REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
29-09-2015	Book Chapter	01/01/2015 – 09/29/2015

4. TITLE AND SUBTITLE

Developing and Validating Practical Eye Metrics for the Sense-Assess-Augment Framework

5a. CONTRACT NUMBER
FA8650-14-D-6501-0009

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Matthew Middendorf (1)
Christina Gruenwald (2)
Michael Vidulich (3)
Scott Galster (3)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0HJ (53290813)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Middendorf Scientific Services, Medway, OH 45341 (1)
Oak Ridge Institute for Science and Education (ORISE), Belcamp, MD 21017 (2)
711 HPW/RHCP, 2510 Fifth Street Wright-Patterson AFB, OH 45433 (3)

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Airman Systems Directorate
Warfighter Interface Division
Applied Neuroscience Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

88 ABW Cleared 11/03/2015; 88ABW-2015-5365.

14. ABSTRACT

The present chapter examines eye measurements that could potentially inform the machine side of the human-machine system about the level of mental workload experienced by the human operator, boosting the machine's ability to aid the human adaptively. To realize this potential, the present work describes algorithms that were developed to detect eye blinks and saccades during real-time mission performance.

15. SUBJECT TERMS
Mental Workload, Psychophysiological Assessment, Adaptive Augmentation, Cognitive Support, Human-Computer Interface

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Michael Vidulich
Unclassified	Unclassified	Unclassified	SAR	18	**19b. TELEPHONE NUMBER** *(include area code)*

Developing and Validating Practical Eye Metrics for the Sense-Assess-Augment Framework

Matthew Middendorf, Christina Gruenwald, Michael Vidulich, & Scott Galster

General Introduction

So far, the field of human factors has generally focused on creating better controls and displays to help the human operator understand the situation and perform the mission. By comparison, there has been relatively little work on how to help the system better understand the human.

Despite the impressive system performance gains that have been accomplished by such work (e.g. Wickens, Hollands, Banbury, & Parasuraman, 2013), it is clearly asymmetric and unbalanced. During the real-time performance of a mission, the traditional approach requires the human to adapt to the machine side of the system. Originally, the machine was merely a conduit for executing whatever the operator commanded; there was relatively little the machine could do to aid the human's understanding or selection of goals. Typically, the machine was limited to providing warnings of potential dangers; such as the "check engine" or "low fuel" lights in automobiles or the "stall warning" indicators in aircraft. Such indicators are very useful in aiding the human's prioritization of goals, but do not really represent the full range of assistance that would be expected from a human teammate. The machine simply lacked a comprehensive understanding of how to assist the human.

In particular, the machine lacked any appreciation of the state of the human operator. Among other things a good human teammate would be sensitive to whether another member of the team was overloaded or stressed and, if so, try to help. A good human teammate would also be more subtly engaged in whether the team was progressing towards achieving appropriate goals, rather than simply announcing that a dangerous situation, such as low fuel, had emerged.

Until recently the machine side of the human-machine team has simply been unable to perform such roles. This relationship is changing as the automation embedded in the machine has gained more and more intelligence by virtue of increased computational power and sophisticated software. Certainly more sensors are often available for the machine to better understand the current situation and provide a larger repertoire of danger warnings to the human. For example, in modern automobiles "blind spot" and "lane change" warnings are presumably helping the human driver improve safety.

However, a good teammate must be sensitive to more than just the emergence of situational dangers; a good teammate must also be sensitive to the emergence of challenges for other members of the team. In human-machine systems this means that the machine must have a better real-time understanding of the human. The likeliest way to provide such understanding to the machine is by exploiting psychophysiological signals from the human operator to inform the machine about the state of the human operator.

This is not a new insight. For example, in 1981the McDonnell Douglas Astronautics Company under the sponsorship of the Cybernetics Technology Office of the Defense Advanced

Research Projects Agency (DARPA) hosted the ambitious "Biocybernetic Applications for Military Systems" conference (Gomer, 1981). In his opening description of the conference's goals, Gomer (1981) defined Biocybernetics as "a real-time communication link from the operator to the system he controls, which uses physiological signals that are recorded as the operator performs assigned tasks" (p. 2). The work presented at the conference tended to highlight the potential role of electrophysiological brain assessments, including event-related evoked potentials. Nevertheless a number of other physiological data sources, such as eye movements and pupillary responses, were considered.

Although the work reviewed at the Biocybernetic conference did not lead to any immediate real-world applications, it was representative of a vision to make machines better teammates for their human operators. Research on various pieces of such a project continued in many laboratories.

One of the most ambitious attempts to bring together a significant portion of the needed research was the "Augmented Cognition" program headed by Dylan Schmorrow (e.g., Schmorrow & Reeves, 2007). Among the agencies that participated in the Augmented Cognition program were the National Science Foundation, the National Institutes of Health, the DARPA, the Office of Naval Research, and the Air Force Research Laboratory (AFRL). The state of the art in technologies required to enable augmented cognition were discussed and critiqued at many well-attended meeting, and viable paths forward were debated. The meetings were very successful in building a critical mass of interest and shared knowledge regarding the viability of building such advanced systems. However, although the various researchers profited greatly from the interchanges to improve their own research and development, no fully-realized application of a system implementing augmented cognition was produced.

To capitalize on all of the progress from previous research it seems appropriate to attempt to combine relevant technologies with the goal of producing robust systems that will be ready to make the transition from laboratory to application. The AFRL has started a program to accomplish just that.

***The Sense-Assess-Augment (SAA) Framework*.** To better close the loop between the human and machine teammates AFRL's Human Performance Wing and Human Effectiveness Directorate has undertaken the "Quantified Warrior" program (Blackhurst, Gresham, & Stone, 2012). The Quantified Warrior program leverages the Sense-Assess-Augment (SAA) framework, which is designed to *sense* a suite of physiological signals from the operator, use these signals to *assess* the operators mental workload, and *augment* performance to optimize mission effectiveness" (Galster & Johnson, 2013). In contrast to the earlier attempts to use psychophysiological measures to improve human-machine teamwork (such as Biocybernetics or Augmented Cognition) the AFRL-SAA research program will attempt to combine all of the necessary components of an intelligent and adaptive human-machine teamwork system within one cohesive research program.

One strategy enabling the SAA framework is to leverage the available technologies as much as possible and to identify the gaps requiring solutions in order to apply those technologies effectively. For example, in regard to the *sense* portion of the SAA framework, there are many off-the-shelf sensing technologies that can provide useful real-time, or near real-time, data streams regarding the human operator's psychophysiological state. These include heart measures, skin conductance, electroencephalogram (EEG), and eye-based measures, among others.

The raw physiological signals from the selected sensors are processed by algorithms to extract features. For example, heart rate variability can be extracted from the electrocardiogram (ECG) and blink rate can be extracted from the electrooculogram (EOG). The extracted features are processed by models in the *assess* portion of the SAA framework to determine cognitive workload. This can provide a clear indication of when the operators need assistance.

If the *assess* portion of the SAA framework indicates a sub-optimal human state, appropriate assistance could be made available. For example, if the *assess* stage determined that an operator's workload measure was at a high level, this information could be available to the operator as bio-feedback, displayed to a commander who could take action, or used by the machine. As a result, this could lead to reassignment of task responsibilities or some form of augmentation, which would ultimately lead to reduced workload and increased performance by the operator.

In the current research project, the goal was to develop and evaluate algorithms that produce measures for potential use within the SAA framework. More specifically, this work presents robust algorithms that reliably and accurately detect eye blinks and saccades. Additionally, these algorithms produce features when blinks and saccades are detected such as blink duration and peak saccade velocity. The selection of eye blinks and saccades for evaluation was inspired by a couple of general attributes of the measures.

First, in laboratory evaluations eye blinks and saccades have been demonstrated to be sensitive to manipulations of mental workload (e.g., Chen & Epps, 2013; Di Stasi, Marchitto, Antolí, & Cañas, 2013; Halverson, Estepp, Christensen, & Monnin, 2012; Gao, Wang, Song, Li, & Dong, 2013) and even to momentary transitions in the chain of thought (e.g., Nakano, 2015; Nakano, Yamamoto, Kitajo, Takahashi, & Kitazawa, 2009; Siegle, Ichikawa, & Steinhauer, 2008).

Secondly, the sensors required to collect the EOG data are simple, un-intrusive, and can easily transition to many real-world applications. Although, other eye measures (such as pupil size and eye-tracking) are also potentially useful (e.g., Fitts, Jones, & Milton, 1950; Yang, Kennedy, Sullivan, & Fricker, 2013), they are considerably more challenging to transition to real-world applications.

Pupil size and eye-tracking are often acquired using off-body cameras, which can impose a restricted field-of-view. A camera-based system can lose its lock on the eye if the participant slouches, changes seating position, or turns their head. Conversely, the EOG signals needed for blink and saccade detection remain continuous regardless of participant movement. One potential issue associated with the acquisition of EOG is that it requires electrodes to be attached to the skin, which may not be desirable in some real-world applications.

To enable these eye-based metrics to be transitioned to a full-scale SAA application, algorithms need to be available to process the raw EOG data and extract the required features. Such algorithms that have been developed and tested in a laboratory setting are reported here. These algorithms must be reliable and accurate. Specifically, the data streams need to be continuous without substantial dropouts and the measures need have sufficient precision to allow small differences to be reported. For the algorithms to be successful in a SAA application, the signal acquisition and processing must be feasible under real-world conditions.

Real-Time Blink Detection Algorithm

The basic shape of a blink in the vertical electrooculogram (VEOG) signal has distinctive features (see Figure 1). Andreassi (2007) describes the waveform as a sharp rise immediately followed by a sharp fall. The duration is short, the peak is rounded, and there is a noticeable overshoot before the signal returns to zero. In the blink detection algorithm discussion that follows, each time the VEOG signal goes above and below a threshold value is referred to as a "bump."

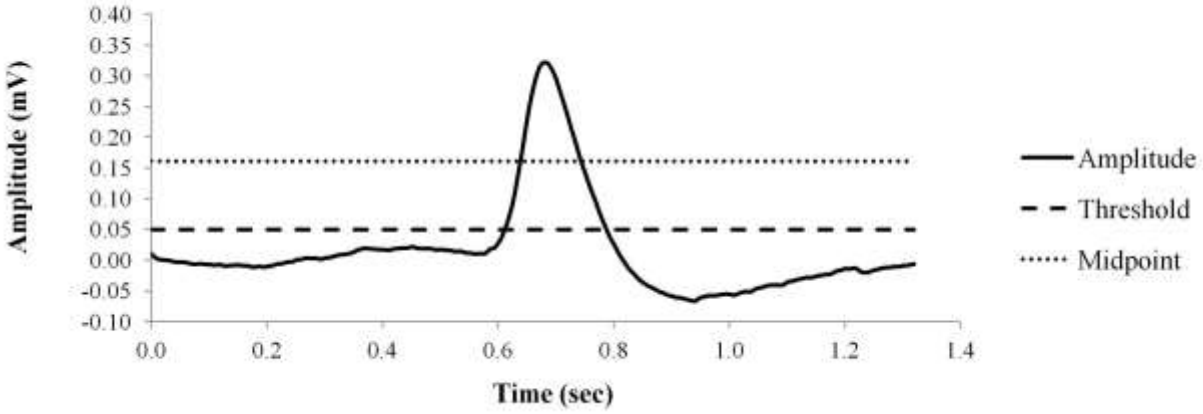

Figure 1 The Basic Shape of a Blink

Blinks are classified in three ways. Voluntary blinks are those that occur with a conscious decision to briefly close the eyes. Involuntary blinks involve both reflexive (startle) blinks, which occur to protect the eye in reaction to an external impetus, and spontaneous (endogenous) blinks, which are also reflexive but serve to maintain corneal moisture. Spontaneous blinks have shown sensitivity to mental workload (Chen & Epps, 2013). Therefore the algorithm presented here was designed for the detection of spontaneous blinks.

The VEOG is an electrical signal that can easily be acquired using a single pair of electrodes. The electrodes are positioned above and below one eye. Because the VEOG is easy to acquire, it is reasonable to assert that this blink detection algorithm could be easily transitioned to real-world applications.

While a blink in the VEOG signal is visually distinct to the trained human classifier, the varied parameters and noisy signal make blink detecting algorithms quite challenging to create. Many researchers in the human factors, psychophysiology, ophthalmology, and human-computer interaction domains have attempted blink detection using different algorithms and for different purposes. For example, Kong and Wilson (1998) filtered the signal four different ways before processing it with their algorithm. Blinks were then determined by finding a negative peak followed by a positive peak within a specified time window, along with other features used to give each potential blink a composite score.

Most of the available literature on blink detection algorithms lacks specific detail. Therefore, there was a research gap to produce an algorithm that could reliably perform and with a well-documented methodology.

Determining primary criteria for blink detection. To start the development of a blink detection algorithm, a VEOG database collected from previous experiments was utilized. The database consisted of 12 participants performing four different tasks. This resulted in a total of 3102 bumps that went above and below threshold. Trained raters manually coded 2020 of these bumps as blinks and 1082 as non-blinks.

This data was used to determine eight criteria values needed to develop the initial blink detection algorithm. These eight values were the minimum and maximum values for peak amplitude, duration at the midpoint, slope up at the midpoint, and slope down at the midpoint. Histograms of the extracted values were created and the data was found to be normally distributed. The data for the two slope features were also examined and found to be normally distributed. The initial values for the primary criteria were determined using the 98th percentile of the distributions and they were slightly adjusted after some testing. The final values of the primary criteria used in the detection algorithm are shown in Table 1. The criteria values define ranges for each of the four variables that are associated with true blinks rather than spurious bumps.

Table 1 Criteria Values for the Primary Features

Blink Feature	Minimum Criteria Value	Maximum Criteria Value
Amplitude (mV)	0.1211	0.6483
Blink Duration at the Midpoint (s)	0.06	0.198
Slope Up at the Midpoint (mV/s)	1.5	13.41
Slope Down at the Midpoint (mV/s)	-10.0	-1.25

An initial blink detection algorithm to perform blink classification using the four defined ranges was created for testing on a different set of VEOG data. In the initial classification algorithm, for any bump to be classified as a blink, all four extracted features must be within associated ranges.

Upon testing the initial algorithm on the new VEOG dataset, the results were promising but there was a tendency towards false positives. In other words, the initial algorithm tended to classify some of the spurious bumps (as classified by the trained human classifiers) as true blinks. This indicated a need for further algorithm refinement.

Determining secondary criteria for blink detection. To improve classification accuracy, it was decided to extract five secondary features from each bump. The five secondary features are the closure duration, the two R^2 values for linear fits at the midpoint, and two additional duration measures. The distance between the two linear fits at the peak is referred to as the closure duration. The distance between the two linear fits at the zero crossing is the blink duration at the zero crossing due to midpoint extrapolation. A similar duration is measured using linear fits about the threshold.

The criteria values for these secondary features were determined in the same manner as the primary criteria. The resulting values are listed in Table 2. These features were used to provide a confidence assessment to refine classification accuracy. Specifically, in addition to all four of the primary features falling within their defined ranges, at least three of the five secondary features had to fall within their defined ranges, except for the R^2 features, which only have a minimum value because R^2 cannot exceed 1.0.

Table 2 Criteria Values for the Secondary Features

Blink Feature	Minimum Criteria Value	Maximum Criteria Value
Closure Duration (s)	0.01	0.10
Slope Up at the Midpoint R^2	0.996	N/A
Slope Down at the Midpoint R^2	0.995	N/A
Blink Duration ZCMP (s)	0.1162	0.3
Blink Duration ZCT (s)	0.1	0.35

Note: ZCMP is duration at the zero crossing due to midpoint extrapolation. ZCT is duration at the zero crossing due to threshold extrapolation.

Algorithm Components. The major components of the blink detection algorithm are threshold generation, feature extraction state machine, scoring and classification, and blink save and false positive detection logic.

Threshold Generation uses a sliding five second window of raw VEOG data. To minimize the effects of blinks and eye movement on the threshold, the data is high pass filtered using a first order Butterworth filter with a break frequency of 10 Hz. This essentially leaves in the "noise" from which the threshold is calculated. The filtered signal is then rectified and the median is taken for the *raw* threshold value. The median is used because the data in the five second window can be highly skewed when there is a blink in the window. A second stage of threshold generation imposes limits on the *raw* threshold and incorporates a threshold reduction value to accommodate double and multiple blinks.

Initially the threshold limits are static, but after ten blinks have been detected, the limits are dynamic and can be adjusted based on the mean amplitude of the previously detected blinks. The threshold reduction value is necessary due to the high pass filter in the signal acquisition hardware, which causes the signal to overshoot zero on the down slope of the blink. The threshold reduction value is based on the amount of overshoot of the previous blink. The threshold returns to its normal (non-reduced) value using a function that is the inverse of the high pass filter implemented in the signal acquisition hardware.

The *Feature Extraction State Machine* uses the threshold to monitor the VEOG signal. The state machine has four values (0, 1, 2, and 3). In state zero the logic waits for the signal to be below threshold. In state one it waits for the signal to go above threshold, at which time upward threshold crossing data is captured and the threshold is frozen. In state two the logic is waiting for the signal to go back below threshold. During this time peak data and downward threshold crossing data are captured and the threshold is unfrozen. In state three the signal overshoot value

is captured and the extracted features are scored to see if the signal excursion above and below threshold is a blink. The state machine then returns to state zero.

The *Scoring and Classification* of any VEOG bump that goes above and below threshold is based on the primary and secondary criteria. One point is awarded when each of the four primary criteria are met and one tenth of a point is awarded when each of the five secondary criteria are met. Therefore, the maximum score for a VEOG bump is 4.5 points. Bumps that score 4.3 points or higher are classified as blinks. This requires that all four of the main features be met, and at least three of the five secondary features be met.

The *Blink Save and False Detection Logic* applies to a small number of VEOG bumps. If a bump fails only one of the four main criteria, but otherwise has a nearly perfect score (3.4 and 3.5), it is given a second look. When a bump fails the maximum amplitude criterion, the criterion can be adjusted upward using amplitude data from previous blinks (minimum of 10 required). Similar adaptive tests are applied when the minimum amplitude or the slope down at the midpoint fails. Currently only one false positive test is performed. Bumps that have two peaks are rejected.

Validation Results. In a recent study participants were asked to track targets using remotely piloted aircraft. Workload was experimentally manipulated and VEOG data was processed using the blink detection algorithm. In the study, workload had a statistically significant effect on blink rate and duration. Blink rate was slower and blink duration was shorter during high workload conditions.

It was encouraging that the blink detection algorithm produced measures that were sensitive to the workload manipulation (see Hoepf, Middendorf, Epling, & Galster, 2015). For further validation, an objective evaluation of the algorithm was also performed. Video recordings were used to generate truth data for comparison purposes. Eight participants were video recorded using a Basler high speed camera while performing two trials in the study. The output of the blink detection algorithm was evaluated by two separate trained raters using the video recordings. Only 2.3 percent of blinks were missed, whereas 1.0 percent of blinks were falsely detected. Overall, the blink detection algorithm had an accuracy rating of 96.7 percent.

Conclusion. The blink detection algorithm was validated against truth data and was found to be very accurate. It has several desirable attributes that make it potentially transitionable to real-time, real-world applications. This blink detection algorithm does not require baseline data or calibration. There is no need (or mechanism) for experimenter adjustments. Furthermore, it is adaptive in the sense that it worked well for a wide selection of individuals. Also, after the algorithm has compiled statistics on a few blinks, it can adapt its criteria to improve classification accuracy. In addition to the algorithm being adaptive, it is dynamic in the sense that the detection threshold will change in real-time in response to changes in the VEOG signal.

Real Time Saccade Detection Algorithm

Similar to blinks, saccades have been demonstrated to be useful in cognitive state assessment, including mental workload assessment (Romero, Mañanas, & Barbanoj, 2008). Saccades have also been used in investigation of drug effects, and clinical applications for neurological and psychiatric disorders (Romero, Mañanas, & Barbanoj, 2008).

For cognitive state research, saccades have been used both directly and indirectly. Literature suggests that peak saccade velocity can be directly used for the evaluation of mental workload (Di Stasi, et al., 2010). An indirect use of saccades is to mediate ocular artifacts in the EEG, which is also often used for workload assessment.

The saccade detection algorithm presented here is performed in polar coordinates. This means that the saccade is reported in magnitude and angle. To detect saccades using polar coordinates requires four electrodes; two for the horizontal electrooculogram (HEOG), and two for the vertical electrooculogram (VEOG).

Algorithm Components. The major components of the algorithm are signal filtering, threshold generation, saccade endpoint detection, dynamic linear fit, mathematical calculations, classification, and saccade queuing.

The *Signal Filtering* of the raw EOG data stream is necessary to improve the accuracy of the measured saccade amplitude. The raw EOG contains saccades that are evident to the naked eye (Figure 2). The distinctive shape of a saccade contains the pre-saccadic spike (Thickbroom & Mastaglia, 1986) followed by a sharp monotonic increase (or decrease for look down and look left). Then there is a slow decay back to zero due to the high pass filter used in the signal acquisition hardware.

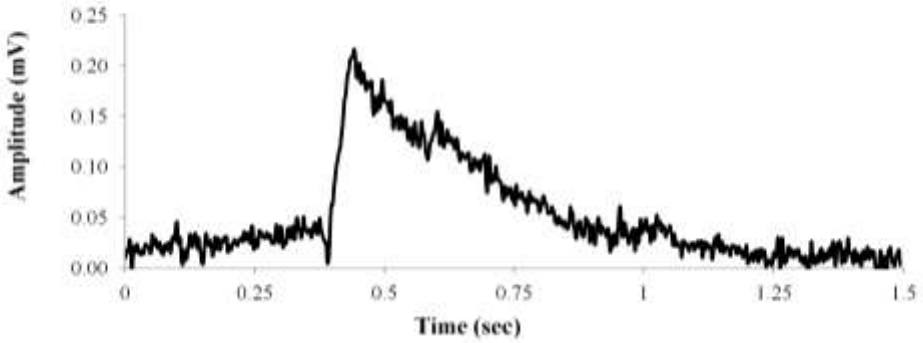

Figure 2 The typical shape of a saccade. This is a horizontal saccade to the right.

The raw EOG also contains micro-saccades, which unlike the major saccades, are very small in amplitude but, occur very frequently. These micro-saccades can occur in the middle of a major saccade (Figure 3). When this happens, the micro-saccades can cause the dynamic linear fit portion of the algorithm to be less accurate. Specifically, the full amplitude of the major saccade may not be reported. To prevent this problem, the raw EOG data is filtered using a first order Butterworth low pass filter with a break frequency of 50 Hz.

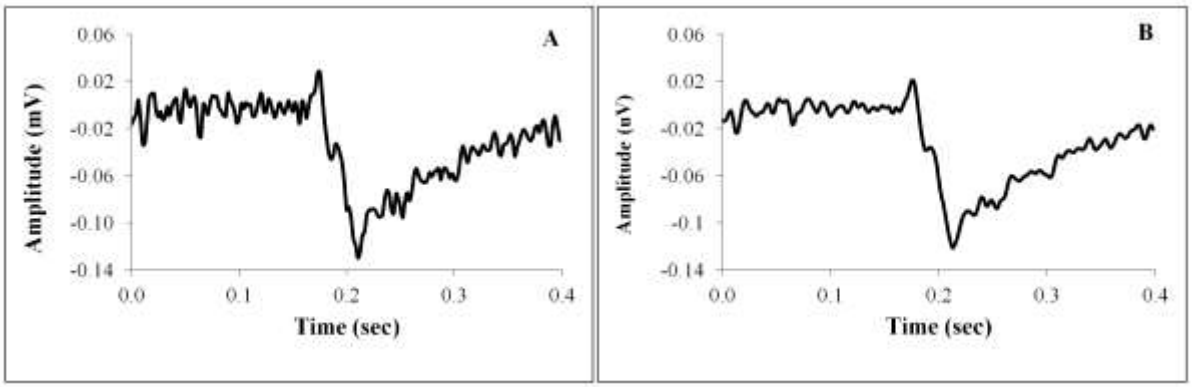

Figure 3 This horizontal saccade is from a leftward eye movement. It has a micro-saccade in the middle of it (A) that causes the linear fit to fall short of the full saccade amplitude. After filtering (B) the micro saccade is reduced enough so that it does not interfere with the full linear fit.

The robust *Threshold Generation* approach that was developed for the blink detection algorithm was adapted for the polar saccade detection algorithm. Similarly to the blink detection algorithm described above, the threshold generation approach uses a sliding five second window of raw VEOG data and is high pass filtered using a first order Butterworth filter. Note that the threshold used in the polar saccade detection algorithm is circular (Figure 4).

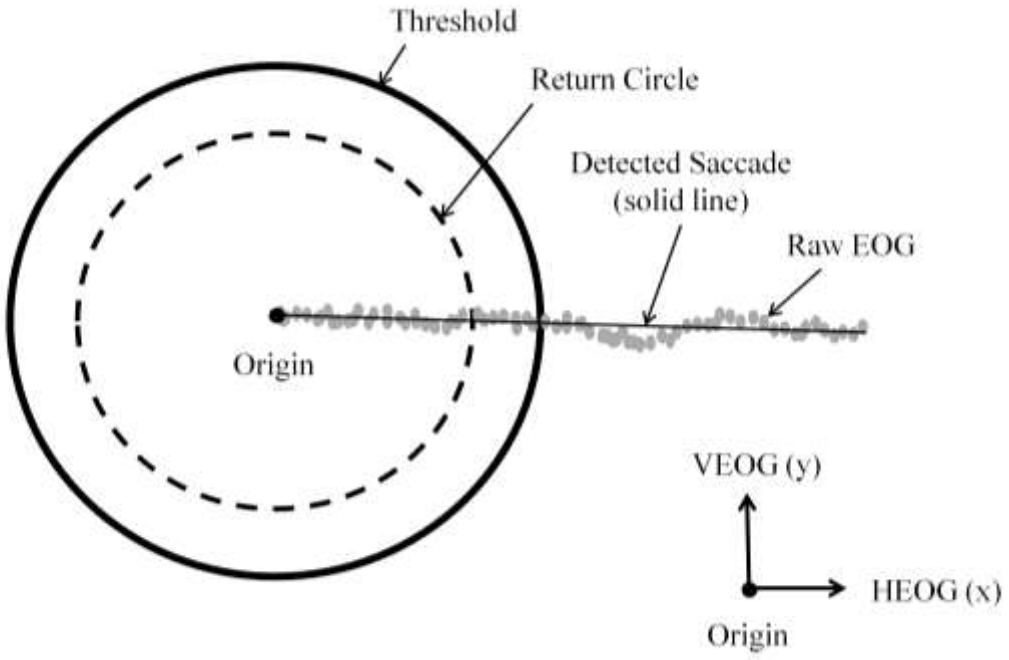

Figure 4 Polar saccade detection threshold, return circle, and raw signal.

The *Saccade Endpoint Detection* is performed in Cartesian coordinates with HEOG on the x-axis and VEOG on the y-axis. The circular threshold is centered about the x/y origin. Initial saccade detection follows three simple steps. First, the x/y position must start inside the circular threshold. Second, the x/y position must travel outside the circular threshold. Third, the x/y position is allowed to move away from the origin for as many samples as possible until it moves back toward the origin for two samples in a row. The last sample that is moving away from the origin is the endpoint of the saccade. To prevent undesirable false positives, a second circle called the return circle was added. This circle is centered about the origin and its radius is equal to two-thirds of the threshold circle. The detection logic required that the signal must return to inside the return circle before it can be tested for traveling outside the threshold. The saccade endpoint detection is accomplished using a state machine.

The *Dynamic Linear Fit* was performed in rectangular coordinates. That is, the VEOG and HEOG were processed separately based on the saccade endpoint. Two vectors were used to find the saccade starting point for each signal (VEOG & HEOG). These two vectors were referred to as the small vector and the big vector. The initial length of both vectors was 20 milliseconds (Chen & Wise, 1996). The heads of the vectors were set to the saccade endpoint and the tails were 20 milliseconds backwards in time. The length of the small vector remained constant but the big vector grew in length backwards in time. The tail of the small vector was anchored to the tail of the big vector. The small vector was used to terminate the growth of the big vector. When the slope of the small vector differed substantially the slope of the big vector, it marked the saccade starting point (Figure 5). After the dynamic linear fits were performed on both axes, the x and y coordinates of the starting point and ending point for the potential saccade are known.

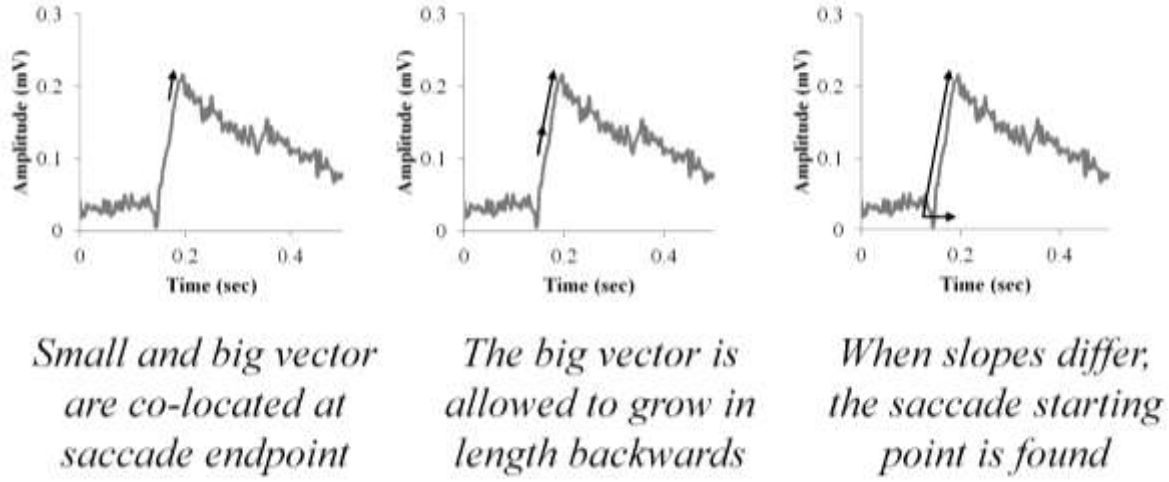

Figure 5 Linear Fit Vectors

Next, *Mathematical Calculations* of each potential saccade were performed to produce several variables. The rectangular coordinates were converted to polar coordinates (magnitude & angle). The amplitude, length, velocity, and peak velocity were computed for the potential

saccade. The two R^2 values from the dynamic linear fits were combined into a single R^2 value. Also, fixation duration was computed.

The *Classification* step used three of the variables computed above (R^2, velocity, and length). These variables were compared to criteria values to determine if the potential saccade was an actual saccade. All three of these criteria had to be met for a positive saccade classification. The initial criteria values were determined using data from a mini-study with four participants. This was accomplished using an EOG playback feature. All of the potential saccades were hand-scored by trained observers to generate truth data. The initial values were refined when data from a follow-on study was available (Mead, Hoepf, Middendorf, and Gruenwald, 2015). The final criteria values are shown in Table 3.

Table 3 Criteria values used for saccade classification

Criteria	Minimum Value	Maximum Value
Combined R^2	0.85	N/A
Velocity (mV/sec)	1.5	11.5
Length (sec)	0.028	0.125

The *Saccade Queuing* logic was needed to ensure that identified potential saccade is not actually the up slope of a blink. While waiting for the excursion of the EOG signal to reach its maximum distance from the origin, the blink detection algorithm was monitored. If the blink detector was active, then a flag was set to indicate that the saccade must be queued. If a blink was detected, then the queued saccade was discarded, otherwise the queued saccade was counted. In essence, this means that blinks "trump" saccades.

Validation Results. For validation purposes a study was conducted to test the performance of the polar saccade detection algorithm. In this study visual stimuli were presented at known angles and distances at regular intervals (1.5 seconds). Two researchers independently reviewed the raw VEOG and HEOG signals to verify that the saccades were present in the signal. This truth data was used to validate the algorithm. Results show that the algorithm had zero false positives, but did have occasional misses. Overall the algorithm had an accuracy rating of 92.6%.

Calibration. Software was developed to allow the two axes of EOG to be normalized, thus improving the accuracy of the reported saccade angles. The calibration process is needed because the VEOG signal is typically smaller than the HEOG signal for the same amount of angular movement of the eye. For some individuals, the difference in amplitude between the two axes can be as much as a factor of two. For others, the two axes have near symmetry. So, the calibration procedure can account for individual differences. The calibration procedure is easy to perform and only takes a few minutes. Results indicate that participant calibration is stable from day to day.

Conclusion. The measures produced by the polar saccade detection algorithm (velocity, amplitude and length) have excellent precision and clearly indicate changes in eye behavior associated with changes in task demands. The saccade detection algorithm produces these measures in real-time and in a computationally efficient manner.

Another positive aspect of the current work is it can support artifact mediation approaches when performing EEG analysis. This algorithm will be used in future research to determine its usefulness for assessing cognitive workload. Also, enhancements need to be made to the algorithm to improve the current overall accuracy of 92.6%.

An Indirect and Practical Use of Blinks and Saccades – EEG Artifact Separation

Blinks and saccades can be used in two different ways. A direct use of the ocular metrics is to treat them as the "signal," and an indirect use is to consider them the "noise." Metrics such as blink rate, blink duration, and peak saccade velocity can be used directly to assess operator cognitive state. Blinks and saccades can be used indirectly to mediate the contaminating effects they have on EEG data.

Physiological measures, such as EEG, have been shown to be an indicator or workload. The ability to reliably assess mental workload is important due to the effect increased workload can have on human operator performance. This is vital due to the ever increasing complexities of technology and systems, and the higher demand they place on the human operator (Hankins & Wilson, 1998).

EEG is a noninvasive electrical sensing technique that uses electrodes placed on the scalp to measure brain activity. Dependent upon the research, different sites may be used. The locations of these sites are based on the International 10-20 system (Jasper, 1958). Researchers have reported the sensitivity of EEG to changes in mental workload (Gevins & Smith, 2003). It has been shown that spectral power increases in the delta (1-3 Hz) and theta (4-7 Hz) bands during high workload (Gevins & Smith, 2003). While in contrast, multiple studies have shown that power decreases in the alpha band (8-12 Hz) during high workload.

Although EEG has often been used as a measure of cognitive workload, it has some functional and practical limitations that must be carefully considered before being applied to operational settings. EEG signals are easily corrupted by a number of artifacts. That is, in addition to the brain's electrical activity recorded at the scalp, the EEG signal can include contaminating potentials from saccades and blinks (Gevins & Smith, 2003).

Considering the effects of artifacts on the EEG signal, a great deal of research has been directed towards artifact mediation (Fatourechi, Bashashati, Ward & Birch, 2006). Common methods of dealing with artifacts in the EEG are artifact avoidance, artifact rejection, and artifact removal. If the blink and saccade detection algorithms are running on the same computer as the EEG signal processing software, then a new technique for ocular artifact mediation can be implemented.

This new technique is referred to as artifact separation (Credlebaugh, et al., 2015). This technique enables the consumer of the EEG spectral results to decide what to do with the artifacts. In the work discussed here, the artifact free data is separated from the contaminated data at the analysis stage, thus the blinks and saccades are used in a *post hoc* fashion. However, the blink and saccade information could be used in real-time. For example, models that use EEG for cognitive assessment may also want to eliminate the contaminated data. The artifact separation technique was used in the experiment discussed next.

Experimental Background. The use of EEG as an indicator of cognitive workload was explored in a recent study. In this study, workload was manipulated while participants performed a tracking task implemented in a remotely piloted aircraft (RPA) simulation. Participants were instructed to track one or two high value targets (HVTs) by continuously clicking on the video feeds while the HVTs traveled by motorcycle. Experimental manipulations consisted of number of HVTs to track (1 or 2), route taken by HVT (country or city), and weather (clear or hazy conditions). A secondary task of answering cognitively demanding math questions was also implemented.

It is noteworthy that the experimental manipulations were determined from interviews with real-world RPA operators. Specifically, tracking two targets is harder that tracking one. Likewise, tracking targets in the city is harder than the country and hazy conditions are harder than clear conditions. Using a factorial design of these manipulations has allowed for a wide variety of workload levels in the experiment.

In the RPA tracking study, participants were outfitted with an electrode cap so that EEG data could be acquired. Seven channels of EEG data were recorded, which included: F7, Fz, F8, T3, T4, Pz and O2. The frequency ranges of the seven bands of EEG were delta (1-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), gamma 1 (31-40 Hz), gamma 2 (41-57 Hz) and gamma 3 (63-100 Hz). The VEOG data were acquired using two electrodes placed above and below the left eye. Mastoids were used as reference and ground points.

In the study, there was a significant effect of workload on frontal delta. However, contrary to literature, these findings were in the wrong direction. Specifically, spectral power in the delta band decreased in the high workload condition. It has been well reported that blink rate decreases under high workload conditions, so it was unclear if the significant frontal delta effects were due to brain activity or EOG artifacts (Wang & Zhou, 2013). To investigate this concern, the blink detection algorithm was used to flag EEG spectral results that are contaminated by blinks. The flags were then used by the aforementioned artifact separation technique to investigate this concern.

EEG Signal Processing. The raw EEG data were split into two second windows and filtered using a 4^{th} order Butterworth band pass filter with pass bands set as described earlier. A Hanning window was applied and power spectral analysis was performed. The resulting power in each window was then averaged. The two second time domain windows had a 50% overlap, thus yielding one average power measure every second for each frequency band and site. This produced a total of 49 measures per second (7 frequency bands at 7 sites). Each EEG spectral measure is accompanied by two flags to indicate the presence of artifacts (blinks and saccades).

Results. The artifact separation technique was applied to see if significant frontal delta effects for the route (country vs. city) manipulation in the study were due to eye activity, or if they were due to an actual neurological phenomenon. When all of the data were used (no artifact separation), there was a significant effect of increased workload at six EEG sites. When blinks were separated, only two sites remained significant. These two sites lost significance when both blinks and saccades were separated.

Discussion. Applying the artifact separation approach caused the significant results of the study to change drastically. One could reasonably argue that artifact separation is the same thing as automatic artifact rejection. One big difference is artifact rejection is typically done in the time domain, and the artifact separation approach is done on the spectral results. Another nuance is

that it is up to the consumer of the data to decide what to do with the artifact flags. The EEG spectral results are available in real time for such applications as machine learning models. In this case, a model could decide if it wants artifact free data using the flags.

General Discussion

The algorithms presented here are quite mature. However, there is always an opportunity to refine their performance. These algorithms will be used in future research to assess operator state in real time. This will allow the researchers to see if the results can be replicated, and when needed, make subtle refinements to improve accuracy. The success of the *assess* portion of the SAA framework is heavily dependent on the precision of the features it receives. Therefore, it is imperative that as this technology is transitioned to the operational community, the input signals retain the same, or nearly the same, levels of resolution and accuracy that is found in the experimental setting.

Because raw physiological signals by themselves are of little use for assessing operator state, algorithms are needed to process the raw signals and extract features. For example, EOG in its raw form has very limited informative value regarding mental workload. However, when blink rate and blink duration are extracted, these features can be reliable indicators of stress associated with high workload (Wang & Zhou, 2013). While this chapter focused on algorithms that use EOG for blink and saccade detection, the SAA framework also requires algorithms to extract features from other physiological signals; such as, ECG, EEG, electromyography (EMG), and respiration. These additional algorithms are currently at various stages of maturity, and will be reported in future publications.

Regardless of which physiological measures are being considered, their relevance to mental workload and/or performance needs to be validated in a laboratory setting. This is needed to support the *assess* phase of the SAA framework.

When multiple algorithms are running concurrently, there is an opportunity for coupling. For example, as discussed in this chapter, the saccade detection algorithm was coupled with the blink detection algorithm to prevent the up slope of a blink from being counted as a saccade. In essence, there was an opportunity for the algorithms to "talk" to each other to improve their accuracy; in this case classification accuracy substantially exceeded 90%.

To transition any algorithms, including the current blink and saccade algorithms, into operational environments, the necessary physiological signals must be reliably and accurately collected in the relevant setting. Some operational communities may balk at the notion of attaching electrodes to the operator's face. If so, the EOG data could potentially be replaced with eye lid position data from off-body cameras if the sample rate is high enough. The sample rate of the EOG used in the current algorithms was 480 Hz. There are already cameras that are capable of frame rates at or above this frequency. Software will need to be written to convert video data to eye lid position data. Testing the current software using eye lid position data is a promising future development.

Conclusion

Many researchers have tried to develop psychophysiological measures to improve the teamwork between human operators and systems that they must control. However, integrating the needed sensors and software was previously beyond the grasp of any individual research and development effort. But with the recent breakthroughs in fairly nonintrusive physiological sensing capabilities, the viability of bringing together all of the necessary components should now be within reach.

The AFRL SAA program is designed to integrate sensors for acquiring physiological data, develop algorithms for feature extraction, assessing the implications of the features, to inform augmentations that aid the human operators to achieve their missions. The current work demonstrated that relatively easily acquired eye activity data has great potential for the SAA framework. The sensing and algorithms for using this eye data will be further developed.

But more importantly, this approach will be extended to developing comparable assessment from other physiological signals to produce robust assessments of the human operator's state. This will ultimately be used to implement appropriate augmentations such as alerting the operator or teammates, shifting automation priorities, or activating other aids.

References

Andreassi, J. L. (2007). *Psychophysiology: Human behavior and physiological response* (5th ed.). Mahwah, NJ: Erlbaum.

Blackhurst, J. L., Gresham, J. S., & Stone, M. O. (2012, December). The quantified warrior: How DoD should lead human performance augmentation. *Armed Forces Journal*, *32*, 14-17.

Chen, S. & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine, 110,* 111-124. doi: 10.1016/j.cmpb.2012.10.021

Chen, L. L. & Wise, S. P. (1996). Evolution of directional preferences in the supplementary eye field during acquisition of conditional oculomotor associations. *The Journal of Neuroscience, 16*, 3067-3081.

Credlebaugh, C., Middendorf, M., Hoepf, M., & Galster, S. (2015). EEG data analysis using artifact separation. In *Proceedings of the Eighteenth International Symposium on Aviation Psychology*, Wright State University.

Di Stasi, L. L., Marchitto, M., Antolí, A., & Cañas, J. J. (2013). Saccadic peak velocity as an alternative index of operator attention: A short review. *Revue Européenne de Psychologie Appliquée, 63,* 335-343. doi: 10.1016/j.erap.2013.09.001

Di Stasi, L. L., Rennar, R., Staehr, P., Helmert, J. R., Velichkovsky, B. M., Canas, J. J., Cantena, A., & Pannasch, S. (2010). Saccadic peak velocity sensitivity to variations in mental workload. *Aviation, Space, and Environmental Medicine, 81,* 413-417. doi: 10.3357/ASEM.2579.2010

Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology, 118,* 480-494.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review, 9,* 24-29.

Gao, Q., Wang, Y., Song, F., Li, Z., & Dong, X. (2013). Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics, 56,* 1070-1085. doi: 10.1080/00140139.2013.790483

Galster, S. M. & Johnson, E. M. (2013). Sense-Assess-Augment: A taxonomy for human effectiveness. Technical Report. Wright-Patterson Air Force Base, OH: United States Air Force Research Laboratory

Gevins, A. S. & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science, 4,* 113-131. doi: 10.1080/14639220210159717

Gomer, F. E. (Ed.). (1981). *Biocybernetic applications for military systems* (Report No. MDC E2191). Saint Louis, MO: McDonnell Douglas Astronautics Company – St. Louis Division.

Halverson, T., Estepp, J., Christensen, J., & Monnin, J. (2012). Classifying workload with eye movements in a complex task. In *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, 56,* 168-172. doi: 10.1177/1071181312561012

Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environment Medicine, 69,* 360-367.

Hoepf, M., Middendorf, M., Epling, S., & Galster, S. (2015). Physiological indicators of workload in a remotely piloted aircraft simulation. In *Proceedings of the 18th International Symposium on Aviation Psychology*, 428-433. Dayton, OH.

Jasper, H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10,* 370-375.

Kong, X. & Wilson, G. F. (1998). A new EOG-based eye-blink detection algorithm. *Behavior Research Methods, Instruments, & Computers, 30,* 713-719.

Mead, J., Middendorf, M., Hoepf, M., & Gruenwald, C. (2015). Polar saccade detection algorithm: Classification criteria refinement. Technical Report. Wright-Patterson Air Force Base, OH: United States Air Force Research Laboratory

Nakano, T. (2015). Blink-related dynamic switching between internal and external orienting networks while viewing videos. *Neuroscience Research, 96,* 54-58. doi: 10.1016/j.neures.2015.02.010

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., & Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. In *Proceedings of the Royal Society B, 276,* 3635-3644. doi: 10.1098/rspb.2009.0828

Romero, S., Mañanas, M. A., & Barbanoj, M. J. (2008). A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical variables: A simulation case. *Computers in Biology and Medicine 38,* 348-360. doi: 10.1016/j.compbiomed.2007.12.001

Schmorrow, D. D. & Reeves, L. M (Eds.) (2007). *Foundations of Augmented Cognition: Lecture notes in artificial intelligence, 4565.* Germany: Springer.

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45,* 679-687. doi: 10.1111/j.1469-8986.2008.00681.x

Thickbroom, G. W. & Mastaglia F. L. (1986). Presaccadic spike potential. Relation to eye movement direction. *Electroencephalography and Clinical Neurophysiology, 64,* 211–214. doi: 10.1016/0013-4694(86)90167-7

Wang, Y., & Zhou, J. (2013). Literature review on physiological measure of cognitive workload. Machine learning research group – NICTA

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (4th ed.). Boston: Pearson.

Yang, J. H., Kennedy, Q., Sullivan, J., & Fricker, R. D. Jr. (2013). Pilot performance: Assessing how scan patterns & navigational assessments vary by flight experience. *Aviation, Space, and Environmental Medicine, 84,* 116-124. doi: 10.3357/ASEM.3372.2013