Psychophysical basis and validation data supporting the Army's Night Vision Integrated Performance Model

RICHARD H. VOLLMERHAUSEN

RHV Electro-Optics, LLC, 1173 Gatwick Loop, Lake Mary, FL
**xyz@optica.org*

Keywords: Image quality metrics, Imaging, Target Acquisition

Abstract: The Army replaced the Johnson Criteria with the Targeting Task Performance (TTP) metric circa 2001. However, the developers of the Night Vision Integrated Performance Model (NVIPM) decided to change the definition of a critical TTP input, and the modified metric has not been validated. Nonetheless, current peer reviewed articles refer to NVIPM as a robust target acquisition model and reference the original validation data in support of that model. This paper describes the change from TTP to NVIPM and summarizes the validation data supporting each metric.

1. Introduction

This paper describes the validation method and data supporting two different image quality metrics (IQM). The first IQM was the basis for the 2001 release of the Night Vision Thermal and Image Processing (NVThermIP) model [1-7]. The second IQM is the basis of the current Army target acquisition model, the Night Vision Integrated Performance Model (NVIPM) [8,9]. The Army model developers use the same name for both metrics. In this paper, the original metric is called Targeting Task Performance (TTP), and the modified metric is labeled NVIPM.

The two metrics are confused in the literature, and validation data supporting the original metric is incorrectly associated with NVIPM [10,11]. The paper [12] described the difference in the metrics and presented a comparison of the accuracy of metric probability predictions for a limited set of validation data. This paper summarizes all the validation data supporting each metric. Also, the psychophysical and targeting task assumptions underlying the two metrics are described.

The original TTP compared the frequency content of the displayed target to single frequency eye thresholds. When measuring eye thresholds at each display luminance, the sine wave pattern has enough periods to provide single-frequency threshold values. Also, the lit area on the display is big enough that the eye adapts to the display luminance. At left in Figure 1, the data from several individuals are averaged to create a Contrast Threshold Function (CTF) of thresholds versus spatial frequency at a variety of display luminances. At right, the observer is represented by the average CTF of several young observers with good eyesight, and the eye thresholds are compared to the Fourier spectrum of the targets.

NVIPM uses thresholds measured using the sine wave patterns illustrated in Figure 2 at left. The CTF sine waves are limited to a specific angle subtended at the eye, and data are taken at many angles. Three angles are shown in the figure, but data are taken one pattern at a time and one subject at a time.

In the NVIPM, the target is viewed on a display the size of the target. If the target angular size is smaller than ten or fifteen degrees, the CTF experiment subject's eyes were not luminance adapted, and that raises threshold values. Also, if there are fewer than seven sine wave periods, then single-frequency thresholds are not obtained using that experiment design.

Figure 1 showing arrangement using single frequency thresholds and luminance adapted eyes.

Figure 2 showing arrangement using few sine wave periods and observer squinting.

In other words, the original TTP is a frequency domain model that uses eye thresholds established with luminance adapted eyes when viewing at least seven sine wave periods. NVIPM uses thresholds measured with non-luminance-adapted eyes viewing sine wave periods that fit on a display the size of the displayed target image. The reason for the change has not been explained.

Both TTP and NVIPM generate CTF values using a numerical algorithm published by Barten [13].When the developers of NVIPM argue that the TTP metric does not use Barten's CTF algorithm correctly, we cite Beaton's analysis stating that Barten's algorithm generates single-frequency thresholds if a wide display angle is input, and Beaton recommends that algorithm because it is both accurate and easy to use [14].

The TTP metric uses Barten's CTF numerical algorithm with a wide-angle input to generate single frequency CTF values. Also, Barten's numerical approximation of the visual cortex spatial filter channels is used in TTP metric calculations, but his IQM metric and other research are not. Whatever the reason that Barten used the CTF data of Carlson for his own research purposes [14], Barten's CTF algorithm must be used with a wide-angle display input to generate the single frequency threshold values needed for the TTP algorithm. Again, the reason for doing otherwise has not been explained by the current Army researchers.

Figures 3 and 4 show an example to illustrate TTP and NVIPM differences. The NVIPM metric has a different range dependence than the TTP IQM. Figure 3 shows metric values versus range, and Figure 4 shows PID when viewing a tactical vehicle target set through a spotting scope with 4.3-inch diameter aperture and 20 magnification. The target set has 0.24 contrast, and the atmosphere has 0.5 transmission per kilometer through aerosol. The two metrics generate different values given the same target set and range, and the behavior of PID versus range is not the same.

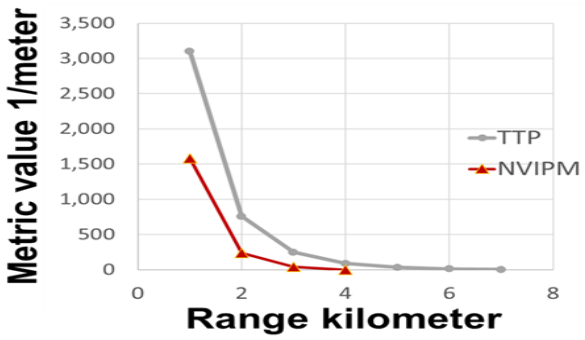

Figure 3. The TTP and NVIPM metrics yield significantly different metric values at each range.

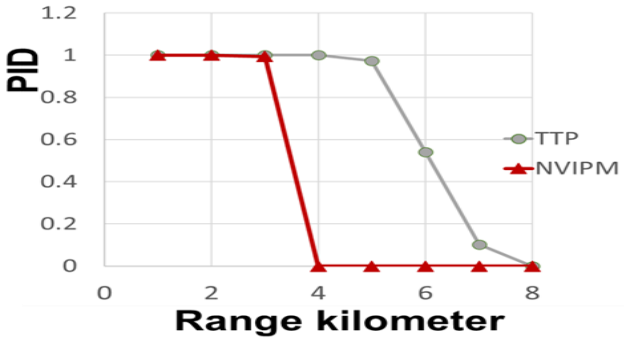

Figure 4. The PID versus range values for either metric can be changed by adjusting a task-difficulty parameter. However, the shape of the range curve will not change.

In the years between 2003 and 2008, that change was tested using the TTP data base and the accuracy of probability predictions were degraded. Only part of that validation exercise has been published, and that available data will be discussed in Section 5. Nonetheless, the angle of individual targets subtended at the observer's eye is now input to Barten's algorithm to calculate the metric values used in NVIPM.

Section 2 describes the function of target acquisition models and the visual task used to test the efficacy of those models. Section 3 provides background on several subjects to simplify the discussion in Sections 4 and 5. The misstatements about the TTP noise model are corrected in the background section. Sections 4 and 5 describe the original TTP IQM and NVIPM validation data, respectively. Conclusions are in Section 6.

2. The function of a target acquisition model and the visual task used for test

An imaging sensor bought today will likely be used for the next three decades. That imager will be used by someone, somewhere in the world, to do some visual task. The cost, reliability, maintainability, weight, size, and power are all factors in a purchase decision, but performance is also a critical consideration. What do we mean by "performance" and how can it be evaluated?

Figure 5 shows a person trying to identify an approaching tactical vehicle. Many factors affect the rendering of the target on the display, including the size and contrast of visual cues like the target silhouette, the atmosphere, the interplay of optics, detector, digital processing, and imager gain and level, plus the display blur, brightness, contrast, and glare. One imager might provide long range imaging under good contrast conditions but does not image when contrast drops, and another imager might provide moderate range target identification under all

expected environmental conditions. We can only evaluate which imager best meets operational needs if we have a way of predicting imager behavior under diverse conditions against a variety of targets. That is, we need an IQM that predicts the effect of myriad factors on a person's ability to visually understand scene details.

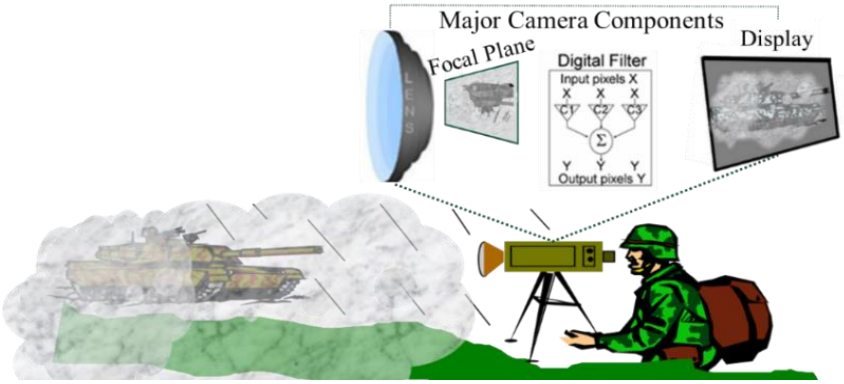

Figure 5 shows a man trying to identify an approaching vehicle, and many factors affect his ability to identify the target.

The ability to identify objects in a scene, the "what is it and where is it" capability of human vision, are higher order visual discriminations, and that image quality is quantified by the ability to perform those visual tasks.

Figure 6 illustrates the probability of identification (PID) task. PID is the average probability that trained observers will identify the objects in a test set. A test set might include tactical vehicles, letters, faces, shapes, or other things. The experiments are always forced choice, and the subjects are trained and tested on their ability to identify all the objects in the test set.

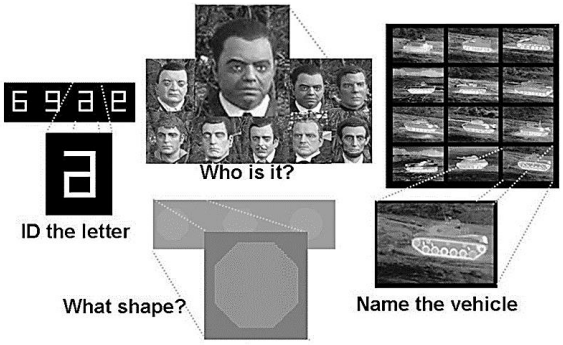

Figure 6 illustrates the probability of identification task.

Note that PID is defined by the set, because we cannot quantify the difficulty of a PID task without specifying the alternatives. For example, Egypt was an ally of the United States in the conflict with Iraq, and Egypt operated both American M60 and Russian T62 tanks. Iraq fields Russian T72 tanks. The three tanks are shown in Figure 7. A friend verses foe decision would be more difficult if the Egyptians fielded T62 tanks than if they fielded the M60.

As another example, consider a patient at the Optometrist's getting a new prescription for glasses. The patient is asked to read the next smaller line on the eye chart. The first letter is an L; that is easy. The next letter is an M, also easy. But the third letter might be a C or maybe an O. If there were no O in the alphabet, the patient would read the next line, but that would not make his or her eyesight better. The probability of identifying the letter C increases because the absence of the letter O makes the visual task easier.

Figure 7 shows three aspects each of eight tactical tracked vehicles used in a metric validation field test [4]. Note that the T62 and T72 look alike but are different than the M60 even though all are tanks. The M113 is a box, and it looks like no other vehicle in the eight-vehicle set. Figure 8 shows the PID versus range curve for one imager configuration. The PID drops to 0.9 at close range because some vehicles like the T62 and T72 Russian tanks look alike at most aspects, and they get confused. At mid-range, some aspects of other vehicles look somewhat alike, and PID value drops. At long range, the little M113 has a high PID, presumably because of its box shape.

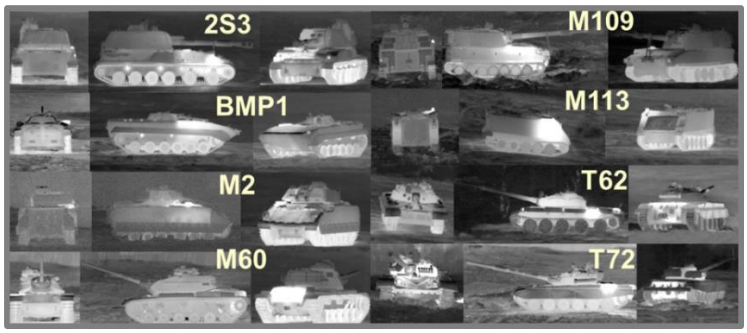

Figure 7 shows the three aspects of eight vehicles used in a field test. Note that some vehicles look alike, and others have a unique shape.

The visual cues that differentiate each tracked vehicle vary from vehicle to vehicle and aspect to aspect. Fender shapes are different, the thermal heat patterns from engine, exhaust, and road wheels varies from vehicle to vehicle, track and wheel configurations vary, hatch shape and locations are different between vehicles, and so is the location of a gun barrel, if present. PID range is determined by the fraction of the set visual cues identified at each range. The "target signature" is the average size and contrast of the visual cues for the entire target set. The vehicles are distinguished by a diverse variety of visual cues associated with individual vehicle aspects. PID is the average probability for the complete set; the PID for tactical vehicles does not apply to each or any member of the target set.

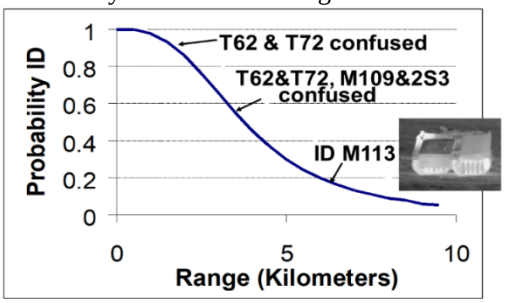

Figure 8. PID versus range for a diverse target set.

The character set in Figure 9 is also a diverse target set because the characters vary in size and contrast. The character set in Figure 5 can be represented by a Normal Distribution with an average size and contrast. Clearly, the PID for the Figure 5 target set does not apply to each member of the set.

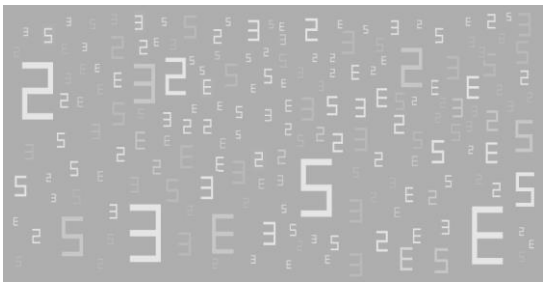

Figure 9 is a diverse visual cue set, and the visual cues can be represented by a Normal Distribution.

Figure 10 shows a tank receding in range from a thermal imager. The red squares represent photo detector pixel instantaneous field of view. As range increases, the tank becomes angularly smaller in the imager field of view and therefore smaller on the display, fewer and fewer pixels intersect the vehicle, and the frequency content of visual cues transitions to higher and higher spatial frequencies within the camera electronics.

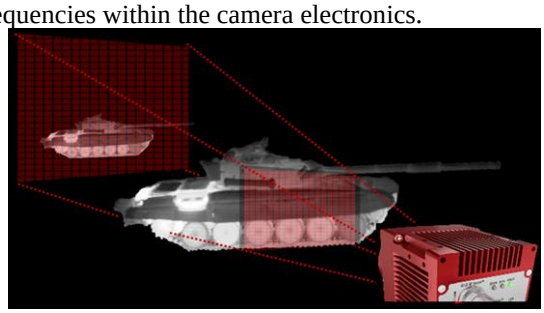

Figure 10. As the tank recedes in range, the Fourier Transform of the spatial structure transitions to higher frequencies in the camera electronics.

The signature of a target set with diverse discrimination cues is represented by a Normal Distribution with average size and contrast. As the set recedes in range, the number of discrimination cues that are visible through the camera decreases and therefore PID for the set decreases.

On the other hand, the signature for the Figure 11 symbol set is the Fourier Transform of one rectangle. Since the characters get smaller on the display as range increases, the Fourier Transform of the rectangle is transformed into imager angle space and is range dependent. In that case, we know the spatial frequency content of the discrimination cue and should be able to predict PID. Again, the PID is the average for the set, but in this case, PID also applies to each member of the set.

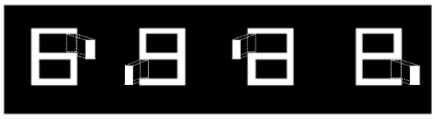

Figure 11 shows a specific object set, and the Fourier transform of a character (or the rectangle) must be included in the metric calculation.

The visual cues that discriminate faces do not have the shape and size diversity of the tactical vehicle set, and the visual cues for facial discrimination cannot be treated as a Normal Distribution. The Fourier Transform of a representative face must be included in the TTP calculation, because we know that the facial discrimination cues reside within that frequency spectrum. When the target set consists of look-alike objects, then PID versus range depends on the camera rendering that specific type of object, and an example Fourier spectrum must be included in the scene-to-display Fourier Domain Transfer Function.

3. Background topics

3.1 Modeling video cameras

This section provides background for the later discussion of TTP versus NVIPM treatment of imager noise.

Figure 12 shows the layout of one image intensifier ocular. Scene light is amplified and presented almost instantaneously to the eye, and that allows our eyes to view the scene in a normal, naked-eye fashion. It is true that the I^2 adds noise and blur and limits field of view, but the primary effect is a brighter scene so that we can see on a dark night.

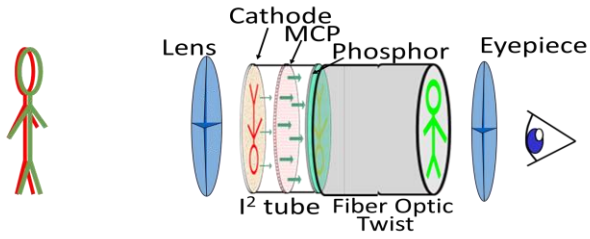

Figure 12 shows the components in one image intensifier ocular.

Both movies and video, whether television or computer video, image the scene quite differently than image intensifiers, because they depend on what psychophysicists call smooth apparent motion [24]. Figure 13 shows a strip of film. When placed in a projector, the first picture, a snapshot of the scene, is moved into place while the projector light is covered by a shutter. Once the movie frame is in place and film motion has stopped, the shutter is opened, and the static scene is projected onto the movie screen. Then the shutter is closed, and the next movie frame is moved into place.

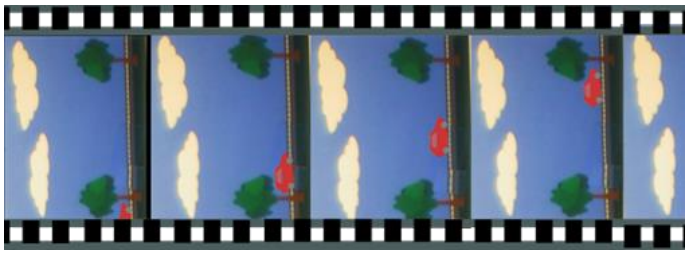

Figure 13. A strip of film is a series of still pictures. When presented in time sequence, however, the eye sees smooth apparent motion.

Apparent motion refers to the perceptual phenomenon where still images displayed in rapid succession are perceived as moving. The eye executes smooth ocular pursuit of the statically projected pictures, and we perceive smooth motion of the car.

Video also depends on smooth apparent motion. See Figure 14. Video is a series of snapshots presented $1/60^{th}$ of a second apart. What the eye sees is smooth apparent motion of the car across the display.

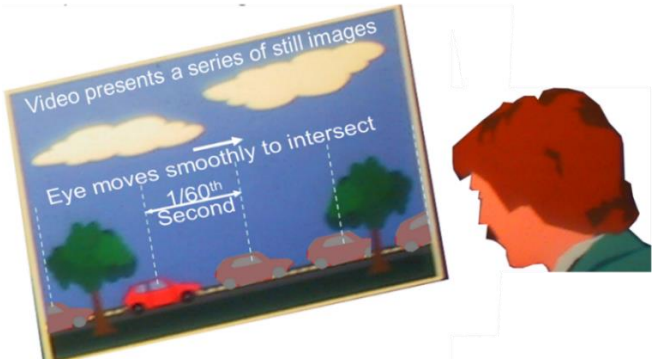

Figure 14. Video, like film, depends on smooth apparent motion.

Smooth apparent motion requires gathering a series of still pictures, and each picture must be of the quality desired in the final image. If the scene moves relative to the camera during exposure, that causes blur. Also, because the eye fixation point is moving smoothly across the display, if the picture presented on the display is held on for too long, then the scene is blurred even if the picture is sharp.

3.2 Predicting the effect of aliasing on target identification

The Aliasing as Noise (AAN) experiments modeled ten imager designs at six ranges [15,16]. The AAN experiments provided data supporting TTP validation. Further, display viewing distance was controlled during the AAN series of experiments.

The NVIPM uses an early version of sampling correction [3,17]. The first set of sampling experiments occurred before we had characterized the displays for perception experiments. Therefore, targets were displayed like the one shown in Figure 15; all the targets were placed at close range. The problem with our first experiment was that the visual cues that differentiate the target set vehicles were not aliased. Only unimportant target details like bolts on the hatch were aliased. The position of road wheels, fender and hatch shapes, heat distribution from the engine and exhaust, those visual cues were not aliased. We incorrectly concluded that in-band aliasing had no effect on imagery interpretation.

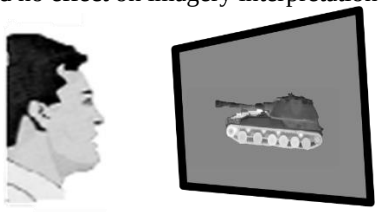

Figure 15. In the Equivalent Blur experiment, all targets were placed at close range.

After the displays were properly characterized, additional experiments were performed with the targets realistically sized on the display by placing them at longer ranges, and the results are shown in the Figure 16 graphs [15,16]. The graphs show PID with no sampling correction (NC), correction using the Equivalent Blur (EB) model [17], and PID correction using the AAN model. Figure 16 show four of the ten imager designs modeled and illustrate the accuracy of AAN predictions.

Note that aliasing is generated over the target area, and that normalizing the aliasing noise requires inputting the actual presented area of the target. The TTP model does not depend on target area unless aliasing is present.

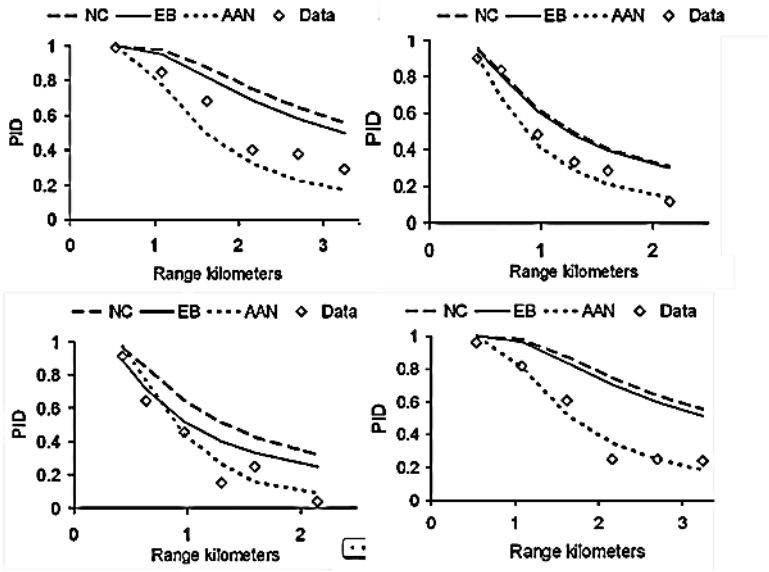

Figure 16. The plots shows PID for no correction (NC), EB, and AAN.

3.3 Fourier Domain representation of target signatures.

Diverse and specific target sets were described earlier in this paper; see Figures 9 and 11. When the target set consists of look-alike objects, the original TTP metric is calculated using the Fourier Transform of a set member. That is, for a set of faces or characters or shapes, the Fourier Transform of a set member is included in the TTP calculation, and the scene-to-display Transfer Function includes the Fourier Transform of a set member.

For diverse visual cue target sets, the TTP metric uses root of the sum of squares (RSS) averaged over the target set. It appears that NVIPM uses area weighted average temperature (AWAT) [9]. RSS for thermal and reflective images are defied by Equations 1 and 2, respectively. AWAT is defined as the area weighted temperature of the target minus the area weighted temperature of the local background.

$$\Delta T = \sqrt{\left(\text{tgt mean - bck mean}\right)^2 + \sigma_{tgt}^2} \qquad 1$$

$$\Delta C = \frac{\sqrt{\left(\text{tgt mean - bck mean}\right)^2 + \sigma_{tgt}^2}}{\left(\text{tgt mean } + \text{ bck mean}\right)} \qquad 2$$

All reflective contrast values in this paper are modulation contrast consistent with a Fourier Domain model. For thermal contrast, ΔT is Effective Blackbody Temperature (EBT) in Kelvin (K). EBT is defined as the imager in-spectral-band-flux generated by the rate of change of Plank's Equation at 300 K; see Equation 3 where L is Plank's blackbody equation, Γ is target contrast in Kelvin (K), T is temperature in K, and λ is wavelength in meters. Although ΔT is called temperature, it is of course a radiometric unit and not the temperature of the viewed object. For both thermal and reflective contrast, the mean difference of the target area and an equal background area surrounding the target are RSS with the standard deviation σ_{tgt} of the target.

$$\Delta T = \Gamma \, \partial L\left(\lambda, T\right) / \partial T \qquad 3$$

Plots of RSS and AWAT contrasts for twelve aspects each of twelve tactical vehicles are shown in Figure 17. AWAT values are always less than RSS, and on occasion, AWAT contrast can be near zero when the target is easily visible. Figure 18 shows tracked vehicles with high and low RSS contrast and high and low AWAT contrast.

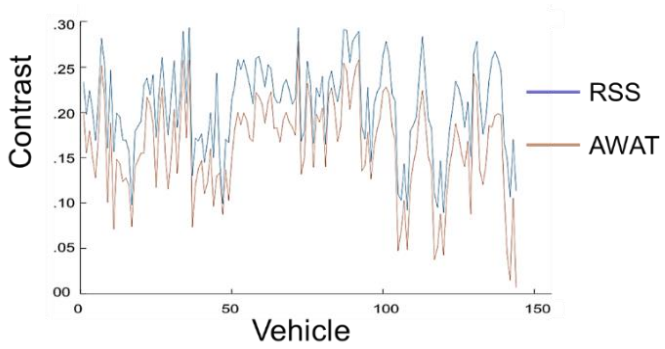

Figure 17. Plot of RSS and AWAT contrasts for 12 vehicle/12 aspect target set.

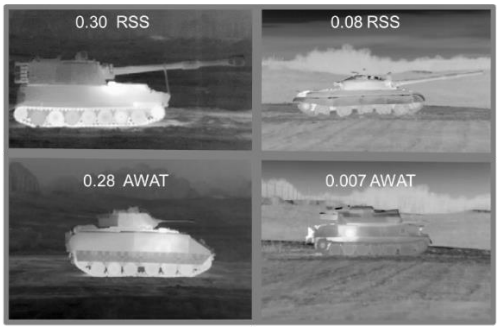

Figure 18. The maximum and minimum contrast vehicles with RSS at top and AWAT bottom.

The developers of NVIPM have not stated why they switched from RSS to AWAT. TTP uses RSS, and that decision was based on a mutual agreement with the authors of the Johnson/Ratches model and NVThermIP [3] and based on the strong recommendation of war game representatives. All TTP validation data is based either on RSS contrast or using the Fourier Transform of a set object.

3.4 Empirically derived task difficulty parameter

The difficulty of identifying one member of a target set depends on how much that object looks like other members of the set. The task difficulty parameter is established empirically by performing a PID experiment and curve fitting metric values to PID data. Once established, the same parameter value is used for all PID experiments, except the parameter values are different for emissive and reflective imagery. The parameters are called N50 when using the Johnson model, V50 when using either the NVThermIP or NVIPM models, and Φ84 when using the TTP in journal literature.

3.5 The functional relationship between metric values and probability.

The Central Limit Theorem states that, regardless of the underlying statistics, if there are many contributing factors to a statistical outcome, then the probability is related to the metric value

by the error function erf. It therefore seems logical to assume that PID should be functionally related to TTP IQM values by the erf function.

TTP metric value ϕ is related to PID by erf. See Equation 4 and Figure 19.

$$PID = \frac{2}{\sqrt{\pi}} \int_{0}^{\Theta/\Theta 84} e^{-t^2} dt \qquad 4$$

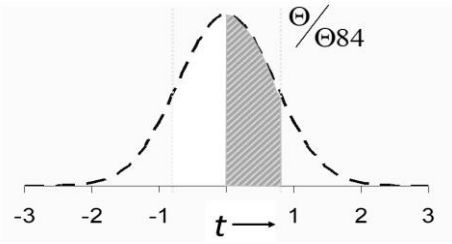

Figure 19 showing area integrated to get erf value.

NVIPM uses Equation 5 to calculate PID from metric values (V).

$$PID = \left(V/V50\right)^{1.5} \Big/ \left(1+\left(V/V50\right)^{1.5}\right) \qquad 5$$

3.6 Misrepresentation of TTP in Night Vision Integrated Performance Model literature

NVIPM literature has described many aspects of the original TTP metric incorrectly, and NVIPM has a wide audience. It seems prudent to dispel misconceptions before describing the TTP metric.

For example, the NVIPM developers state that TTP uses different noise parameters in the thermal and reflective models. The thermal model treats noise in the frequency domain and the reflective model treats noise in the time domain, so the values of the TTP noise parameter are different because the units are different.

We RSS eye noise and detector noise as shown in Equation 6

$$n_{total}^2 = n_{eye}^2 + \alpha_{\text{proportional}}^2 n_{\text{detector}}^2 . \qquad 6$$

For reflective imagers, the signal E_{av} photo electrons per second is integrated for an eye integration time t_{eye}.

$$signal = E_{av} t_{eye} \qquad 7$$

and the noise is

$$n_{\text{detector\_reflective}} = \sqrt{E_{av} t_{eye}} . \qquad 8$$

For thermal imagers, the signal and noise are:

$$signal_{thermal} = E_{av} \quad \text{per second} \tag{9}$$

$$n_{detector\_thermal} = \sqrt{E_{av}} \Big/ \sqrt{t_{eye}}. \tag{10}$$

The signal to noise is the same, namely $\sqrt{(E_{av}t_{eye})}$, but we RSS noise and not signal to noise. In our reflective equations,

$$\alpha_{photon} = \alpha_{proportional}\sqrt{t_{eye}} \quad = 169.6 \tag{11}$$

and there is no explicit dependence on eye integration time in those equation. For thermal, however,

$$\alpha_{detectivity} = \alpha_{proportional} \quad = \ 862 \tag{12}$$

and explicit t_{eye} dependence is in the thermal imager metric equations. The proportionality constant is the same, but the reflective equations count photons while the thermal equations model current. The reflective and thermal models use different units and treat eye noise differently mathematically, and that requires a different value for the proportionality constant when calculating the TTP metric value. If the noise parameters representing the eye were different, that would be a major flaw, but they are not. The need for different values for the reflective and thermal models was explained in a report included with the NVThermIP computer model [3].

Another example of misinformation about the TTP metric is that the NVIPM documentation states that we do not differentiate between static and dynamic video noise; refer to the first line on page 17260 of [7] and to Equation 13 to understand how we treat the two differently. If V_{static} is single frame video metric$_{eye}$ is eye integration time in seconds, and F_{rate} is video frame rate per second, then

$$V_{dynamic} = V_{static}\sqrt{t_{eye}F_{rate}} \tag{13}$$

NVIPM literature states that TTP does not correct for the number of eyes used to view the display, and that is incorrect. TTP increases CTF by the square root of two when only one eye is used.

NVIPM literature states that TTP do not predict the effect of noise on the Contrast Threshold Function correctly, but [7] describes TTP predictions of empirical data collected by independent researchers, and the comparisons are accurate. Some of the comparisons are shown in Section 4.

NVIPM developers suggest that we have ignored retinal quantal noise. Not at all, we just do not agree with their treatment. The retinal photon noise is in the TTP noise model once and in the NVIPM model twice. That issue is addressed further in Sections 4 and 5. In [9], current Army researchers state that adding a noise term prior to the visual cortex hurts rather than helps model predictive accuracy. The authors find that result illogical and add the retina photon noise term regardless. That result is logical if one realizes that the photon noise is already included in the CTF threshold values.

The NVIPM documentation suggests that the TTP metric has trouble with modeling target images that subtend different angles at the observer's eye. The data they showed, however, was only a fraction of the experimental data, and the model curve was not generated using TTP. In

other words, they did not show the model to data comparison available in the TTP literature. TTP has accurately predicted PID for over forty target display angles ranging from 0.9 to 40 degrees, and all those instances are documented in journal literature published prior to NVIPM release.

NVIPM authors claim that their Fig. 21 shows the poor predictive accuracy of the old TTP noise model. There are three problems with presenting that figure and making that statement. First, if the two-hand, hand-held target set is diverse, the authors of [9] should use the RSS contrast with the TTP model. Second, the two-hand, hand-help target set is not diverse. True, that is a judgement based on just looking, but diverse target sets have visual discrimination cues that are Normally distributed in size and contrast; few target sets are diverse. Third, those authors have made statements in earlier conferences and conference papers that they use their own version of the TTP noise model, not the model in the TTP literature [7]. The legacy NVThermIP model incorporated into NVIPM is not the NVThermIP model and cannot be used to compare the two metrics.

4. Targeting Task Performance calculation and validation

The TTP IQM is calculated using either Eq. 14 or Eq. 15 where ξ is spatial frequency in imager angle space and ξ_{cut} and ξ_{low} define integration limits where CTF_{sys} is less than target RSS contrast (C_{TGT}). See [7] for the parameters needed to calculate CTF_{sys}. The CTF_{sys} algorithm requires calculating the imager transfer function and then applying the display modulation transfer function (MTF), eye MTF, and visual cortex filter MTF to the imager noise. Supplying those details here would not significantly contribute to the current discussion.

The second version of TTP is found in the later references. Both equations return the same metric values except when the displayed image is barely visible due to glare. CTF is

$$\phi = \int_{\xi_{low}}^{\xi_{cut}} \sqrt{\frac{C_{TGT}}{CTF_{sys}(\varepsilon)} d\varepsilon} \qquad 14$$

$$\phi = \int \sqrt{\delta\left(\frac{C_{TGT}}{CTF_{sys}(\varepsilon)}\right) \frac{C_{TGT}}{CTF_{sys}(\varepsilon)} d\varepsilon} \qquad 15$$

In these equations, ϕ is the TTP IQM value in imager angle space. The function δ of the ratio of C_{TGT} to CTF_{sys} is the probability that the sine wave at spatial frequency ξ is visible. The δ function is derived from CTF experiment data. A similar TTP value is calculated for vertical, and the geometric mean used to calculate the final ϕ. See [7] for the parameters and formulas needed to calculate CTF_{sys}.

The viewing scenario is foveating and fixating display details at mid-mesopic to photopic luminance. Even on dark nights, people do not set display luminance below 0.1 footlamberts (fL), and cones mediate vision down to about 3E-4 fL. We assume that the observer is fixating the target image and seeing the color of the display phosphor, and that describes cone vision. TTP is a cone model, and the only dependence on luminance enters TTP calculations via the observers CTF. To be clear, if the noise parameters in the TTP calculation changed based on either target signatures or imager type, that would be considered a fatal flaw in the IQM.

The authors of [9] state that the TTP does not distinguish static from dynamic video, and that is not the case. As stated in [7], a correction factor $\sqrt{(0.04 \text{ Frame rate})}$ is applied to the video noise. The 0.04 eye integration time is based on [18], and our understanding of the

psychophysical literature is that there is no evidence that cones change temporal behavior until they no longer strongly mediate vision at about 3E-4 fL.

Predicting the effect of noise on CTF uses the engineering model of the eye shown in Figure 20. As showed by that figure, the MTF of the display, eyeball, and cortical bandpass filters are applied to the camera signal prior to the root-of-the-sum-of-squares (RSS) of display and cortical noise. The figure shows the single point where noise is injected into the visual signal. Based on the experiments of Nagaraja [19] and others [20-22], the effect of noise can be explained by assuming the brain is taking the RSS of display noise and some internal eye noise. Further, for display luminance above the de-Vries Rose Law region and for foveated and fixated targets, the RSS of eye and display noises occurs in the visual cortex.

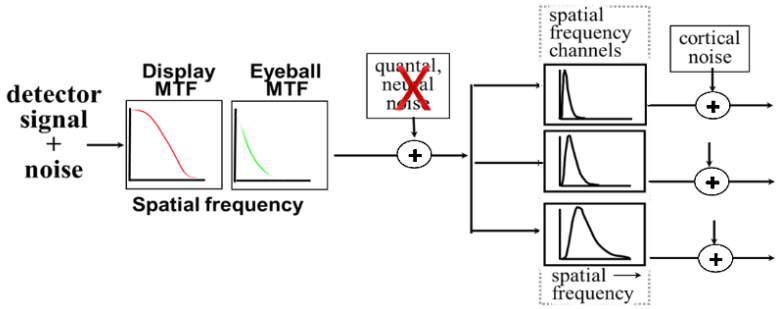

Figure 20. Engineering model showing blurring of image signal before RSS with cortical noise.

We do not include retinal photon noise. The TTP noise model modifies the Barten CTF, and that CTF was measured at the same display luminance. Therefore, the photon noise contributed to the naked-eye CTF and including it again would be a mistake.

The photon noise contributes to the subject's CTF, so the retinal quantal noise has already degraded the CTF used in our model. That is different than Barten's treatment, because he is predicting the CTF while we are modifying it. We start by using Barten's numerical approximation to generate CTF, and then we degrade that CTF based on the blur and noise characteristics of the camera and display.

The authors of [9] suggest that the TTP noise model does not predict their CTF data, but no examples of that failure are described, so addressing their concerns directly is not possible. In any event, their CTF data were not taken in the generally accepted manner, and it is not clear why the TTP noise model should predict their data.

However, the TTP noise model was compared to several examples of CTF in noise data found in the literature, and those comparisons support the accuracy of the TTP metric calculations [7]. Four examples are included from [7] and shown in Figures 21and 22. The Contrast Sensitivity Function (CSF) is the inverse of CTF. The TTP predictions of CTF_{sys} are not as good as those of Barten [23]. However, Barten modifies the naked eye CTF of each observer, while TTP modifies the Barten CTF representing a single nominal observer.

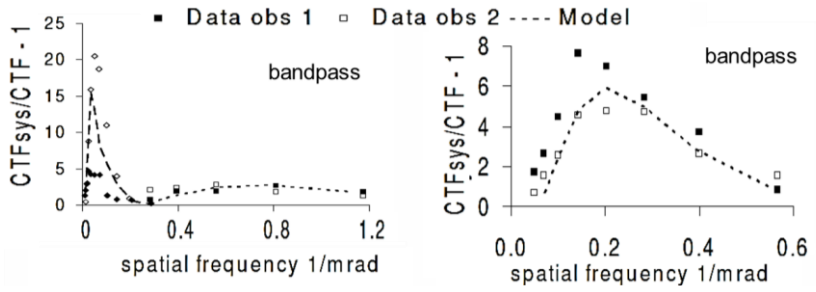

Figure 21. These plots compare TTP prediction of the effect of noise on naked-eye CTF [7].

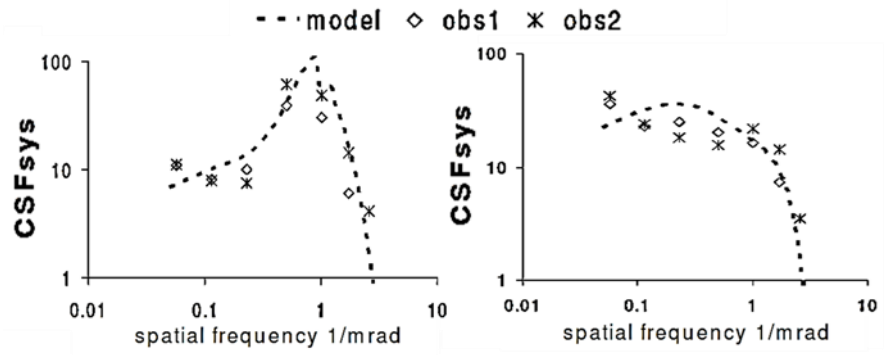

Figure 22. The plots compare CTF_{sys} to empirical data of independent researchers [7].

The first validation of the CTF model used data from a limiting light experiment where three experienced observers viewed Air Force 3-bar charts to establish limiting resolution of image intensifiers versus chart illumination [1-3]. The experiment used chart contrasts of 1.0 and 0.4. Data were collected with and without laser protective eyewear situated between the goggle and eye that reduced apparent luminance by a factor of ten. Light to the eye varied from as little as 3.4E-4 fL to 35 fL Chart illuminance varied from 2.88E-6 to 3.39E-3-foot candles, and that variation in illumination means that the image intensifier tubes were operated from noise limited to resolution limited conditions. Details on image intensifier tube data and other goggle parameters can be found in [1-3]. The plot in Figure 23 shows data on the abscissa and model predictions on the ordinate. The line represents perfect predictions.

The TTP calibration constant labeled α was derived from the image intensifier data, and α did not change when Barten's visual cortex spatial filters were introduced to predict the effect of colored noise on CTF. As previously stated, the α parameter does not change. However, in the reference literature, the value of α is different for the thermal model and reflective model because the units are different.

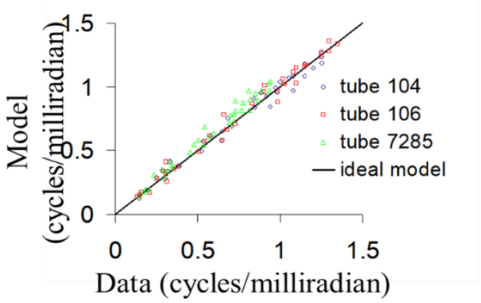

Figure 23 compares metric predictions to limiting light data.

In Figure 24, each PID is based on the average of 20 calls by an active Army aviation or tank gunner, tank commander, Tow operator, or other soldier with an appropriate military specialty. The low-contrast and high-resolution data points were displayed on a medical monitor with a special computer interface so that the target set had a minimum five-bit gray scale. Targets were displayed with four varied sizes from 0.9 to 4.1 degrees.

The PID data in Figure 24 were taken on two display types, one color and one high resolution monochrome with a special computer digital interface. There were twelve sizes of

target displayed to represent 6 ranges on each display. The data included four different blur shapes, five contrasts, four noise spectrums, and five target set contrasts.

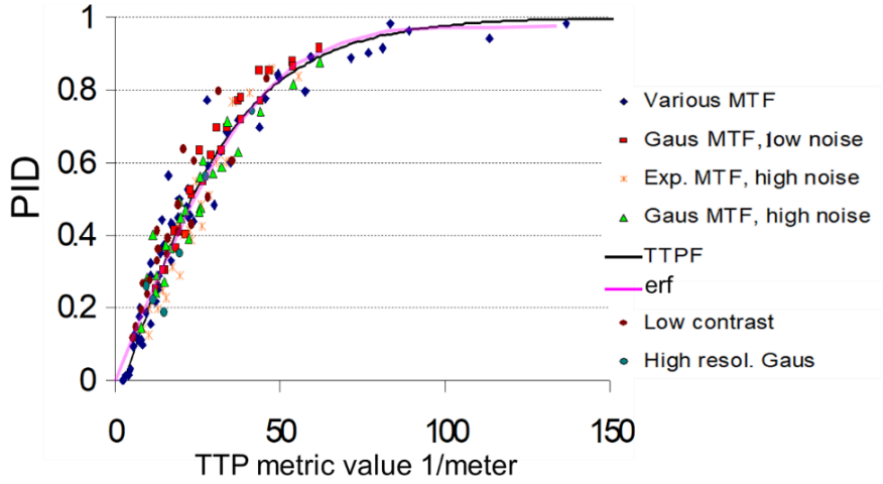

Figure 24. The target set was twelve aspects each of twelve tracked, tactical vehicles. TTPF is a curve fit to all data using the Logistics Function. The erf is now used to calculate PID given a metric value, and that is true regarless of target type. See [4].

The data in Fgure 24 are re-plotted in Figure 25 to show that the TTP metric does not rely on presenting the target with a specific size on the display. Targets were presented with twelve different sizes in the alaising experiments, and in that case the observers viewing distance was controlled. The target angular size on the display also changed twleve times during the field test to be described below. In other tests described below, the target angular size changed at least twenty times.

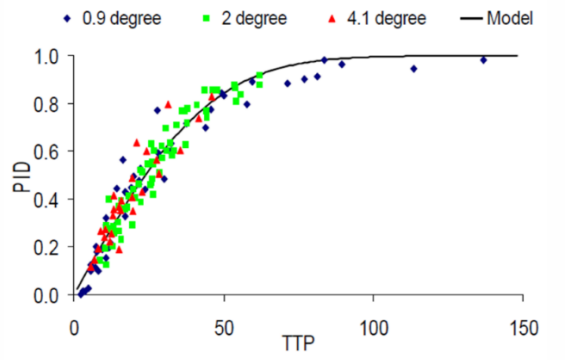

Figure 25 showing that PID predictive accuracy does not depend on the size of target on display.

Figure 26 shows predicted PID for various colored (that is, non-white) noise [4].

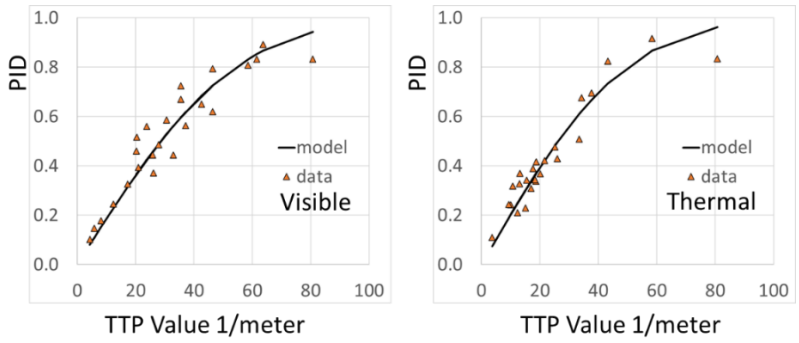

Figure 26 shows TTP PID predictions with several types of noise. See [4].

Figure 27 shows PID predictions of the characters in Figure 11 using the Fourier Transform of a rectangle as targt signature [5].

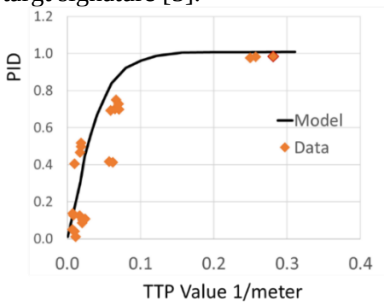

Figure 27. PID of characters in Figure 11 versus TTP metric value. See [5].

Figure 28 shows facial identification for two sets of ten faces. These results are for naked eye, two spotting scopes, and binoculars [6]. Figure 29 shows facial PID when the faces were digitally captured and displayed on a computer screen or printed on paper and then identified outdoors at range [6]. Facial PID predictions use the Fourier Transform of one face in the target set as target signature.

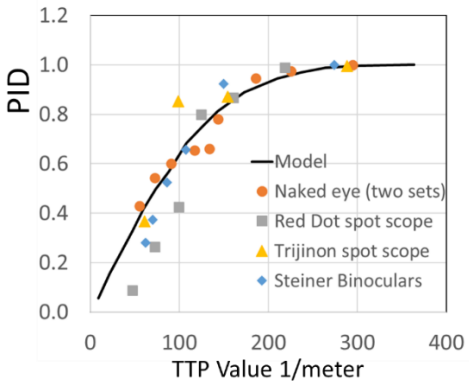

Figure 28. plots facial PID versus TTP value for faces viewed on a computer screen with and without noise and for faces printed on paper and identified at range outdoors. See [6].

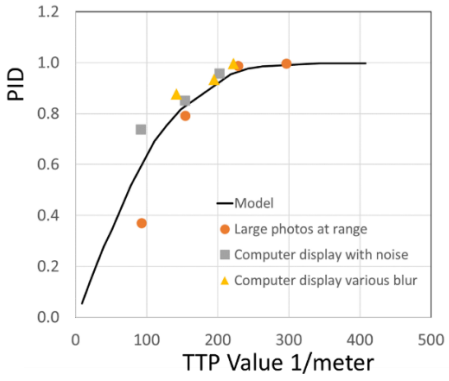

Figure 29. plots facial PID verus metric value when faces were displayed on comuter monitors or printed on paper. See [6].

Figure 30 shows shape discrimination for the three shapes in Figure 11 plus a square [5].

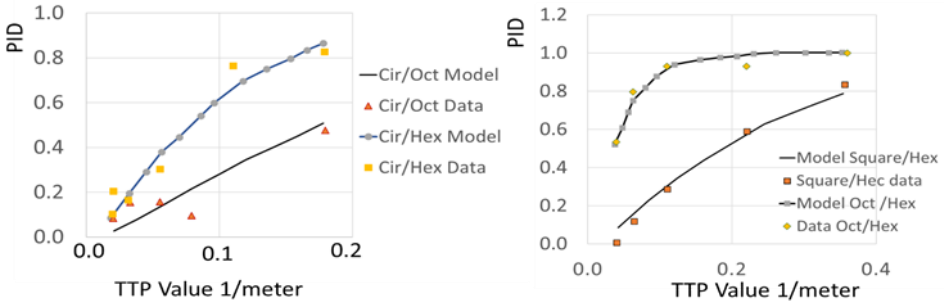

Figure 30. Shape discrimination predicted by TTP. See [5].

Two thermal imagers were field tested using eleven vehicle and eight aspect target set described in [3]. One imager was an F/1.1 long wave sensor with a 2.2 degree field of view. The second was a 2nd Generation F/3.6 long wave imager with a 3.1 degree field of view. The PID verus NVThermIP model predictions are shown in Figure 31. Again, the NVThermIP model incorporated into NVIPM is not NVThermIP.

The TTP metric also predicts the effect of environment on the utility of pilotage imagers. In this case, no ϕ84 is needed because we are not comparing PID performance but rather comparing the utility of different types of imagers to do a visual task. See Table 1 for a summary of TTP metric value predictions for the imagers and aviator assessments of imager utility [24].

ANVIS are Aviator's Night Vision Imaging System consisting of two Third Generation image intensifier oculars. First Generation thermal is 180 element Common Module thermal technology. Second Generation thermal is time-delay-and-integrate thermal technology. Both first and second generation thermal are cryogenically cooled, long wave imagers. Both thermal systems are head-tracked. The pilot assessment is based on two surveys of over one hundred experienced aviators taken ten years apart at the end of the 20th Century. We cite these surveys because they represent field-feedback of the efficacy of the TTP metric.

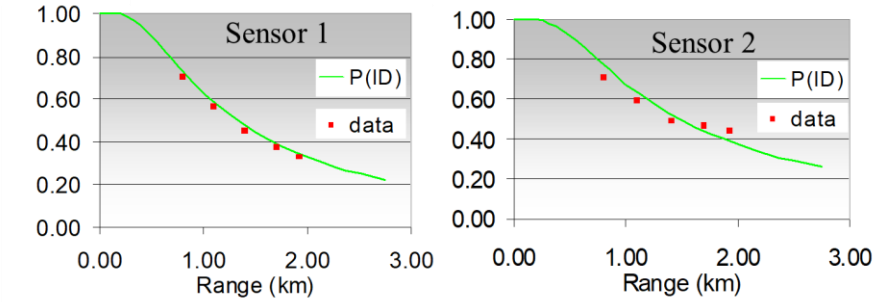

Figure 31 shows TTP PID from a field test of two thermal imagers. See [3].

Table 1. Aviator assessment of imager performance.

Sensor	Condition	TTP	Pilot assessment
ANVIS	overcast starlight	1.05	Poor
ANVIS	starlight	3.3	Fair
ANVIS	Quarter Moon	5.8	Good
1st Gen thermal	0.1 K scene	1.0	Poor
1st Gen thermal	1.0 K scene	2.6	fair
1st Gen thermal	5.0 K scene	3.1	fair
2nd Gen thermal	0.1 K	2.4	fair
2nd Gen thermal	1.0 K	5.9	Good
2nd Gen thermal	5.0 K	6.7	Good

5. Night Vision Integrated Performance Model validation

The subject of [9] is a re-vamping of the TTP noise model because of unspecified and undocumented problems with that model. The authors of [9] reject their calibration of the noise model using their CTF data. If they had not, the CTF experiment design described in [9] would receive more discussion here, but the authors themselves dismissed the results.

A authors describe a two-hand, hand-held object experiment and use the results to calibrate the noise model. However, in doing that, their data are used for calibration, and they change γ, β, and target contrast to fit PID predictions to data. In other words, that experiment is model calibration and not validation.

The Army has a data base with tactical vehicles, faces, characters, and shapes, but NVIPM is validated against only those tactical vehicle experiments with target contrast of 0.21 and display target size 4.1 degrees at the eye. That data from experiments 13, 19, and 20 are compared to NVIPM predictions in Figure 33. Note that Figures 25 and 26 are PID with image display angle from 0.9 to 4.1 degrees, whereas Figure 33 compares NVIPM predictions to images displayed with a 4.1-degree angle.

NVIPM predictions of face, character, and shape PID have not been published.

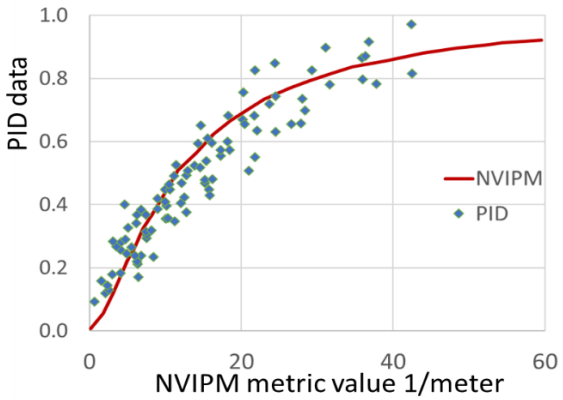

Figure 32 shows NVIPM predictions of PID data from experiments 13, 19, and 20 [4]. In these experiments, target RSS contrast is 0.21 and target size on the display is 4.1 degrees. See [9] for original NVIPM versus PID plot.

The authors of [9] incorporated a two-parameter noise model into NVIPM with γ equal to 240 and β equal to 4.3. The parameter γ serves the purpose of α, and β is added to model retinal photon noise. Those numbers are based on the two-hand, hand-help PID experiment. They note, however, that the CTF data calibration best fits for those two parameters is 330 and 4.9, respectively. They also note two features of the model that they consider illogical. The first is that the photon noise best fit works best if the photon noise is inserted after the visual cortex spatial filters. The second anomaly is that, for high noise levels, one-eye predicted performance is better than two-eye. These anomalies are accepted, however, because the new noise model works better than the old TTP noise model.

6. Conclusions

As can be seen by examining Eq. 1, CTF_{sys} is a critical factor in the calculation of TTP values. The original TTP modeled imagers in the frequency domain and compared the threshold amplitude of the eye at each spatial frequency to the range dependent Fourier Transform of the targets. Psychophysical research suggested that the effect of imager noise on the eye's thresholds should be modeled by RSS of display and eye noise in the visual cortex, and that model was successfully validated by comparing TTP CTF_{sys} with the CTF_{sys} data measured by independent researchers.

However, the robustness of the TTP model can only be proven by comparing TTP PID predictions with experiment data. In support of the TTP model, seventy-six experiments and field tests were performed, and the results published in journals and books. The experiments assessed both thermal and visible targets, targets as complex as tactical vehicles and faces and as simple as characters and shapes.

The TTP metric predicts image intensifier bar thresholds for image intensifier operation from overcast starlight to moonlight. Also, the bar thresholds tested luminance at the eye from 3E-4 to 35 fL. Some TTP experiments assessed TTP predictions when the image was limited by colored (non-white) noise. Also, TTP predictions use the erf function to derive PID from metric values, meaning that one of the two empirical fits used by most IQM modelers has been surrendered. The original TTP metric also predicts the experience of Army aviators using night vision pilotage aids.

The IQM in NVIPM has been tested against only three of the seventy-six experiments and field tests, and those PID predictions are scattered when compared to TTP PID predictions of

all the experiments. No other validation of the revised TTP incorporated into NVIPM has been offered.

The reason that the TTP metric was changed, substituting the target angle for display angel, has never been explained. Several of Barten's Square Root Integral (SQRI) metric papers are listed in the references of [9], but Barten's SQRI is not used in NVIPM. In any event, it is the TTP validation that supports the TTP metric and not Barten's research.

We understand that Barten's interest was in predicting the effect of small display size when viewed in a dark room. We understand why Barten used Carlson's CTF data when developing his CTF numerical algorithm. We use Barten's CTF algorithm and his numerical approximation to the visual cortex spatial filters in the original TTP metric and appreciate the quality of his work.

NVIPM does not use Barten's SQRI and depends for credibility on the validation of the original TTP metric. If the current Army modelers have an idea for a new metric or wish to incorporate Barten's SQRI, then they have that authority.

Instead, in both conference papers and journal articles, the NVIPM modelers have discredited the TTP model. Although the authors of NVIPM have offered no evidence of the failings of the original TTP model, their approach of continual discrediting that model has been successful. The Army is the authority on these models, and technical support for their statements is apparently unnecessary.

We understand that the current modelers represent Army authority, and we cannot sway the Army to examine their current approach. We can, however, ask engineers and scientists interested in these models who might review future articles to question statements that discredit other researchers without saying why. The TTP model has been discredited without offering any examples to prove inadequacy.

We also request future reviewers to consider the Optical Engineering paper [9]. We do not conclude that their CTF_{sys} model is not validated, they do. We do not conclude that their noise model has two major flaws, they do. We do not conclude that their new noise model needs correction, they do. In the last paragraph of their conclusions, however, they state without explanation that the new model is better than the validated TTP model and that NVIPM has been thoroughly validated.

It took years and commitment and funds to gather the TTP validation data and publish several conference papers, six journal papers, and four books to document TTP validation. The Army has all that experiment data, and validation of NVIPM should not be a problem. Instead, the Army continues to use discrediting of the original TTP as a reason for changing their target acquisition model. Simultaneously, journal articles are appearing crediting NVIPM with TTP validation.

Funding

None

Disclosures

The author has both a personal and professional interest in defending his published research.

Data availability

Data underlying the results presented in this paper is held by the Army's Futures Command at Fort Belvoir, Virginia.

References

1. R. VOLLMERHAUSEN, "INCORPORATING DISPLAY LIMITATIONS INTO NIGHT VISION PERFORMANCE MODELS," IN IRIS PASSIVE SENSORS, PROCEEDINGS OF THE INFRARED INFORMATION SYMPOSIUM, ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN, VOL. 2, PP. 11–31. 13, (1995).

2. ELECTRO-OPTICAL IMAGING: SYSTEM PERFORMANCE AND MODELING, LUCIEN M. BIBERMAN, ED., CH 12, "MODELING THE PERFORMANCE OF IMAGING SENSORS," RICHARD VOLLMERHAUSEN, SPIE PRESS, BELLINGHAM (2000).

3. R. H. VOLLMERHAUSEN, E. JACOBS, J. HIXSON, AND M. FRIEDMAN, "THE TARGETING TASK PERFORMANCE (TTP) METRIC: A NEW MODEL FOR PREDICTING TARGET ACQUISITION PERFORMANCE," TECH. REP. (DTICAMSEL-NV-TR-230, U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND, COMMUNICATIONS AND ELECTRONICS RESEARCH, DEVELOPMENT AND ENGINEERING CENTER, NIGHT VISION AND ELECTRONIC SENSORS DIRECTORATE, 10221 BURBECK RD., FT. BELVOIR, VA 22060-5806 (2005).

4. R. H. VOLLMERHAUSEN, E. JACOBS, AND R. G. DRIGGERS, "NEW METRIC FOR PREDICTING TARGET ACQUISITION PERFORMANCE," OPT. ENG. (BELLINGHAM) 43, 2806–2818 (2004).

5. R. VOLLMERHAUSEN AND A. L. ROBINSON, "MODELING TARGET ACQUISITION TASKS ASSOCIATED WITH SECURITY AND SURVEILLANCE," APPL. OPT. 46, 4209–4221 (2007).

6. RH VOLLMERHAUSEN, S MOYER, K KRAPELS, RG DRIGGERS, JG HIXSON, "PREDICTING THE PROBABILITY OF FACIAL IDENTIFICATION USING A SPECIFIC OBJECT MODEL," APPLIED OPTICS 47 (6), 751-759 (2007).

7. RICHARD H. VOLLMERHAUSEN, "REPRESENTING THE OBSERVER IN ELECTRO-OPTICAL TARGET ACQUISITION MODELS," OPTICS EXPRESS 17 (20), 17253-17268 (2009)

8. BRADLEY L. PREECE, JEFFREY T. OLSON, JOSEPH P. REYNOLDS, JONATHAN D. FANNING, "IMPROVED NOISE MODEL FOR THE US ARMY SENSOR PERFORMANCE METRIC," PROC. OF SPIE VOL. 8014 801406 (2011)

9. BRADLEY L. PREECE, JEFFREY T. OLSON, JOSEPH P. REYNOLDS, JONATHAN D. FANNING, AND DAVID P. HAEFNER, "HUMAN VISION NOISE MODEL VALIDATION FOR THE U.S. ARMY SENSOR PERFORMANCE METRIC," OPTICAL ENGINEERING 53(6), 061712 (2014)

10. ROBERT SHORT, DUKE LITTLEJOHN, JOHN BAILEY, AND RONALD DRIGGERS, "INFRARED SENSOR PERFORMANCE WITH BOOST AND RESTORATION FILTERING," APPLIED OPTICS, VOL. 60, ISSUE 3, PP. 571-579 (2021)

11 RONALD DRIGGERS, GLENN GORANSON, STEVE BUTRIMAS, GERALD HOLST AND ORGES FURXHI, 'SIMPLE TARGET ACQUISITION MODEL BASED ON FA/D," OPTICAL ENGINEERING, VOL. 60(2), 023104-1 (2021)

12. RICHARD H. VOLLMERHAUSEN, "NIGHT VISION INTEGRATED PERFORMANCE MODEL: IMPACT OF A RECENT CHANGE ON THE MODEL'S PREDICTIVE ACCURACY," OPTICS EXPRESS, VOL. 24, NO. 21, 23654-23666 (2016)

13. P. G. J. BARTEN, "FORMULA FOR THE CONTRAST SENSITIVITY OF THE HUMAN EYE," PROC. SPIE 5294, 231–238 (2004) (PAPER AVAILABLE ON THE WEB AT HTTP://WWW.SPIE.ORG).

14. R. J. BEATON, AND W. W. FARLEY, "COMPARATIVE STUDY OF THE MTFA, ICS, AND SQRI IMAGE QUALITY METRICS FOR VISUAL DISPLAY SYSTEMS," ARMSTRONG LAB., AIR FORCE SYSTEMS COMMAND, WRIGHT-PATTERSON AFB, OH, REPORT AL, TR-1992–0001, DTIC ADA252116, (1991)

15. RICHARD H. VOLLMERHAUSEN, RONALD G. DRIGGERS, AND DAVID L. WILSON, "PREDICTING RANGE PERFORMANCE OF SAMPLED IMAGERS BY TREATING ALIASED SIGNAL AS TARGET-DEPENDENT NOISE," J. OPT. SOC. AM. A/VOL. 25, NO. 8 (2000)

16. RICHARD H. VOLLMERHAUSEN, RONALD G. DRIGGERS, AND DAVID L. WILSON, "PREDICTING RANGE PERFORMANCE OF SAMPLED IMAGERS BY TREATING ALIASED SIGNAL AS TARGET-DEPENDENT NOISE: ERRATUM," J. OPT. SOC. AM. A/VOL. 25, NO. 8 (2000)

17. RG DRIGGERS, R VOLLMERHAUSEN, B O'KANE, "EQUIVALENT BLUR AS A FUNCTION OF SPURIOUS RESPONSE OF A SAMPLED IMAGING SYSTEM: APPLICATION TO CHARACTER RECOGNITION," J. OPT. SOC. AM. A, VOL. 16 (5) ,1026-1033 (1999)

18. R. A. MOSES AND W. M. HART, "THE TEMPORAL RESPONSIVENESS OF VISION,' IN ADLER'S PHYSIOLOGY OF THE EYE: CLINICAL," APPLICATION, MOSBY (1987)

19. N.S. NAGARAJA, "EFFECT OF LUMINANCE NOISE ON CONTRAST THRESHOLDS, J. OPT. SOC. AM. 54(7), 950-955 (1964)

20. D. G. PELLI, ''WHY USE NOISE?'' J. OPT. SOC. AM. A 163, 647–653 (1999)

21. M. RAGHAVAN, ''SOURCES OF VISUAL NOISE,'' PHD DISSERTATION, SYRACUSE, NY, (1987)

22. H. DAVSON, PHYSIOLOGY OF THE EYE, 5TH ED., MACMILLAN, LONDON (1990)

23. PETER G. . BARTEN, CONTRAST SENSITIVITY OF THE HUMAN EYE AND ITS EFFECTS ON IMAGE QUALITY, [PHD THESIS 1 (RESEARCH TU/E / GRADUATION TU/E), INSTITUTE FOR PERCEPTION RESEARCH, EINDHOVEN]. TECHNISCHE UNIVERSITEIT EINDHOVEN. HTTPS://DOI.ORG/10.6100/IR523072 (2005)

24. RH VOLLMERHAUSEN, T BUI, "USING A TARGETING METRIC TO PREDICT THE UTILITY OF AN EO IMAGER AS A PILOTAGE AID," INFRARED IMAGING SYSTEMS: DESIGN, ANALYSIS, MODELING, AND TESTING XVII 6207, 15 (2006)