Measuring Situational Awareness with the "Ideal Observer"[1]

Marc Green[2,3], J. Vernon Odom[2], and J. Terry Yates[4]

[2] University of West Virginia Medical School
[3] York University
[4] Brooks Air Force Base

Background

Situational awareness requires the human operator to quickly detect, integrate and interpret data gathered from the environment. In many real-world conditions, situational awareness is hampered by two factors. First, the data may be spread throughout the visual field. Operators are then limited by attention, memory and ability to combine data seen in the same or different formats. Second, the data are frequently noisy.

We have been investigated application of the "Ideal Observer" model as a means for measuring situational awareness under these conditions. According to the model, task performance is limited only by three factors, "external noise," noise in the environmental data, "internal noise," noise inside the observer and "efficiency," the ability to sample environmental information. The model makes precise predictions about the relationships among these 3 variables.

There are many advantages in using an Ideal Observer model to measure situational awareness. The Ideal observer provides a measure of statistically optimal performance, so that situational awareness can be evaluated against an absolute rather than a relative standard. Moreover, the Ideal Observer model allows independent analysis of both low level information processing and high level cognitive decision-making within the same framework. Further, the Ideal Observer permits direct comparison of optimal behavior across different environments - it is possible to compare situational awareness with vastly different displays and tasks. There are several other benefits of the Ideal Observer, but these will not be obvious until the model is explained below.

We are using the Ideal Observer in developing a battery of tests which evaluate the ability to make decisions when confronted with noisy data. The displays contain information, perturbed by Gaussian noise, which is spread throughout the visual field. The observer must integrate the noisy data and then make a decision. We are investigating factors which minimize observer and decision noise and maximize efficiency.

In this paper, we will describe a preliminary application of the Ideal Observer to a task from our test battery, dot estimation. We have chosen this task both because the application of the Ideal Observer model is simple and direct and because several studies (Endsley and Bolstad, 1994; T. Caretta, cited in Endsley and Bolstad, 1944) have shown that dot estimation performance correlates well with other measures of situational awareness.

The derivation of an Ideal Observer for dot estimation has been described elsewhere (Barlow, 1978; Burgess and Barlow, 1983, but, for convenience, we will provide a brief review and highlight the significance for measurement of situational awareness.

Derivation of the Ideal Observer for Dot Estimation

In Signal Detection theory, performance is limited only by noise. Given this view and a statistical representation of environmental information, it is possible to construct a model, the Ideal Observer, which has no internal noise and uses all available signal information. Such a hypothetical observer performs optimally in the sense that it is limited only by the information in the environment.

In the dot estimation task, the observer sees two boxes, one contains N dots and the other $N + \Delta N$ (Figure 2). The task is to say which box has the ΔN. Optimal performance can be expressed as:

$$d' = \frac{\Delta N}{\sigma_N}$$

where d' is the measure of detectability, ΔN is the difference in dot number between target and background (noise) and σ_N is the noise standard deviation – a number of extra dots which have been added to or subtracted from each box.

Real observers, however, seldom achieve ideal performance. Previous work has suggested two general factors to account for the suboptimality of human performance. The first factor producing suboptimal behavior is internal noise. Although the Ideal Observer is presumably limited only by external noise, real detection devices also have *internal* noise which decreases performance. In the dot estimation task, this can be modelled as:

$$d' = \frac{\Delta_N}{\sqrt{\sigma_N^2 + \sigma_N^2}}$$

where σ^2_N refers to observer's internal noise. (Of course, this assumes independence of internal and external noise.) In other words, detection is a joint function of external and internal noise, which increases the denominator and decreases performance.

The second factor affecting performance is reduced information gathering efficiency. The Ideal Observer is a Bayesian classifier which determines p(hypothesis|data) for each potential signal. It computes a likelihood ratio for any two signals from the ratio of their probabilities. A decision rule uses the likelihood ratio or its monotonic transform to specify when the observer should say "yes" or "no." The response transition point, the criterion, depends on what aspect of performance the observer wishes to optimize). For most applications, this is assumed to be maximum percent correct.

The main problem, as in all Bayesian classification tasks, is to derive an estimate of p(data|hypothesis). The Ideal Observer obtains p(data|hypothesis) by sampling the data in the environment and comparing them to an internal model. The importance of proper sampling is apparent from Barlow's (1978) definition of sampling efficiency, F, as:

$$\frac{\textit{sample size required by the ideal observer}}{\textit{sample size required by the actual observer}}$$

where F=1 is optimal efficiency and lower values represent suboptimal sampling. The less efficient observer will require more information samples, meaning a longer sampling period, and therefore longer reaction time, or better information. This notion of efficiency reflects how well an observer uses external information. The observer samples visual input and tries to match this information to an internal model, much like a template matching procedure. In the broad application to situational awareness, the templates are presumably a patterns of information which reflect possible states of the aircraft (or any other dynamic system). Efficiency reflects the observers ability to use external information to detect the appropriate "template."

The F value has two practical uses. One is that it can be used to discriminate people who have high and low ability at accepting and integrating visual information. Another is to use F as a tool for designing information displays. That is. when comparing different display formats, high F also suggests highly efficient information display.

F can be estimated directly from sensitivity as:

$$F = (d'_e / d'_i)^2$$

where d'_e is experimentally obtained sensitivity and d'_i is the sensitivity of an Ideal Observer who uses all available information. Finally, by substitution and solving for a d'_e value of 1, the final model for a real observer, taking into account efficiency and internal noise, becomes:

$$\Delta N_T^2 = (1 / F)\,(\sigma_N^2 + \sigma_N^2)$$

where ΔN_T^2 is the square of the difference in dot number required to produce a d'_e of one - 76% correct in a two alternate forced-choice test. (If this step was too large, see Burgess and Barlow, 1983 for more details.) In English, this says that performance is a joint function of two additive factors, internal and external noise, and a multiplicative factor, efficiency[1].

One way to visualize the relationship between internal noise and efficiency is to plot data from a test which measures the square of the of the threshold (dot difference needed to achieve a performance of d'=1) as a function of external noise variance (dots added/subtracted from the display). Figure 1 shows hypothetical results for such an experiment. Internal noise is a constant factor which alters the X intercept (not shown) while efficiency is a multiplicative term which alters slope. The Ideal Observer, A, who has an intercept of zero (is noiseless) and a slope of one (uses all information). C shows a family of observer with internal noise (change in intercept) but efficiency close to the optimum (slope of 1). The internal noise produces a horizontal shift which can be quantified by the negative of the curve's intersection with the abscissa. Slope increases reflect suboptimal efficiency. Curve B shows an observer with no internal noise but a lowered efficiency, which is quantified as the reciprocal of the slope. Of course, an observer could exhibit both lowered efficiency and internal noise.

There are two important points highlighted by this graph. First, the Ideal Observer provides two different measures, efficiency and internal noise level, for each test. At 0, or any single external noise level, a given level of performance could be caused by different combinations of the two factors. Two observers could perform equally on the task, yet one might have low efficiency and the other high internal noise. Situational awareness might correlate with only one of these two factors. To distinguish the two factors, observers must be tested at a minimum of two noise levels. Once done, however, the Ideal Observer models allows a more precise and detailed analysis of individual abilities because it reveals not just overall performance but also individual differences in the factors underlying task performance. Although situational awareness correlates moderately with dot estimation, for example, there could be a higher correlation with observer efficiency and a lower correlation with internal noise. Looking only at overall performance would then reduce the correlation.

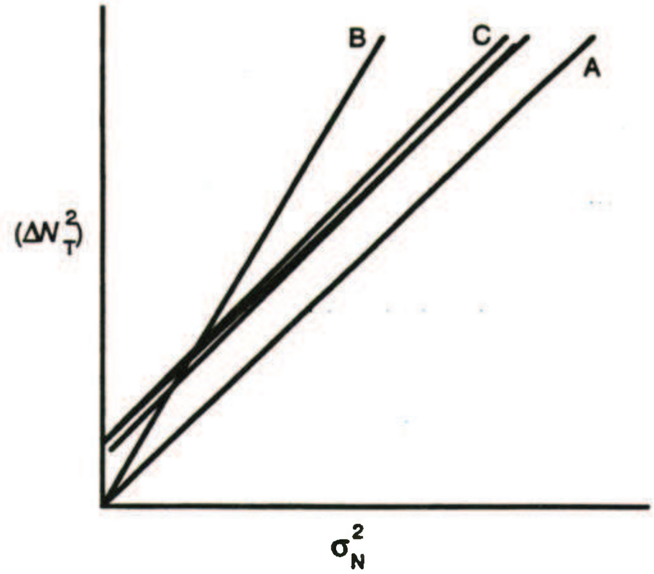

Figure 1. Schematic data from Ideal Observer test.

Second, note that in the low external noise conditions, high and low efficiency observers will be less readily differentiated. At high external noise levels, however, low and high efficiency observers are more easily distinguished. Addition of noise magnifies individual differences in some observers, making it easier to discern those with high situational awareness.

The Dot Estimation Test

Method

In our version of the task, the viewer sees two red rectangular boxes (Figure 2) containing differing numbers of black dots on a grey background. Following each 667 msec exposure, the observer responded by pressing the left or right mouse button to signal whether the left or right box had more dots. Observers were tested in a series of two-alternative spatial, forced-choice trials in which task difficulty, the difference in the number of dots in the two rectangles, was modulated by a tracking rule. The standard number of dots (N) was 100. The dot difference (ΔN) between the boxes was then perturbed by adding "noise" (σ_N), i. e., increasing/decreasing dots from each box. The number of noise nots was randomly chosen from a Gaussian distribution with a mean of 0 and a variance of 0, 25, 100 or 400 dots.

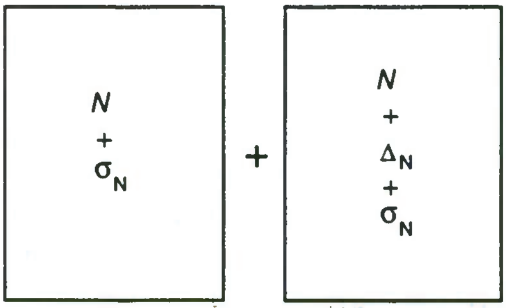

Figure 2. The right hand box contains the signal Δ_N

Results

Table 1 and Figure 3 show the results for three observers. Each observer shows both the presence of internal noise and a slope greater than one indicating suboptimal efficiency. Different observers, however, exhibited different degrees of noise and efficiency. Observer 3 had the lowest efficiency, for example, while observer 2 had highest efficiency but greatest internal noise.

Figure 3 shows the results plotted as a function of the squared dot threshold variance. and external noise The regression lines through the points were fit by least-squares method. For comparison, we also show the predicted performance of the ideal observer. The points cluster close to the regression line, showing that the data are in good agreement with the model for observers with internal noise and suboptimal efficiency.

Table 1.

Observer	X Intercept	Slope
1	290.5	1.77
2	343.7	1.82
3	197.5	2.12
Mean	273.1	1.90

Conclusion

Our goal has been to both describe the advantages of Ideal Observer analysis and to demonstrate its application to a test which is known to correlate with situational awareness. By decomposing test scores into subcomponents of efficiency and internal noise, Ideal Observer models may provide a finer grained analysis of individual skills and capacities and reveal more fundamental components of high situational awareness ability. Moreover, by stressing visual performance with noise, individual differences in ability should become more apparent.

We do not claim that the particular task demonstrated here, dot estimation, would alone be sufficient to measure situational awareness. A complete test battery would be comprised of several perceptual and attentional tests. The ideal Observer, however, can be applied to virtually any test where it is possible to create a statistical description of the task and to detection, reaction time and even physiologically derived dependent measures. We have already applied the Ideal observer to a large battery of behavioral and physiological tests which are being used to evaluate pilot visual abilities.

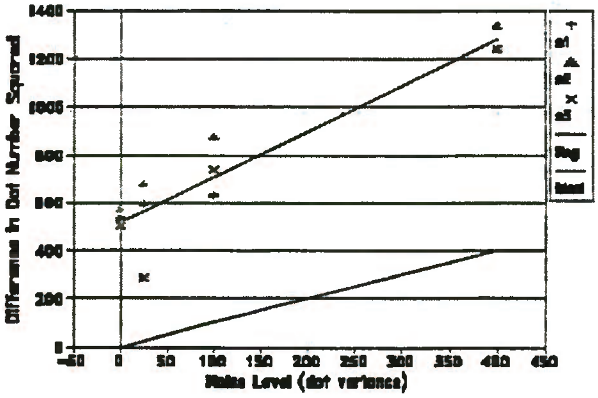

Figure 3.

References

Barlow, H. (1978) The efficiency of detecting changes of density of random dot patterns. *Vision Research, 18*, 637-650.

Burgess, A. and Barlow, H. (1983) The precision of numerosity discrimination in arrays of random dots. *Vision Research, 23*, 811-820

Endsley, M.R. and Bolstad, C.A. (1994). Individual differences in pilot situation awareness. *International Journal of Aviation Psychology, 4*, 241- 264.

Hartman, B and Secrest, G (1991). Situational awareness is more than exceptional vision. *Aviation, Space and Environmental Medicine*, November, 1094-1089

Secrest, G. and Hartman, B. (1993). Situational awareness: The trainability of near- threshold information acquisition dimension. *Aviation, Space and Environmental Medicine*, October, 885-889.

Footnotes

[1]Funded by contract F41624-92-DS-4001 DO-0017 from the U. S. Air Force Human Systems Center.

[5] There is another multiplicative parameter, decision noise, which we have not explicitly modelled. Because it would require far more involved testing procedures, we have chosen, as most researchers do, to collapse it with efficiency in the F parameter.