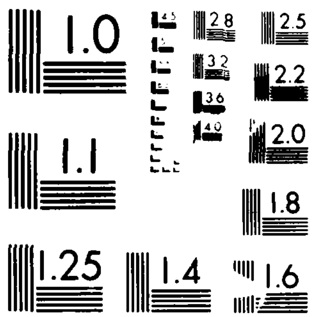
AD-A183 576 PSYCHOPHYSICAL CONSIDERATIONS IN MEASURING MRTD (MINIMUM RESOLVABLE TEMPE..(U) ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN (ENGLAND) K S MURPHY MAY 87 1/1

UNCLASSIFIED RSRE-MEMO-4031 DRIC-BR-102762 F/G 17/5.1 NL

END
DATE
FILMED
9 87



MICROCOPY RESOLUTION TEST CHART

AD-A183 576

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 4031

TITLE: PSYCHOPHYSICAL CONSIDERATIONS IN MEASURING MRTD WITH STARING ARRAYS

AUTHOR: K St J Murphy

DATE: MAY 1987

SUMMARY

Large variations in MRTD data can partly be attributed to inconsistent observer response biases. Psychophysical techniques which can minimise the errors introduced by such observer response bias are briefly described. These techniques are applied to measurements of MRTD performance for staring array thermal imagers.

PSYCHOPHYSICAL CONSIDERATIONS IN MEASURING MRTD WITH STARING ARRAYS

SUMMARY

Large variations in MRTD data can partly be attributed to inconsistent observer response biases. Psychophysical techniques which can minimise the errors introduced by such observer response bias are briefly described. These techniques are applied to measurements of MRTD performance for staring array thermal imagers.

1. GENERAL INTRODUCTION

One of the most frequently quoted parameters used in the general evaluation of the temperature sensitivity and spatial resolution of thermal imaging systems is the Minimum Resolvable Temperature Difference (MRTD). MRTD figures are usually obtained using standard NATO four-bar targets (see Figure 1), in conjunction with experienced human observers, the observers being required to judge when a four-bar pattern of a given spatial frequency is just resolvable on some CRT display. The results from such measurements are, as a consequence, heavily dependent not only on the characteristics of the thermal imager itself, but additionally on the abilities of the observers to make a sensible judgement about the visibility of the four-bar pattern, and the extent to which they are prepared to co-operate. It is therefore apparent that unless great care is taken to provide an observer with some sensible criterion against which he or she can made a reproducible judgement about the visibility of the bar patterns, that MRTD data for different observers will vary considerably.

Psychophysical techniques can minimise the variations in subjectively assessed measurements, but need to be carefully applied and controlled so that meaningless data is not inadvertently obtained. This memorandum is intended to provide readers with a basic, non-mathematical, introduction to pyschophysical techniques, with the emphasis in later sections placed on applications to MRTD measurement with staring array thermal imagers.

2. PSYCHOPHYSICAL MEASUREMENT

Haber and Hershenson (1973) define classical psychophysics as "an area of psychology that attempts to determine the functional relationship between a perceived or subjective magnitude and a physical magnitude". In the case of MRTD measurements the perceived magnitude is the apparent difference to the observer in the brightness between parts of the 4-bar patterns on the CRT, the corresponding physical magnitude being the measured temperature difference between the "hot" and "cold" bars of the four-bar pattern.

2.1 Frequency of Seeing

An elementary concept in psychophysics is the frequency of seeing (FOS) curve, an example of which is illustrated in Figure 2. If one presents an observer with some stimulus and plots changes in the observer's detection rate as a function of the physical variable being manipulated in the experiment, then a curve similar to that in Figure 2 should result. For example, if one presents a light stimulus to an observer, the probability that the observer will detect the presence of the stimulus should increase if one increases the brightness of the light flash, i.e., increasing stimulus strength results in a detection rate increase. Eventually a point is reached where the observer will be able to detect 50% of the flashes. This point is usually termed the "threshold". In theory, the threshold is the point of the beginning of perceptual experience. Below the threshold no awareness of stimulation is possible and above it there is always some experience of stimulation to greater or lesser degrees. Although some information about an observer's sensitivity to stimulation can be obtained from the FOS curve, one factor which cannot be taken into account is the manner in which the observer's response bias, i.e., readiness to say "Yes" or "No" to the presence of stimuli affects the outcome of the experiment. Very different FOS curves and, therefore, detection thresholds can be produced by different observers, or a single observer at different instants, simply because one observer's willingness to say "Yes" to stimulation is greater than another, more cautious observer, who may require further evidence of the presence of a stimulus before being prepared to say "Yes". In such cases, particularly if stimulus strength has been varied, the experimenter is unable to decide whether or not to attribute differences in threshold to changes in the physical variable, ie, stimulus strength, or to alterations in the observer's response bias during the experiment. Such data is, therefore, at best somewhat unreliable, and at worst absolutely useless. Eliminating serious errors due to observer response bias is one advantage of employing a sound psychophysical method of collecting subjective data. A description of some of the more commonly used techniques follows in the next section after some concepts of Signal Detection Theory are introduced.

3. SIGNAL DETECTION THEORY

Signal detection theory stands for a set of procedures and analyses that is explicitly designed to measure perceiver biases and assess precisely their contributions to the measurement of sensitivity. In Signal Detection Theory it is assumed that the observer sets up some criterion level for the response to stimulation and if the level of neural activity during a trial exceeds some critical value then a detection is registered, otherwise no response is elicited. More precisely, the observer is required to distinguish beween a signal plus noise during the trial and noise alone. It is prudent, and more often than not vital to introduce catch-trials so that the observer is unaware of whether or not a stimulus will appear during a set interval. This eliminates any tendency the observer may have to cheat. After a series of trials and catch-trials a set of responses are obtained which can be classified as follows. A "hit" occurs when the observer says "Yes" in the presence of a stimulus (trial) and a "false-alarm" occurs when the observer says "Yes" in the absence of a stimulus (catch-trial). If the observer says "No", the response is recorded as a "miss" if a stimulus was actually present and a "correct reject" if not present.

For a given criterion level, keeping stimulus strength constant, a hit rate and false-alarm rate can be calculated. By asking the observer to alter his/her criterion, ie, to be a little more or less risky in responding, both the hit rate and false-alarm rate can be altered. If the hit rate is plotted against the false-alarm rate, for a range of criterion levels, a Receiver Operating Characteristic - ROC - curve is produced. ROC curves represent the various degrees of response bias adopted when signal strength is held constant and are therefore useful for comparing differences in observer sensitivity without confusing variations in response bias.

3.1 Forced-Choice Response Indicators

An alternative paradigm to the "Yes-No" type of response discussed earlier is the forced-choice procedure (Blackwell, 1953), in which the observer is not allowed to say "No" and consequently cannot bias the "Yes-No" decision. Usually in this type of procedure, the experimenter tells the observer in advance that a stimulus will always appear in a given trial in either one of two possible locations. The observer's task is to decide in which of the two possible locations the stimulus appears. This type of task is called a Two-Alternative Spatial Forced-Choice task, the procedure being called "forced-choice" since the observer is always forced to say "Yes" no matter how uncertian he/she is about the spatial location of the stimulus. An alternative - the Two Alternative Temporal Forced-Choice-Task-removes the uncertainly about the spatial location of a stimulus during a trial, but the observer does not know in which of two different time intervals the stimulus will be present. The observer's task is to say in which of the two equal time intervals the stimulus was present. Again, the observer is always forced to say "Yes" irrespective of the level of uncertainty. The FOS curve obtained using a two-alternative forced-choice paradigm (2 AFC for short) always starts at 50%, since the observer can achieve this score even with his/her eyes closed. The threshold is usually taken to be the point at which the observer detects the stimulus correctly on 75% of trials.

4. PSYCHOPHYSICAL EXPERIMENTAL METHODS

In this section some psychophysical experimental methods are described which can be used in conjunction with Signal Detection Theory. The "Yes-No" and forced-choice indicators have been described and are the principal ones used in psychophysics. It has been pointed out that since the "Yes-No" indicator is sensitive to the effects of bias, it should never be used without some kind of bias control - either using some blank stimuli or a full signal-detection analysis. In any event, the 2AFC type of task is preferable and should be used whenever possible.

4.1 Method of Adjustment

The method of adjustment (MOA) is the simplest "psychophysical" method. In MOA, the observer is allowed to continuously adjust the strength of the stimulus until he believes that the stimulus is just visible, or in the case of MRTD, until some fraction (usually 75%) of the 4-bar pattern is resolved.

Although the method of adjustment is the quickest method to obtain a psychophysical function, there is no way of controlling observer response bias. Comparisons between subjects are therefore nearly always confounded with criterion differences. The method of adjustment should, therefore, be avoided wherever possible.

4.2 Methods of Limits

The method of limits is similar to the method of adjustment except that in this case the experimenter adjusts the stimulus parameters. One major difficulty with this technique is that the observer knows in advance that the next trial will contain a stimulus and that the energy of the stimulus will be higher than the previous one. The method of limits is often disregarded, therefore, in favour of the following method.

4.3 Method of Constant Stimuli

The method of constant stimuli is similar to the method of limits, but in this technique the sequence of presentation is randomised, the experimenter deciding which energy values to use. The observer, therefore, cannot predict the sequence of intensities of the stimuli. This method lends itself to use with a 2AFC (ie spatial or temporal) indicator, but results unfortunately require many trials. However, the data obtained by this method are generally more stable and reliable than by other methods.

4.4 Staircase Method

This method is only useful if the experimenter is interested in the region of threshold rather than the whole psychophysical function from "always detected" to "never detected", ie, the staircase method is designed to concentrate most measurements in the region of the psychophysical function where the observer will be least certain of the response. The procedure usually adopted is to start with the stimulus energy at some low value well below the expected threshold. If the observer does not perceive the stimulus during the trial, the stimulus energy is raised by a set amount (usually 0.1 log units) and the trial is repeated. This continues until the observer reports having seen the stimulus. When the observer sees the stimulus the energy is decreased by one unit. If the stimulus is still seen the energy is decreased again by one unit until a stage is reached where the observer again states that the stimulus can no longer be seen; then it is increased again. The whole procedure is repeated until the observer has adopted a fairly uniform rythm. The statistical threshold is then defined as the average of the stimulus intensities needed to change from "No" to "Yes", and from "Yes" to "No". To prevent the observer from setting the threshold simply by alternating the sequence of responses, and adopting a response strategy, it is common to violate the general rule about 25% of time, ie, a "No" response will cause a decrease in stimulus energy instead of the usual expected increase. Further improvements still can be introduced by running a simultaneous staircase. In a simultaneous staircase two sequences of trials are intertwined so that on any given trial the subject will not know whether a stimulus presentation will be from a low to high or a high to low sequence. The staircase method is particularly effective if used in conjunction with a 2AFC response indicator.

5. CONTRAST SENSITIVITY

The eye, or visual system if one wishes to consider higher level processing, responds primarily to contrast. Measurement of contrast sensitivity is, therefore, a good indicator of the condition of the visual system in general. Two types of contrast sensitivity measurement are reported widely in the literature; one to isolated stimuli (Blackwell, 1946), the other to sinusoidal grating patterns (eg Campbell & Robson, 1968). The measure of contrast sensitivity most relevant to current MRT measruement is the latter of the two since both involve determining contrast sensitivity to periodic stimuli. Efforts will therefore be directed towards a discussion of this measure alone.

The well-known Fourier theory shows that any square wave can be mathematically represented by a function of the form:-

$$f(x) = \frac{4}{\pi} \{\sin kx + \frac{1}{3} \sin 3kx + \frac{1}{5} \sin 5kx + \}$$

where k = wave number = $2\pi/\lambda$.

Objects in space that have sharp well-defined edges contain high spatial frequencies, whereas blurred edges contain lower spatial frequencies or a detrimental phase shift has occurred. Our ability to detect edges depends on the sensitivity of the visual system to the spatial frequencies of which the edge is composed.

Recent decades have seen the introduction and eventual widespread acceptance of a test of spatial vision which is based on the assumption that the visual system can be described using a linear systems analogy for small signals (ie low contrast). A lens system is described as linear if a spatial sine wave input of a given frequency and amplitude produces an output sine wave of the same frequency. In practice, the output sine wave is often reduced in amplitude and phase shifted. The variation in output amplitude as a function of spatial frequency of the input sine wave is called the Modulation Transfer Function (MTF). A related function in physiological optics is the Contrast Sensitivity Function (CSF).

The CSF is measured using stimuli with a sinusoidal luminance distribution in space. Such stimuli are commonly called gratings. In an experiment to determine the CSF for an individual, the observer is required to look at a grating pattern of a given spatial frequency and adjust the contrast of the grating until the pattern is no longer visible. This test is then repeated for a wide range of spatial frequencies up to the limit of visual acuity.

The basic shape of the CSF is that of a band-pass spatial filter. Over the range of spatial frequencies 0-60 cycles/deg (of visual angle) or 0-3.4 cy/mR (of visual angle), the typical CSF is shown in Figure 3, where the upper limit of visual acuity is at about 50-60 cy/deg.arc (correpsonding to about 1 min.arc resolution at the fovea).

The shape of the CSF curve shows that the visual system is optimally sensitive to spatial frequencies around 3-5 cy/deg (0.1-0.2 cy/mR). Campbell and Robson (1968) studied contrast sensitivity to stimuli composed of multiple spatial frequencies (ie square and triangular waves) in addition to sinusoidal waves. Some of their results are shown in Figure 4. Their main conclusions were that at medium and high spatial frequencies, detection is a function of the fundamental frequency component alone, but that discrimination of square waves and triangular waves of the same fundamental frequency from sinusoidal waves, could only be achieved when the higher harmonics present in the square and triangular waveforms were themselves of sufficient amplitude to permit detection. At low spatial frequencies all waves tested, apart from sine waves, tended to have a constant contrast detection threshold. The evidence from Campbell & Robson's work gave rise to the now widely-held belief that there are tuned spatial frequency channels in the visual system which operate in parallel, and that the overall CSF curve represents the envelope of these parallel spatial frequency channels.

The implications for measurement of MRTD are clear:-

a) The recognition threshold for the 4-bar pattern will be determined by the higher spatial frequency components as well as by the fundamental itself. Detectioni s achieved via the lower spatial frequencies.

b) If the observer is allowed to alter his/her viewing distance during MRTD measurements (as is common) then the spatial frequency of the bar pattern subtended at the eye will alter correspondingly and with it the observer's contrast sensitivity to the bar pattern on the CRT.

c) If the observer knows in advance to expect a 4-bar pattern, the task becomes one of detection, ie awareness of the presence of the stimulus, rather than recognition.

Although it is accepted that normal practice in conventional operation of an indirect-view thermal imager is for the observer to continually alter the viewing distance to aid in target acquisition, it nevertheless introduces another undefined quantity into MRTD performance measurement and, therefore, is undesirable. In the case of staring array imagers, where the imaging optics is more likely to be direct-view, the viewing distance must be kept fixed. In addition, where one is interested in comparing/ contrasting thermal and spatial resolution properties of different types of 2-D focal plane array, it is important that differences in performance can be attributed to the physical properties of the system rather than to observer variables.

5.1 MRTD Measurement on Staring-array Imagers

Future thermal imagers are likely to be of the staring array type, as opposed to the present day imagers which rely on a scanning mechanism and consequently are rather complex and expensive. A major factor when considering MRTD data for staring array systems is aliasing. Thermal imagers which use 2D focal plane arrays are limited in their spatial frequency response. The highest spatial frequency that can be reproduced by such an imager is determined by the detector pitch, and is commonly referred to as the Nyquist frequency (N_f) where:-

$$N_f = 1/2P_s \; : \; P_s = \text{sampling pitch}$$

Spatial frequencies greater than the Nyquist frequency are aliased back into the 0 to N_f spatial frequency region and will therefore appear at a lower spatial frequency. Recognition thresholds will, to some extent, be dictated by the observer's sensitivity to high spatial frequencies. If these spatial frequencies needed to effect recognition are aliased into the lower spatial frequency region, then the probability of an observer incorrectly identifying features of a possible target must be considered and evaluated. Current MRTD procedures have no firm basis for estimating the probability of incorrectly recognising targets, primarily because the observer always knows in advance that the MRTD target will be a four-bar pattern only. In this case, the observer's task resembles that of detection only. It is therefore advisable, from the psychophysical point of view, to ensure that observers working with four-bar MRTD targets only are not allowed to have an "a-priori" certainty that the four-bar pattern will appear during a trial, ie, some catch-trials must be introduced. An alternative approach, proposed by Mort and Schnitzler (1985), involves presenting the observer with varying types of target similar to the standard four-bar MRTD target but with one or more bars missing. The observer's task is to adjust the strength of the signal until he is able to identify correctly on 50% of occasions all the bar patterns of a given spatial frequency. The point at which this occurs is then recorded as the minimum resolvable temperature difference for that spatial frequency. This approach not only allows a far more objective and, therefore, reliable assessment of an observer's ability to recognise targets of a given spatial frequency, but also permits some estimation of the degree of aliasing present in the system to be made.

6. RECOMMENDATIONS

The method used to obtain MRTD data will depend on the availability of suitable equipment and on the task required of the observer. If the MRTD measurement involves resolving all or part of a bar pattern it is the author's opinion that a staircase method be adopted since in MRTD measurements only the psychophysical function in the region of threshold is of interest to the experimenter. Unfortunately, a 2AFC response indicator will be of little use with staring array imagers because the observer ideally needs to pan the 2D focal plane array around the scene to the reduce errors from aliasing, particularly in the region of the Nyquist frequency. However, for other types of imager, where the effects of aliasing are less severe, a 2AFC paradigm should be considered. For determining the MDTD (Minimum Detectable Temperture Difference), which is primarily a detection task, a simultaneous staircase method in conjunction with a 2AFC response indicator would eliminate criterion difficulties and, if applied correctly, reduce observer threshold variability.

7. CONCLUSIONS

The reliability of MRTD data can only be as good as the performance of the observer(s) used to collect the figures. Evidence from the literature shows that variations of 30% or more are found in MRTD data - much of this may be attributed to the methods used to obtain the data. In development stages, where performance of individual prototype imagers may require careful assessment, efforts should be directed towards controlling observer variations using well-established psychophysical methods. Variables such as gain, offset and viewing distance should be clearly defined. These parameters should be kept constant if comparing two or more types of imager. Furthermore, it is sound practice to use at least six observers in a psychophysical study so that sensible statistical analysis can be carried out on the results.

It is accepted that on a production line, it is neither practical nor efficient to take weeks or even days to arrive at an MRTD curve and to this end, efforts are being directed toward the development of an objective method of measuring MRTD performance. In the meantime on the production line basis the method of adjustment may prove to be the only quick method for obtaining MRTD data from human observers even though its reliability is questionable.

REFERENCES

Blackwell, H R (1953)
"Psychophysical thresholds: Experimental studies of methods and measurements".
Bulletin of the Engineering Research Institute, University of Michigan, No 36.

Blackwell, H R (1946)
"Contrast thresholds of the human eye".
J. Opt. Soc. Am Vol 36, pp 624-643.

Campbell, F W and Robson, J G (1968)
"Application of Fourier analysis to the visibility of gratings".
J. Physio l Vol 197, pp 551-566.

Mort, M S and Schnitzler, A D (1985)
"Quantitative Relationship between sampling effects and the performance of sampling imagery video communication systems".
BDM Report No BDM/W-85-0491-TR.

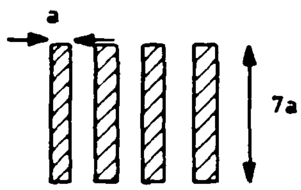


FIG.1. STANDARD NATO 4-BAR TARGETS USED IN MEASUREMENT OF THE MINIMUM RESOLVABLE TEMPERATURE DIFFERENCE (MRTD)

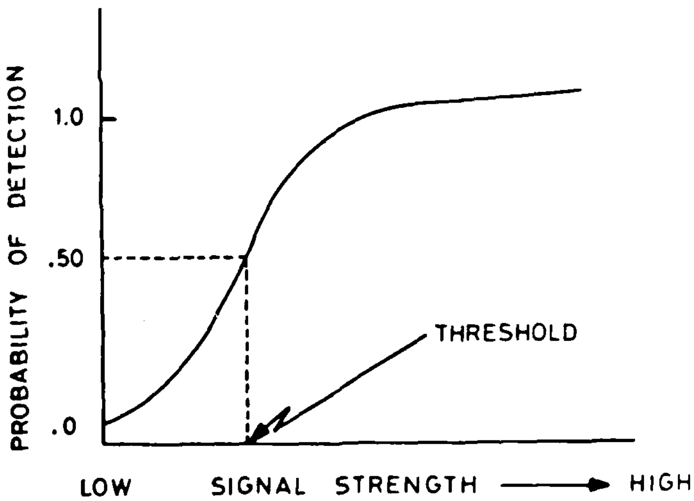


FIG.2. TYPICAL FREQUENCY-OF-SEEING CURVE DESCRIBING THE RELATIONSHIP BETWEEN THE PROBABILITY OF DETECTING A SIGNAL AND SIGNAL STRENGTH

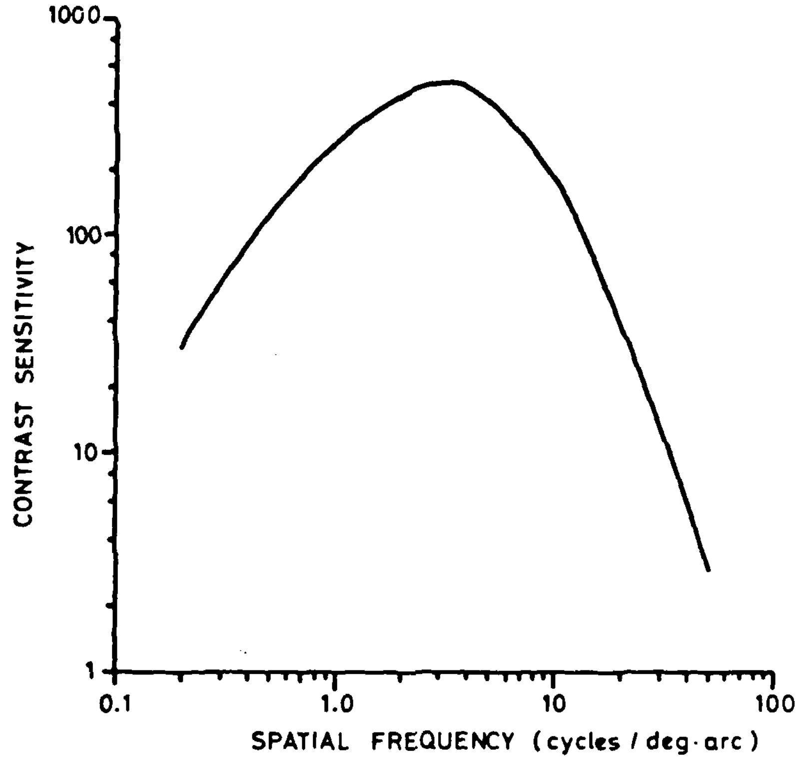


FIG. 3 TYPICAL CONTRAST SENSITIVITY CURVE FOR PHOTOPIC LEVELS OF LUMINANCE

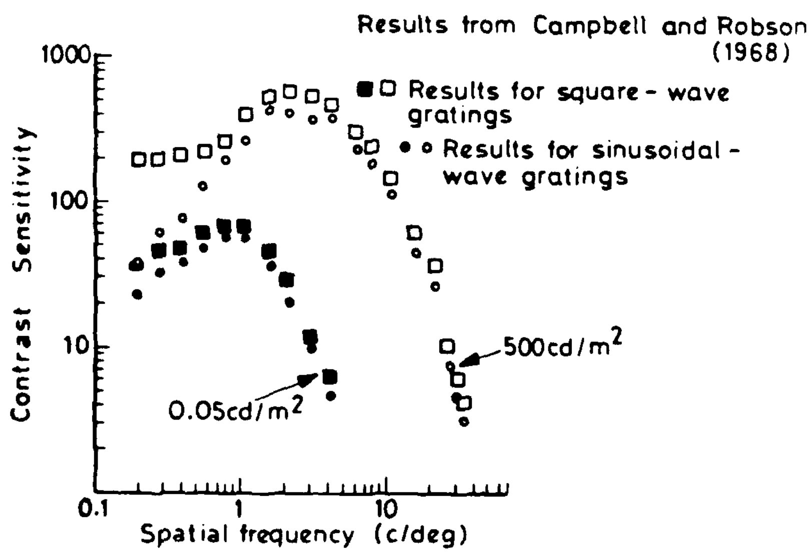


FIG. 4.

END

DATE

FILMED

9-87

DTIC