2009

NT-SEEV: A Model of Attention Capture and Noticing on the Flight Deck

Dr. Christopher Wickens

Angelia Sebok

Julia Bzostek

Kelly Steelman-Allen

Jason McCarley

See next page for additional authors

Authors

Dr. Christopher Wickens, Angelia Sebok, Julia Bzostek, Kelly Steelman-Allen, Jason McCarley, and Nadine Sarter

NT-SEEV: A model of attention capture and noticing on the Flight Deck

Dr. Christopher Wickens, Angelia Sebok, Julie Bzostek
Alion Science and Technology
Boulder, CO
Kelly Steelman-Allen, Jason McCarley,
University of Illinois Human Factors Division, Champaign, IL
Nadine Sarter
University of Michigan Dept of IOR. Ann Arbor, MI

N-SEEV is a model that predicts the noticeability of events that occur in the context of routine task-driven scanning across large scale visual environments such as the flight deck or ATC work station. The model is an extension of the SEEV (salience, effort, expectancy, value) model, incorporating the influence of attentional set and allowing the possibility of a dynamic environment. The model was validated against two empirical data sets. In a study of pilot scanning across a high fidelity automated 747 cockpit, the SEEV component of the model predicted the distribution of attention with a correlation of 0.85. In a lower fidelity study of pilot noticing of the onset of critical cockpit events (flight mode annunciators) the model predicted differences in noticing time and accuracy with correlations (across conditions) above 0.95. Other properties of the model are described

Failures of pilots or controllers to notice critical events, such as low altitude warnings, have been responsible for serious accidents with numerous fatalities (Wiener, 1977; NTSB, 2006). To be effective, visual and aural alerts must capture attention (SAE- ARP, 2007). Guidelines for the design of alerts typically emphasize the importance of display characteristics such as flashing, color, brightness, etc. But even a signal that is highly salient, however, can often go unnoticed, as illustrated by the phenomena of change blindness (Simons & Levin, 1997; Rensink, 2002; Stelzer & Wickens, 2006: Martens, 2007) and inattentional blindness (Mack & Rock, 1998;), effects which demonstrate the **failure** of events (e.g., changes) in the environment to capture attention. We can thus characterize human performance, as reflected in **noticing time (NT)** and the **probability of noticing** (or its inverse, **miss rate**), as falling along a continuum between rapid attentional capture and inattentional blindness. Effective cockpit design must be supported by models that can predict the noticeability of visual alerts, and therefore a assure that noticeability increases with the importance of the alert.

Such noticeability is governed by two classes of factors: bottom up and top-down. Bottom up factors are defined primarily by event **salience**. For example repeated flashes are more salient than single onsets, and these more salient than single offsets. Foveal events are more salient than peripheral ones. Top down factors can be subdivided into those related to expectancy and value. In addition, the effectiveness of both bottom up and top-down factors is modulated by workload, with high workload diminishing noticeability vis attentional tunneling (Wickens & Alexander, 2009). While all of these factors have been clearly identified to effect noticeability in isolation, there have been few studies examining their influence in combination, and there appears to be no valid computational model, that can predict their interaction, in a manner that might be useful for flight deck certification. The purpose of the research we describe here is to validate a computational model of noticing, N-SEEV, against two sets of existing data, and to demonstrate how the model can be applied to certification of visual warnings in safety-critical environments such as the cockpit or the ATC work station. We note that the model described and validated here has been subsequently used to **predict** the detection rate of off-nominal or unexpected events, in highly realistic flight simulations, and as projected to occur in NextGen technology and procedures, as described in detail in Gore et al (2009; see also Hooey et al, 2009)

Method: The Model

The N-SEEV model, described in more detail in Steelman-Allen et al (submitted) has two components, N and SEEV. The SEEV component describes the steady state allocation of visual attention (e.g., scan path) across any large scale environment such as the cockpit (Wickens McCarley, et al., 2008; Wickens & McCarley, 2008). Attention is directed to <u>**S**</u>**alient** events and locations, is inhibited in its movement by the <u>**E**</u>**ffort** needed to undertake long scans or head movements, and is directed to areas where there is <u>**E**</u>**xpected** to be information (e.g., high bandwidth areas), particularly if that information is **Valuable.** The first letter of each of the above terms, defines SEEV. As examples, a red flashing alert will have a high salience. The effort required for a pilot to scan an overhead panel will be greater than scanning a HUD indicator. The expectancy of change on the ADI in turbulence will be higher than the change of a heading indicator, or, in particular, a fuel gauge. The value of sampling displays that support aviating (e.g., angle of attack, pitch, speed above stall speed) is greater than those that support communication (e.g, radio frequencies).

The four elements of SEEV are additive, although in the current application, Salience and Effort are lumped into a single parameter. The model drives gaze around a simulated environment in a probabilistic (i.e., Monte-Carlo) fashion, such that the probability of the scan moving from one location to the next, is proportional to its relative "attentional attractiveness" compared to all other areas, as this attractiveness is defined by SEEV. This model of steady state scanning then predicts a distribution of locations of gaze at the moment when the to-be-noticed-event (TBNE) occurs, and therefore, it predicts distribution of the degree of *retinal eccentricity* of the TBNE. Such eccentricity is shown to be a powerful modulator of noticeablity. Our meta-analytic review of the small amount of existing human factors literature on the topic revealed the eccentricity effect on miss rate to be approximately 8% per 10 degrees, although this eccentricity cost is greatly modulated by clutter and expectancy. The N component of N-SEEV then predicts the time to notice the TBNE as a function of eccentricity, expectancy & value of the TBNE, and its salience, as the latter is quantified by a model derived from Itti & Koch, (2000).

Results: Model Validation

A set of data from three experiments was first employed to fit the parameters of N-SEEV, and then the model was validated against the scanning and noticing time data from two experiments. In the parameter fitting phase, we examined a set of scanning data from single-pilot general aviation simulations in which self-separation responsibilities were supported by a cockpit display of traffic information (CDTI) (Wickens, et al., 2002); and also augmented by a data link display (Helleberg & Wickens, 2003; Wickens, Goh, et al., 2003). Using these data we assured that the current version of SEEV, a Monte-Carlo simulation, predicted pilot scanning with the same precision as the analytic (equation) version of the SEEV model used in those three studies. Our results indicated a favorable fit, with the current Monte-Carlo model predicting the distribution of fixations across areas of interest in the three studies, with correlations of 0.93, 0.96, and 0.94.

Following parameter fitting, the first formal model validation was carried out against the pilot scanning data collected by Sarter, Mumaw & Wickens (2007), describing the scan data of 21 commercial aircraft pilots as they flew realistic missions using the flight management system in a high fidelity Boeing 747 simulator, across phases of take-off, departure, cruise, descent and final approach.. We applied the model to the cockpit image shown in figure 1[1], and populated the different areas of interest with estimates of bandwidth (frequency of change, driving expectancy) and value (importance of the task served by the AOI), that would characteristize the automated cockpit. Effort was inherent in the display layout, with greater effort characterizing areas that were farther apart. Other than adjusting AOI parameters for expectancy (bandwidth) and value, the other model parameters were set to the same values as were established in the model fitting exercise described above from the GA cockpit

[1] This cockpit model was created under NASA NRA# NNX08AE87A; See Gore et al., 2009

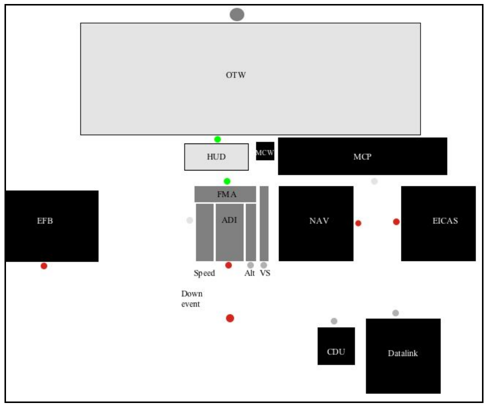

Figure 1: Cockpit image employed for automated cockpit validation model runs. Note that the printed labels on the image were not present during model runs. The dots in the figure represent different possible locations of a TBNE.

Figures 2a and 2b depict the scatter plot of model-predicted versus simulation-obtained percentage dwell time (attentional interest) across the different active areas of interest for the four flight phases of takeoff, cruise, descent and final approach whose point shapes are defined in the legend. The different areas of interest within a phase are not identified within the figures. Figure 2a includes the local areas within and around the primary flight display (altimeter, vertical speed, ADI, airspeed, nav display). Figure 2b includes larger areas more globally, including those specific to automation (primary flight display, navigation display, control display unit, mode control panel and outside world). Both figures reflect reasonably strong correlations between predicted and obtained attentional interest of r= 0.85 (local) and r= 0.88 (global). While both correlations are less than those observed in the model fitting exercise described above, we note that this is the first time that the SEEV model has been applied to the automated cockpit, so we had no firm basis on which to estimate expectancy and value parameters for the different FMS-specific AOIs.

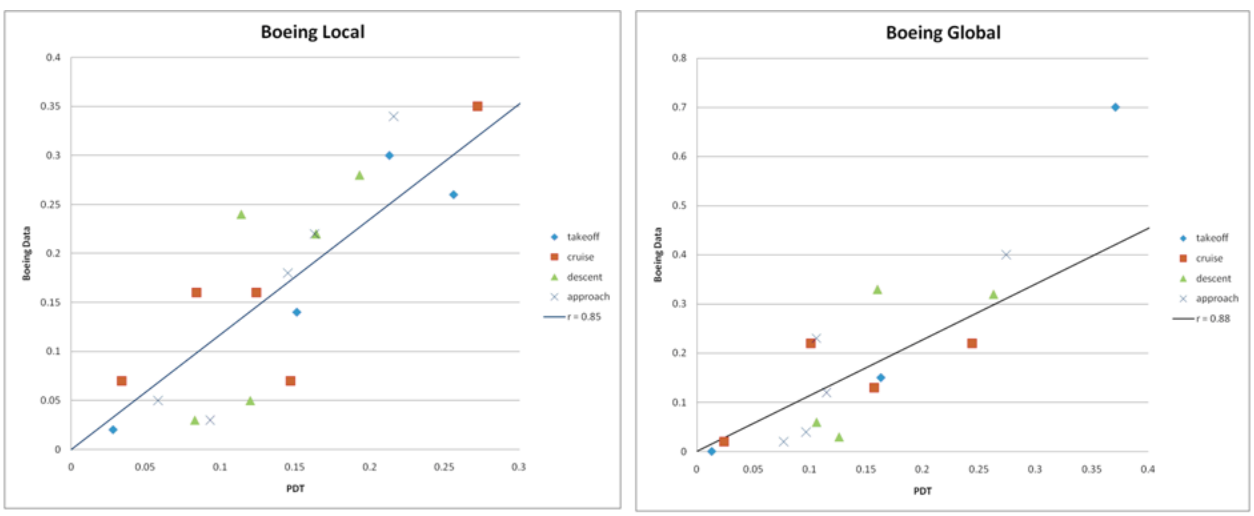

Figure 2a (left): Results of model validation runs from automated cockpit. Predicted scan (X axis) versus obtained scan (Y axis) for **local** areas of interest (r=0.85). All points represented by a common symbol type are measured in a single flight phase. Different points within this symbol type represent different areas of interest. Figure 2b (right): Results of model validation runs from automated cockpit. Predicted scan (X axis) versus obtained scan (Y axis) for **global** areas of interest (r=0.88).

Our second validation study employed the data collected in a simulation of flight management annunciator (FMA) noticing data carried out by Nikolic, Orr and Sarter (2004). The investigators examined the properties of the

FMA that led pilots often to fail to notice the onset of a green box, surrounding the FMA, that indicated an important change in FMS flight mode (i.e., in the way that the automation was flying the aircraft).

This simulation was less realistic than that employed in the Boeing cockpit study, but offered the unique attributes that noticing time and miss rate for events whose salience was varied were measured across two different levels of eccentricity. Our model simulation of the experimental paradigm used by Nikolic et al is shown in figure 3. Participants in the original experiment were asked to play a highly demanding game of Tetris (whose display is shown on the left side of the image), requiring the heavy focus of attention, corresponding to that imposed by high visual engagement with the primary flight display. In parallel, they were to monitor the right side of the image and to report any onset of a green box. Figure 3 depicts an example of the image that we provided to the model for validation. Heavy participant attention is focused on the Tetris box to the left (bandwidth = 0.8, value = 0.8 on a scale of 0- 1.0). The two green boxes between the circles to the right are the locations where a green box onset occurred in the near (35 degrees eccentricity from the Tetris game) and far (45 degrees) condition. Because they were of value to be noticed, they were assigned a value parameter of 0.2. The remaining AOI's represent "clutter displays" which, when present, were green, and rendered dynamic, (to mimic the rotation of dials in the original experiment) by providing them with 0.20 bandwidth settings. In a non-clutter condition, these other instruments were colored pale yellow and had near 0 bandwidth. All clutter AOI's had the value parameter set to 0.

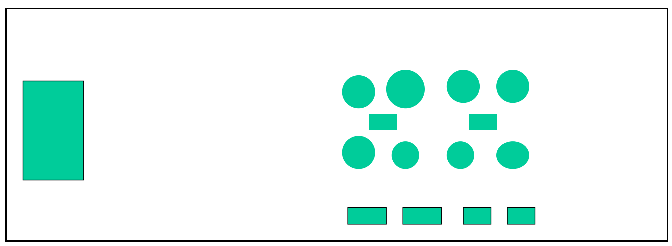

Figure 3: Cockpit image employed to validate noticing predictions, from Nikolic, et al., 2004.

We exercised the model in the near and far conditions, crossed with clutter vs no-clutter conditions. Each model run generates a single output of the number of scans to notice; thus across repeated model runs, a distribution of scans-until-noticing is created. These distributions, not unlike real data, tend to be positively skewed. We made two further assumptions: (1) that participants make approximately two scans per second, so that the conversion of scans-to-notice to noticing time is division by two. (2) Not all event onsets trigger a detection. We made a plausible assumption that since the green box remained illuminated for 10 seconds, any model run for which the TBNE fixation was not achieved within 10 seconds (20 fixations) was considered a "miss". Hence we were able to compute a miss rate. Figure 4a presents the scatter plot of model-predicted versus experimenter-obtained RT (for those responses less than 10 seconds), and figure 4b presents equivalent data for miss rate. Four conditions are represented in each figure: low and high clutter, crossed with near and far display locations.(Note that only that only three points are visible in each graph since, for each, there is a case of two data points directly overlaying each other). Both sets of data show a very high (r > 0.97) correlation, and both also show the approximate values of predicted RT and miss rate, to be within the same range as those of the obtained data.

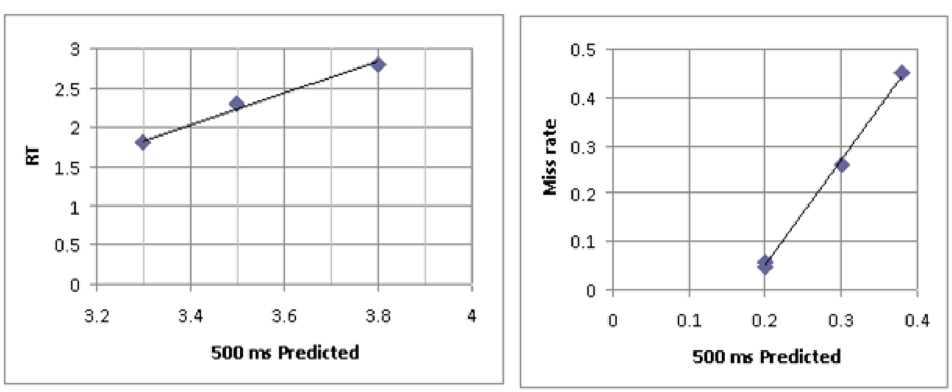

Figure 4a: Noticing predictions and observations across four conditions from Nikolic, et al. (2004) for noticing (or response) time; Figure 4b: Noticing predictions and observations across four conditions from Nikolic, et al. (2004) for miss rate, assuming all runs with fixation time on the green box greater than 10 seconds were misses.

Discussion and Conclusion

The current version of the N-SEEV model appears to do a competent job of integrating both bottom-up (salience and effort) and top-down (expectancy and value) parameters to predict either or both noticing time and event detection rate in a complex visual environment. The model was effectively validated on two sets of experimental data, one predicting scanning, and the other predicting noticing time as inferred from scanning.

In conclusion, we also note the added capabilities of the N-SEEV model, not validated here against empirical data:

- The effects of visual workload can be simulated by increasing the bandwidth at an AOI in question. For example the difference between low and high turbulence can be simulated by increasing the bandwidth at the ADI from 0.2 to 0.8.
- The effects of cognitive workload can be simulated by narrowing the field of view within the model, to mimic the well-known effects of cognitive load on attentional tunneling.
- The effects of attentional set (e.g,. greater for a red than an amber or white alert) are achieved by adjusting a color-tuning parameter within the model.
- The effects of human skepticism in false-alert-prone can be rendered by lowering the expectancy of the alert – essentially mimicking a lower expectancy for a **valid** alert (Dixon & Wickens, 2005). These and other features offer promise for a valid computational model of this vitally important cognitive process.
- The effect of surprise (low expectancy) can be simulated by setting bandwith of the TBNE to 0. As such this mimics the response to off-nominal events. Gore et al (2009) describes how model-predictions of off-nominal events are validated against flight simulation data obtained from a meta-analysis (see also Hooey et al, 2009).

Finally, there is a potential to input the scanning output of the current model into further models of situation awareness (Wickens McCarley et al, 2008; Wickens Sebok et al 2008) and, indeed, into overall pilot performance models (Gore et al, 2009).

Acknowledgements

This research was supported by grant # NASA Research Opportunities in Aeronautics (ROA) 2007, NNX07AV97A, awarded to the University of Illinois. The authors wish to acknowledge Dr. Jeffrey Mulligan of NASA Ames for his support of this work.

References

Helleberg, J. & Wickens, C. D. (2003). Effects of data link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology*, 13(3), 189-210.

Gore, B. F., Hooey, B. L., Wickens, C. D., & Scott-Nash, S. (2009). A Computational Implementation of a Human Attention Guiding Mechanism in MIDAS v5. To appear in *Proceedings of the 13th International Conference on Human-Computer Interaction*.

Gore, B. F., Hooey, B. L., Salud, E., Wickens, C. D., Sebok, A., Hutchins, S., Koenecke, C., & Bzostek, J. (2009). *Identification Of NextGen Air Traffic Control and Pilot Performance Parameters for Human Performance Model Development in the Transitional Airspace.* (Final Report: NRA # NNX08AE87A). San Jose, CA: San Jose State University Research Foundation.

Hooey, B. L., Wickens, C. D., Salud, E., Sebok, A., Hutchins, S., & Gore, B. F. (2009). Predicting the Unpredictable: Estimating Human Performance Parameters for Off-Nominal Events. Proceedings of the 15th International Symposium of Aviation Psychology. Wright State U: Dayton, OH

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Martens, M (2007). The failure to act upon important information. PhD dissertation. Free University, Netherlands.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why Pilots Miss the Green Box: How Display Context Undermines Attention Capture. *The International Journal of Aviation Psychology*, *14*(1), 39–52.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

SAE-ARP (2007). Society of Automobile Engineering Aerospace Recommended Practice ARP4102/4. Flight Deck Alerting Systems. Warrendale Pa:. SAE International.

Sarter, N.B., Mumaw, R., & Wickens, C.D. (2007). Pilots Monitoring Strategies and Performance on Highly Automated Glass Cockpit Aircraft. *Human Factors*, *49*(3), 347-357.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*(7), 261-267

Steelman-Allen, K, McCarley, J, and Wickens, C. (submitted). A salience model of event noticing. Submitted to 2009 Annual meeting of the Human Factors & Ergonomics Society.

Steltzer E. M. & Wickens, C. D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks*. Human Factors.48*(1), 166-181.

Wickens, C.D & Alexander, A.L. (2009) Attentional tunneling and task management. *International Journal of Aviation Psychology, 19(2), 1-17.*

Wickens, C. D., & McCarley, J. M. (2008). *Applied Attention Theory.* Boca Raton, FL: CRC Press.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380.

Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.

Wickens, C. D., McCarley, J. S., Alexander, A., Thomas, L., Ambinder, M., & Zheng, S. (2008). Attention-situation awareness (A-SA) model of pilot error. In D. C. Foyle, & B. L. Hooey (Eds.). *Human Performance Modeling in Aviation*. Boca Raton, FL.: CRC press.

Wickens Sebok et al (2007) Modeling advanced flight deck displays. *Human Factors and Ergonomics Society Annual Meeting Proceedings*. Santa Monica, CA: HFES.

Wickens, C. D., Helleberg, J., & Xu, X. (2002). Pilot maneuver choice and workload in free flight. *Human Factors, 44*(2), 171-188.