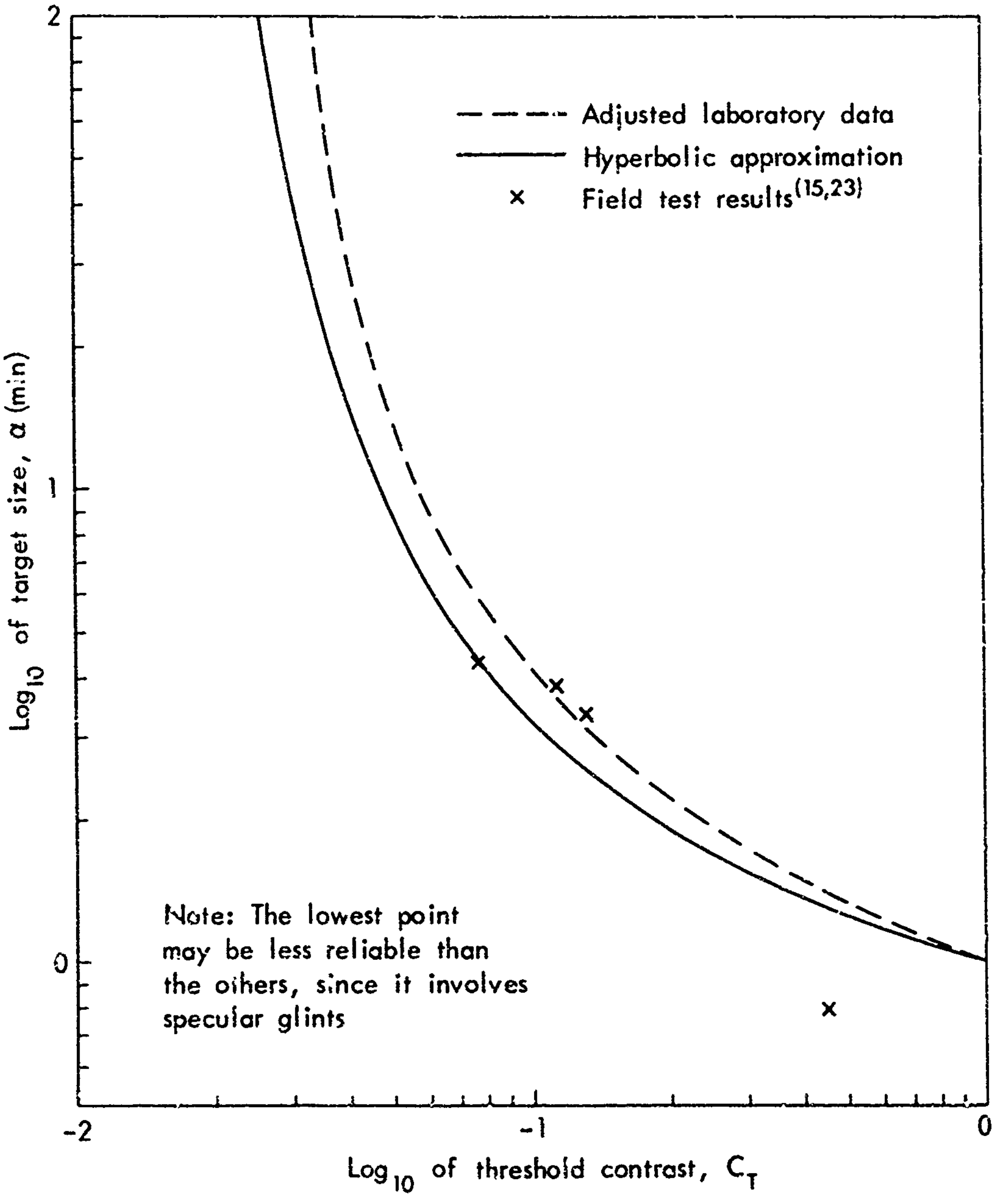

Fig. 1—Threshold performance of the eye: the "demand" contrast function

which is shown by the solid curve in Fig. 1. This simplification is often convenient and usually adequate, but whenever contrasts greater than unity are important (for example, on certain electro-optical displays), a more accurate curve with an asymptotic slope of -1/2 should be used.

It is obvious from Fig. 1 that a better fit could be obtained between the two curves. Very simply, if the hyperbola were shifted 0.1 to the right, by changing the 2 to 1.9 in Eq. (2), an excellent though still not optimum agreement with the "true" curve would result. But, in accordance with the old-fashioned concept of significant figures, one should not imply a precision of results that is not justified. In view of the way the dashed curve was derived, it may be no more accurate than 20 or 30 percent--so it really should be drawn with an air brush. Accordingly, with the present state of our knowledge, no greater accuracy should be inferred for Eq. (2) than is indicated.

The probability of detection, P_2, at the threshold contrast is, by definition, 50 percent. The probability of detection for other values of observed contrast, C, has been shown by Blackwell and McCready[16] to depend only on the ratio C/C_T and to have the form of the cumulative normal distribution with $P_2 = 0.9$ for $C/C_T = 1.5$. This is equivalent to setting the value of the Gaussian standard deviation equal to 0.39, and it indicates that on the average Blackwell's subjects chose to operate at a false-alarm rate of about 1/200, corresponding to an S/N of roughly 2.6:1. Further support for the general form of the dependence, based on statistical decision theory, is provided by Ory.[26] Accordingly, we write

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\left\{[(C/C_T)-1]/0.39\right\}} e^{-u^2/2} \, du \qquad (4a)$$

A useful approximation that is more suitable for machine computation is the following:

$$P_2 = \frac{1}{2} \pm \frac{1}{2} \left\{ 1 - e^{-4.2[(C/C_T) - 1]^2} \right\}^{1/2} \qquad (4b)$$

where the minus sign is used when $C < C_T$. C is the actual contrast available at the eye after atmospheric effects(25) and (when pertinent) equipment gains and display settings are accounted for; C_T is computed from Eq. (3) with α giving the average angular subtense, in minutes of arc at the eye, of an object or its displayed image.

Line-of-sight masking by terrain or foliage is really outside the purview of this model; however, since it has the effect of reducing the observed target area, it can be thought of as reducing P_2. Similarly, camouflage may reduce the observable contrast to some very low value or may alter the apparent shape of an object. The subject of shape recognition is discussed next.

THE RESOLUTION TERM

The third term, P_3, has to do with the more subjective act of deciding what particular image forms represent in the real world. But since we are primarily concerned with shape recognition of known or briefed objects, as distinct from the interpretation of unfamiliar imagery, the problem can be reduced to the visibility--or detectability in the sense of the previous subsection--of sufficient geometrical detail for shapes to be compared with memory and thereby recognized. The concept of "sufficient" detail might lead one into the morass of "critical details"--those unique features that permit various classes of objects to be distinguished one from another. However, when all portions of an image are equally detectable so that the whole shape is either visible or not, Johnson(27) has demonstrated the remarkable fact that, for a variety of military objects,* a single parameter--namely N_r, the number of resolution cells contained in the shortest dimension across a target--is all that is required to describe what constitutes "sufficient" detail for detection or for recognition. He found values of N_r between 3.3 and 4.8, or 4.0 ±20 percent, for high-

*One should probably add "in a military context." The amount of detail required to distinguish a truck from an oxcart is far less than that required to discriminate between various truck models; but the simple separation of objects into classes is usually sufficient for designating targets.

confidence recognition. This important simplification has been further confirmed by Brainerd et al.(28) for several target shapes, and these authors also provide enough data points to support a simple Gaussian form for the dependence of P_3 on the parameter N_r. Oatman(29) has performed similar experiments which are only slightly more pessimistic (N_r larger by 25 percent) than the first two, provided that his numerical results are corrected for the deterioration of his TV display away from the center of the tube face. We adopt a conservative value, close to Oatman's, and write

$$\begin{aligned} P_3 &= 1 - e^{-[(N_r/2)-1]^2} & N_r &\leq 2 \\ &= 0 & N_r &< 2 \end{aligned} \qquad (5)$$

which makes $P_3 \approx 0.9$ when $N_r = 5$.

It is important to emphasize the meaning of N_r. As previously defined, it is the number of resolution cells contained in the minimum dimension (e.g., width or height) of the projected image of an object to be recognized. In the present context, "resolution cells" means independently detectable spots--the subject of the previous discussion. Pure resolution (in the original sense of separating two spots), though related, is not directly involved here, nor is resolution as determined from a bar chart the appropriate measure to be used in calculating N_r. The proper procedure is to calculate first, from Eq. (3) or Fig. 1, the size of the smallest spot that can be seen--at the contrast level with which the target is presented to the observer. Next, for reasons discussed in the following paragraph, this spot size should be corrected for a 90-percent probability of detection, rather than using the threshold (50-percent) value. Finally, the number of these 90-percent detectable spots contained in the shortest dimension of the target image then gives the value of N_r. This procedure is illustrated graphically in Fig. 2 (page 19) and is described more fully starting on page 18.

The choice of 90 percent as the level of detection probability to be used in determining the effective resolution in any specific situation is to some extent arbitrary. However, if it were a^ low as, for example, 50 percent, :learly only half of the spots would be visible at any instant. This would violate the condition (stated on page 14) that the "whole shape" should be either visible or not. Although exact results would depend on the properties of any noise that might be present and on the ability of the observer to integrate out these effects, it can be expected (in the absence of detailed experimental evidence) that the individual spot-detection probabilities should be, say, 85 percent or greater in order for the cited measurements to apply and for the simple form of Eq. (5) to be valid. The variations permitted within the remaining uncertainty fall well within the overall accuracy limits claimed for this model.

The need for distinguishing carefully between the various possible measures of resolution, as was done in the preceding paragraphs, arises from the fact that bar-chart resolution, particularly when observed with converging bars as in the common TV test patterns, is quite different from--and significantly more optimistic than--the resolution determined from random spot detection. As pointed out explicitly by Rosell,[30] the difference lies in the ability of the human visual system to integrate over a completely known and heavily ·dundant bar pattern--to accept gaps in a bar or even whole missing bars--and so to effectively operate at a much lower S/N ratio than is possible when almost every "corner" or other detail of an arbitrary shape must be detected independently in order for the shape to be correctly observed. The difference seems to be a factor of about 4 or 5. This number can be derived from a direct comparison of the value of S/N = 1.2 quoted by Parton and Moody,[31] loosely based on their bar-chart measurements, with the classical work of Rose[21] on spot detection; or it can be simply estimated, as was apparently done by a group of RCA engineers,[32] from the fact that overall S/N is proportional to the square root of the area observed and from the finding of Coltman and Anderson[33] that the eye uses efficiently the area of about 5 bars or line pairs.

An interesting confirmation of both the concept that has been described and the numerical value that has been adopted in Eq. (4) can be derived from the work of Steedman and Baker.(34) In their experiment, the resolution is clearly defined by the cell size of their computer-generated patterns, increased by some fraction (which, within limits, does not make much difference) of the blur circles that are artificially added. If their data (see their Table I) are examined in detail, it is observed that those targets with the small angular subtense--which require the longest search times and induce the most errors--are also those that consist of a small number of elements or cells. Furthermore, at their well-known "cutoff" size of 12 min of arc,* the average number N_r of resolution cells (with due allowance for the blur circles) is between 5 and 6. Above this cutoff, they found an almost constant search time for a given shape, and an error rate of 2 to 4 percent; below this value, they found a marked increase in both quantities. Correspondingly, Eq. (4) predicts a $P_3 = 0.95$ for this value of N_r, which drops rapidly to about 0.3 for half that value of N_r and to zero for $N_r = 2$. The latter corresponds to Johnson's(27) criterion for detection only, with no shape recognition per se.

A special case is that of long, narrow objects which, in the limit, reduce to lines. These are a great deal easier to recognize, primarily because of the same redundancy effect mentioned above. This effect in one dimension, combined with moderate (not threshold) levels of contrast, gives rise to the commonly observed value of $N_r \approx 0.2$ for this case.

In the process of applying the foregoing model of an observer to a practical situation involving an artificial electronic or electro-optical sensor, it would be helpful to construct a diagram similar to

*They actually use the longest target dimension, which subtends 12 min of arc under ideal conditions, and they suggest that 20 min of arc might be a more practical value. We interpret the 12 min of arc as the subtense across the minimum dimension of the target under realistic conditions, which is probably reasonable for most commonly shaped objects for which the "aspect ratio" is less than 2:1.

Fig. 2. A description of this diagram will serve as a good summary of the model as it has been described up to this point. First, one must calculate the displayed contrast, C, for a hypothetical target with intrinsic (zero-range) contrast, C_o, with respect to its contiguous background. This calculation involves power levels, receiver or detector sensitivity, atmospheric attenuation and path luminance effects (if any), the transfer characteristic of the system for the particular "gain" and "contrast" settings chosen by the operator, and the modulation transfer function (MTF), i.e., the system response as a function of spatial frequency or reciprocal target size. The result is an overall transfer function plotted on a graph of contrast versus target image subtense at the eye, α. Typical curves for various possible measured (or postulated) values of C_o are plotted as thin solid lines in Fig. 2.* Next, one computes the actual target (image) average subtense at the eye, say α′, and enters Fig. 2 at this abscissa. Reading the appropriate contrast curve, one finds the value, C′, with which that target will be presented to the observer. C'_T, the threshold contrast for an object of apparent size α′, is obtained from Eq. (3), and the ratio C'/C'_T permits calculation of P_2 through Eq. (4a) or (4b). Equation (3) can also be plotted on Fig. 2 for all values of α; this demand contrast is shown as the heavy solid curve. The stippled area covers the band of $0.5 \leq (C/C_T) \leq 1.5$, which, by Eq. (4a), represents the region for which $0.1 \leq P_2 \leq 0.9$. This can be used for finding P_3 in the following manner. If the appropriate displayed-contrast curve is followed to its intersection with the stippled area, the abscissa of that intersection (say, α″) will represent the useful resolution that can be achieved on the subject display (with targets of inherent contrast C_o). The ratio α′/α″ (corrected, if necessary, for target aspect ratio) is N_r, the parameter which, when inserted in Eq. (5), yields the value of P_3.

It was implied at the beginning of this section that recognition in unfamiliar situations may be much more complicated, and far more

*For unaided vision, only the atmospheric reduction in contrast need be computed, and the left-hand intercepts determined accordingly; the transfer "functions" will then be horizontal straight lines on Fig. 2 out to the point where shimmer sets in.

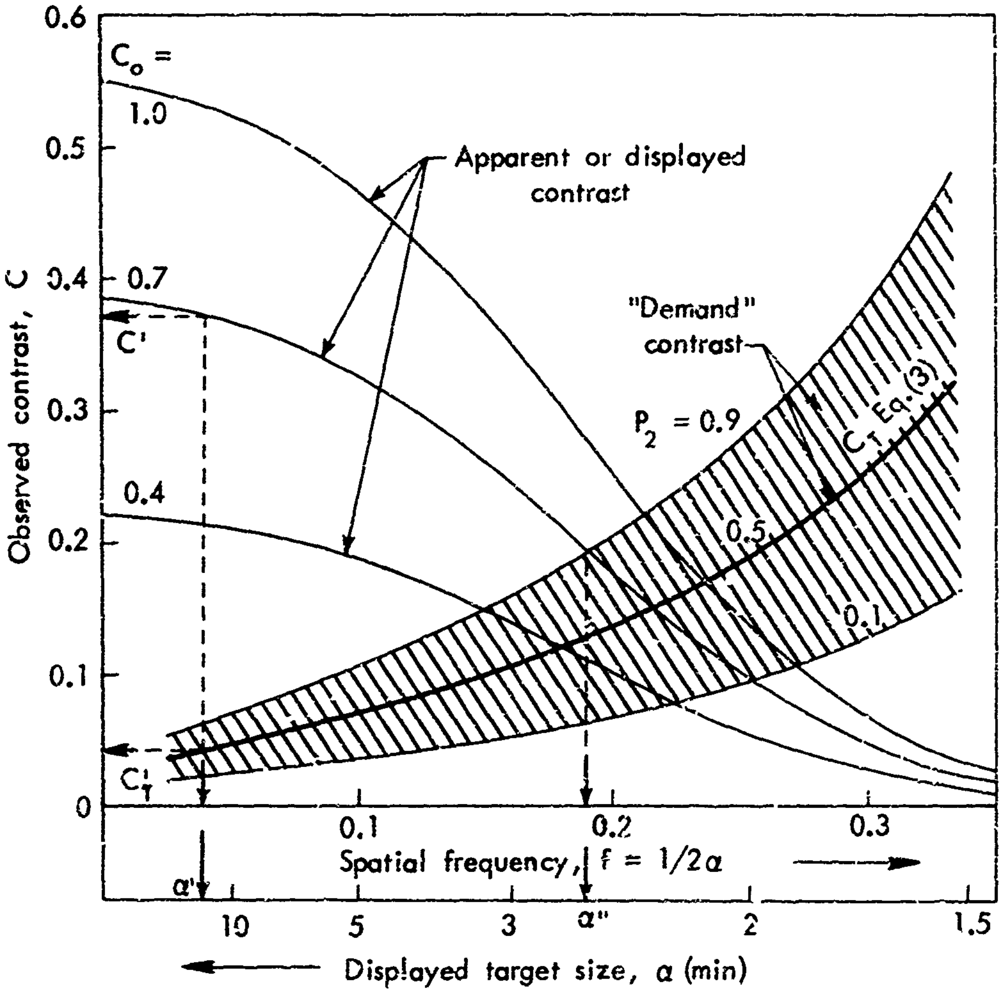

Fig. 2—Schematic representation of displayed and "demand" contrast versus displayed target size

difficult to predict, than the mere detection of shape details. An extreme example might be the classical one of the photo-interpreters searching for completely unknown elements of the Peenemunde launching areas during World War II. No attempt is made to extend this model to cover such cases. It should also be mentioned, however, that under certain other circumstances recognition may be very much easier than this model would predict. Consider the approach of unauthorized aircraft, or the presence of vehicles along a road in enemy territory. Both are cases in which the mere detection of objects might be sufficient to justify the decision, "There is a target!" These cases can be handled by assigning artificially high values to P_3 (when the prior information so justifies), thus effectively equating detection as given by P_2 to recognition. This point is discussed further in Section III. Our model of P_3 covers the more common intermediate cases in which shape provides the primary criterion for recognition.

THE NOISE TERM

The last term of our model, η, describes the ability of an observer to integrate out those unwanted fluctuations usually referred to as noise. More accurately, it describes the difficulty of reading through any noise that may be present in the image being viewed. This includes both equipment-generated noise and real but unpredictable fluctuations in the scene itself. Amplifier noise, TV beam effects, and photographic grain are examples of the former; amplified photon noise and the graininess of coherent imagery (laser or synthetic-aperture radar) are examples of the latter. True photon noise is not pertinent at this point, since the model in its present form applies only to photopic vision--observing daylight scenes or bright displays--which is apparently processor-limited and thus sensitive to contrast rather than being noise-limited.(26)

Image noise, whatever its source, affects the recognition processes in many ways. First, it increases the apparent congestion of a scene, G, and thus reduces P_1. Second, it increases the threshold contrast, C_T, required for spot detection, and so reduces P_2. Third, by distorting contrast boundaries and generally lowering gradients or acutance, it increases the required value of N_r (essentially, the denominator in the exponent of Eq. (5)) and so reduces P_3 as well. Rather than specifying each of these effects in detail, we take the expedient course of proposing a single overall degradation factor, η. Whenever image noise exists, this factor is to be applied to the recognition probabilities estimated by the $P_1P_2P_3$ product.

Most of the work, both analytical and experimental, on the effects of noise on image interpretation is concerned only with threshold conditions for which the probability of detection is 0.5. Data on the effect of other than threshold values of noise on the probability of detection are not easy to come by, but there are a few. Coltman and Anderson(33) show that the S/N per unit area that is tolerable for detection of an image is inversely proportional to the linear dimension of the image. Since total S/N can thus be traded directly for image size, one can conclude that the dependence of detection probability on S/N should have the same form as that of image size, namely the form of Eq. (5). However, in view of the paucity of good empirical data on this point, the author prefers a slightly more conservative formulation (predicting lower probabilities at modest values of S/N), which can be had by reducing the exponent in the exponential from 2 to 1. In addition, a few measurements by Schade,(35) as replotted by Stathacopoulos et al.,(11) do fit very well on the resulting curve. We therefore adopt the form

$$\begin{aligned} \eta &= 1 - e^{-[(S/N)-1]} & S/N &\geq 1 \\ &= 0 & S/N &< 1 \end{aligned} \tag{6}$$

In direct vision (at high light levels), with no equipment or image noise to be accounted for, this S/N is infinite and $\eta = 1$.

III. DISCUSSION

Equations (2) to (6), combined as indicated by Eq. (1), constitute the proposed model of a human observer. The fundamental concepts and the basic product formulation are explained at the beginning of this Memorandum. The result, after analyzing each of the four terms, is an expression for the probability of recognition as a function of several observable quantities--the apparent size and contrast of the target, the required search rate and the false-target density in the scene, and the resolution and noise appearing on any intervening display.

An important and useful property of the model is the separation of variables that has been achieved. Each of the terms is expressed as a function of a rather small number of input parameters, and target size is the only parameter which appears in more than one term. This rather significant simplification arises from a careful consideration of the consequences of the product formulation and a detailed evaluation of each of the terms over only the ranges of the input variables for which that term is controlling or otherwise of interest.

For example, the model is not applicable to a target that is so isolated or whose contrast is so high (relative to the background clutter) that it can easily be seen with peripheral vision,* since in that case the search rate can be very much faster than postulated in Eq. (2) and P_1 will be very high. But then P_2 and η will also be very high (essentially unity), and the problem is almost trivial. The search model assumes only that target contrast is *not* that high, so that fairly systematic and fine-grained search must be carried out. In fact, the actual search rate employed by an observer is determined by some sort of average false-target density over the scene. If the actual contrast of a specific target against its contiguous background turns out to be less than sufficient for recognition to take place during a single properly directed glimpse, this fact will show up in P_2 and P_3, which will correctly reduce the value of P_R.

*This is essentially what is achieved by multispectral cueing or by MTI radar, as indicated on p. 8.

The relationship between the conditional probabilities, P_2 and P_3, can be discussed in a similar way. While it is clear that they are intimately related, they are separated, with further separation of variables, for several reasons. Ordinarily, detection not only precedes but also dominates shape recognition. That is, unless P_2 is rather high, there is probably no point in even calculating P_3 or η, since P_R will be too low to justify the sortie. When P_2 is high, then P_3 controls. On the other hand, as has been mentioned, there are cases for which a priori or contextual information may suffice to obviate the need for shape recognition per se. In such cases--boats on a river or trucks on a road, for example--P_3 can be ignored (i.e., set to unity without regard for Eq. (5)) and P_2 will control. By keeping the two terms separate, model flexibility is preserved. Further arguments for this separation revolve around the role of resolution. First, as a practical matter, most man-made sensor systems are resolution-limited, since resolution always costs something. (This is true at least of systems whose displays are properly designed.) Accordingly the sometimes-difficult calculation of system MTF need be applied only once (namely, when it is most critical) in the shape-recognition term. More importantly, there are many cases with multiscaled or zoom-capable systems in which the combination of a priori information and required search area may make a two-step identification of the target desirable. In such cases an initial and tentative detection on a wide field of view is followed and confirmed (or denied) by shape recognition on a magnified image. At the first step P_2 controls, but P_R is incomplete; at the second step P_3 controls.

Finally, η affects P_1, P_2, and P_3 as has been mentioned, but it is kept separate merely for convenience. In fact, all four terms, as they are defined, are not only functions of different variables, but are also subject to different kinds of uncertainties and will require different experiments for their future refinement. Yet the product of the four provides a viable model for a wide variety of circumstances; it can be used in predicting the capabilities of a broad class of manned systems, since it deals only with the observer and the information presented to him, whether this be directly to his unaided eyes or through optical aids or sophisticated artificial sensors.

It has been emphasized, nevertheless, that the applicability of this model is restricted to structured search (as in air-to-ground applications) for fixed objects whose appearance is at least approximately known (as in acquiring pre-briefed targets) under conditions of time-urgency (e.g., prior to weapon delivery). In quite different contexts, such as monitoring static situations, unstructured search (as in looking for aircraft against a completely homogeneous sky background), and examination of unfamiliar imagery by photo-interpreters, this model will be quite inadequate. Also, the accuracy of these predictions, or the lack thereof, should be kept firmly in mind. As judged from the degree of consistency of the available experimental data, it has been indicated that most of the terms of the model are correct to within some 20 to 30 percent (1 σ, measured at the inputs--contrast, number of resolution cells, etc.) and that the search rates may well be in error by a factor of two or so in either direction. Hence the real utility of the model is in setting bounds on what should be expected of observers in "real-time" situations.

No overall "validation" of this model, in the sense of completely controlled field tests, is known to exist. Of course, the several pieces of the model are based on experimental evidence, including such flight tests as are pertinent, but better operational data are badly needed. Field trials, carefully designed with some sort of predicting model in mind, and with *all* the pertinent parameters recorded, are a necessity. If such programs could be funded, it could be hoped that eventually there might emerge a quantitative understanding of observer performance along the lines of Ory's[26] treatment of threshold visual performance. At present, however, this appears to be no more than a distant gleam.

The difficulties encountered in attempting to predict recognition probabilities are manifest and well known. Nevertheless, this simplified model of the observer, when properly combined with data on targets, backgrounds, the atmosphere, and the performance of specific sensors, is believed to be capable of setting bounds on feasibility that are practically useful. When applied reiteratively to successive system designs, the model serves to define--albeit loosely at present--the requirements placed on any system which is to be operated by a human observer.

REFERENCES

1. Koopman, B. O., *Search and Screening*, U.S. Navy, Chief of Naval Operations, Operations Evaluation Group, Report 56, 1946.

2. Bradford, W. H., *A Mathematical Model for Determining the Probability of Visual Acquisition of Ground Targets by Observers in Low-Level High-Speed Aircraft*, Sandia Laboratory, Report TM-66-54, February 1966.

3. Classified reference, available to official users.

4. Ford, A. C., C. T. White, and M. Liechenstein, "Analysis of Eye Movements During Free Search," *J. Opt. Soc. Am.*, Vol. 49, March 1959, p. 287.

5. Classified reference, available to official users.

6. Williams, L. G., "Target Conspicuity and Visual Search," *Human Factors*, Vol. 8, February 1966, p. 80.

7. Classified reference, available to official users.

8. Boynton, R. M., and W. R. Bush, *Laboratory Studies Pertaining to Visual Air Reconnaissance*, Parts I and II, Wright Air Development Center, Report TR-55-304, September 1955, April 1957.

9. Boynton, R. M., C. Ellsworth, and R. M. Palmer, *Laboratory Studies Pertaining to Visual Air Reconnaissance*, Part III, Wright Air Development Center, Report TR-55-304, April 1958.

10. Nygaard, J. E., G. K. Slocum, et al., *The Measurement of Stimulus Complexity in High Resolution Sensor Imagery*, AFSC, Aerospace Medical Division, Report AMRL-TDR-64-29, May 1964.

11. Classified reference, available to official users.

12. Simon, C. W., "Rapid Acquisition of Radar Targets from Moving and Static Display," *Human Factors*, Vol. 7, June 1965, p. 185.

13. Bennett, C. A., S. H. Winterstein, and R. E. Kent, "Image Quality and Target Recognition," *Human Factors*, Vol. 9, February 1967, p. 5.

14. Blackwell, H. R., "Contrast Thresholds of the Human Eye," *J. Opt. Soc. Am.*, Vol. 36, 1946, p. 624.

15. Davies, E. B., *Contrast Thresholds for Air to Ground Vision*, Royal Aircraft Establishment, Report TR 65089, 1965.

16. Blackwell, H. R., and D. W. McCready, Jr., "Foveal Detection Thresholds for Various Durations of Target Presentation," *Proc. NAS-NRC Vision Committee*, November 1952.

17. Taylor, J. H., *Visual Contrast Thresholds for Large Targets*, Parts I and II, Scripps Institute of Oceanography Visibility Laboratory, Reports SIO 61-25 and 61-31, 1961.

18. Blackwell, H. R., "Recent Laboratory Studies of Visual Detection," *Proc. NAS-Armed Forces-NRC Vision Committee*, April 1953.

19. Vos, J. J., A. Lazet, and M. A. Bouman, "Visual Contrast Thresholds in Practical Problems," *J. Opt. Soc. Am.*, Vol. 46, 1956, p. 1065.

20. Richards, E. A., "Fundamental Limitations in the Low-Light-Level Performance of Direct-View Image-Intensifier Systems," *Infrared Phys.*, Vol. 8, 1968, p. 101.

21. Rose, A., "The Sensitivity Performance of the Human Eye on an Absolute Scale," *J. Opt. Soc. Am.*, Vol. 38, 1948, p. 196.

22. Taylor, J. H., "Use of Visual Performance Data in Visibility Prediction," *Appl. Opt.*, Vol. 3, 1964, p. 562.

23. Heap, E., "Visual Factors in Aircraft Navigation," *J. Inst. Navigation* (London), Vol 18, 1965, p. 257

24. Sheppard, J. J., Jr., R. H. Stratton, and C. Gazley, Jr., *Pseudocolor as a Means of Image Enhancement*, The Rand Corporation, P-3988, January 1969.

25. Bailey, H. H., and L. G. Mundie, *The Effects of Atmospheric Scattering and Absorption on the Performance of Optical Sensors*, The Rand Corporation, RM-5938-PR, March 1969.

26. Ory, H. A., *Statistical Detection Theory of Threshold Visual Performance*, The Rand Corporation, RM-5992-PR, August 1969.

27. Johnson, J., "Analysis of Image Forming Systems," *Proc. Image Intensifier Symposium*, U.S. Army Engineers Research and Development Laboratory, Fort Belvoir, Virginia, October 1958.

28. Brainerd, R. W., E. C. Hanford, and R. H. Marshall, *Resolution Requirements for Identification of Targets in Television Imagery*, NAA Report NA-63H-794, 1965.

29. Oatman, L. C., *Target Detection Using B/W TV, Study II: Degraded Resolution and Target Detection Probability*, U.S. Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, Report TM-10-65, July 1965.

30. Rosell, F. A., "Limiting Resolution of Low-Light-Level Imaging Sensors," *J. Opt. Soc. Am.*, Vol. 59, May 1969, p. 539.

31. Parton, J. S., and J. C. Moody, "Performance of Image Orthicon Type Intensifier Tubes," *Proc. Image Intensifier Symposium*, Fort Belvoir, Virginia, Report NASA SP-2, October 1961.

32. *Electro-Optics Handbook*, RCA Aerospace Systems Division, Burlington, Massachusetts, 1968.

33. Coltman, J. W., and A. E. Anderson, "Noise Limitations to Resolving Power in Electronic Imaging," *Proc. IRE*, Vol. 48, May 1960, p. 858.

34. Steedman, W. C., and C. A. Baker, "Target Size and Visual Recognition," *Human Factors*, Vol. 2, August 1960, p. 120.

35. Schade, O. H., "An Evaluation of Photographic Image Quality and Resolving Power," *J. Soc. Motion Picture and Television Engineers*, Vol. 73, February 1964, p. 81.