SYSTEMS APPROACH TO THE DESIGN OF LOCOMOTIVE FATIGUE MANAGEMENT TECHNOLOGIES

by

Anton Aboukhalil

B. Eng. Honours, Electrical Engineering
McGill University, 2004

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2006

Author: _____
Anton Aboukhalil
Department of Aeronautics and Astronautics
January 27, 2006

Certified by: _____
Charles M. Oman
Senior Lecturer, Department of Aeronautics and Astronautics
Thesis Supervisor

Accepted by: _____
Jaime Peraire
Professor of Aeronautics and Astronautics
Chair, Department Committee on Graduate Students

SYSTEMS APPROACH TO THE DESIGN OF LOCOMOTIVE FATIGUE MANAGEMENT TECHNOLOGIES

by
Anton Aboukhalil

Submitted to the Department of Aeronautics and Astronautics
on January 27, 2006, in partial fulfillment of the requirements for the Degree of
Master of Science in Aeronautics and Astronautics

Abstract

Falling asleep while operating a vehicle leads to serious accidents and loss of lives. The challenge of detecting drowsiness nonintrusively stems from the absence of a single marker, and the existence of diverse signs and symptoms that collectively but not uniquely characterize it. Current alerters for locomotive cabs are inadequate partly because they a) often monitor one modality b) fail to consider the inherent characteristics of the locomotive operator tasks and physical environment, and c) are typically developed without quantitative techniques to assess performance and optimize components as part of an overall system, rather than at the individual level.

Based on an estimation theory framework, a new systems approach is here proposed to design locomotive cab alerting technologies. The main idea is to combine information from an infrared eyelid monitor and a generic Train Sentry class activity monitor, to isolate the common drowsiness component and obtain an improved estimate of the operator's state.

A study first quantified the important physical aspects of the locomotive cab and engineer behavior pertinent to the performance of image-based eye closure monitors. A bench test study evaluated Attention Technology's current infrared eye closure monitor prototype front-end image analysis, Model DD-850, to verify whether its performance was a good match to the locomotive physical environment and engineer behavioral characteristics. Data from these studies were used to develop a simulation software tool in MATLAB/SIMULINK. The goal was to assess the proposed tandem detector solution and to support rational design, development and optimization of future locomotive alerting systems. A signal detection theory (SDT) approach was employed. However, the detectors were nonlinear, had multiple alerting levels and displayed non-Gaussian noise characteristics. Therefore, Monte Carlo methods were used to compute their SDT parameters on both a standalone and tandem basis.

Investigation through simulation showed that adopting an architecture using tandem detectors and an "AND" logic based arbiter reduces the false alarm rate by an order of magnitude and improves the total time to alert, at the expense of only a few percent in missed alarm probability. Detection performance may be further enhanced using a speed dependent arbiter with "AND" logic above a speed threshold and "OR" logic below it. In the simulation, the speed threshold was found to be 25 mph. This system-level study provides enough ground to build a prototype system and test the proposed solution in a simulator.

Supported by FRA through Department of Transportation Volpe Research Center Contract DTRS57-04-Q-80164 PR 79-3354 and by a Postgraduate Scholarship from the Natural Sciences and Engineering Research Council of Canada.

Thesis Supervisor: Dr. Charles M. Oman.
Title: Senior Lecturer. Director, Man Vehicle Laboratory. Department of Aeronautics and Astronautics.

"This is what happens when you don't finish your thesis..."

Dan Buckland
-on the July 10th 2005, ~ 4 A.M. freight train collision in Mississippi

Acknowledgements

I came to MIT aspiring to do science and conduct vestibular research with application to spaceflight. However, this was not the case, and I was presented with this project sponsored by the US DOT Volpe Center. All I knew was that it dealt with "fatigue" and "trains". It had seemingly nothing to do with space or science! I had no clue what to expect when I signed up for it. However, looking back, I must say that it was a truly rewarding intellectual journey! I have learned a great deal about conducting multi-modal research, with immediate and practical application to diverse areas of transportation, and not only to railway or aerospace.

As the saying goes, "*When you see a frog at the top of a pole, you know he has not made it on his own.*" That's why I would like to thank the many great people who have helped me reach this stage and finish this work.

First, I would like to sincerely thank the 'one and only' Dr. Chuck Oman. With your brilliant mind, unvarying enthusiasm and passion, you have continuously challenged me to think broadly, to look at things from different perspectives, and to keep asking the important questions. You gave me outstanding guidance and phenomenal support, and constantly provided remarkable feedback during our regular discussions and by thoroughly editing this thesis. On top of all that, you have been an extremely kind mentor. During the Cab Characterization Study, I will never forget how you drove all the way to Worcester to give me a ride back just so that I didn't have to wait for the train and return to Boston at 1am. Working with yourself has been a truly great and enjoyable journey, and I really do not know how to ever thank you. In this respect, I have been by far the luckiest graduate student at MIT! And I truly mean every single word I wrote.

I would like to also thank the Volpe team in particular, Dr. Stephen Popkin and Dr. Heidi Howarth for editing this work and for the invaluable discussions we held as the project progressed and took shape. A million thanks to Mr. John K. Pollard, a most skillful and fine engineer. Certainly, without your help and experience, the Cab Characterization Study could have never been accomplished. You taught me a lot about everything school does not teach about real-life applications: using practical devices and making measurements in a non-laboratory/field setting. Thank you very much for all your help; it was a pleasure working with you! Matt Isaacs, Drew Kendra, and Dr. Michael Zuschlag have also been of great assistance at the various stages of the project. I am also very obliged to our engineer subjects and to their railroad companies (Amtrak, Mass Bay Commuter Rail, and CSX).

I am grateful for the financial support from the U.S. Federal Railroad Administration who funded the project through Department of Transportation Volpe Research Center Contract DTRS57-04-Q-80164 PR 79-3354, and for Canada's National Sciences and Engineering Research Council Postgraduate Scholarship.

To all my labmates, the MVL'ers, I would need one more page just to thank you all for making my life as a grad student exciting. I learned so much from every single one of you, both socially and intellectually. The only part of my Masters I regret is not having the chance to collaborate and interact with any of you at the research level. Nevertheless, thanks for the great times!

I must really thank my parents for their full and unconditional support. They have always encouraged me and guided me on the road of life. Although I am quite certain they will not read this thesis (or even this page), I would like to thank them for all their sacrifices and for being fantastic pillars of support.

Finally, I wish I could tell my grandma, Marie, that I finished my Masters and am starting a PhD; I learned that she passed away just after I learned that I passed the Quals. Her steady encouragement and loving support over the course of my life will always be remembered.

Anton

[This page is intentionally left blank]

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Physiological and Psychological State Estimation: A General Problem

Fighting drowsiness or falling asleep while operating a vehicle inevitably leads to accidents, property damage and death. Monitoring operator drowsiness level and judiciously issuing alerts can aid in reducing catastrophic outcomes in all areas of transportation – including earth, air, sea and space – hence saving millions of lives. This thesis will focus on fatigue management in locomotive operations and will propose a systems approach.

Certainly, the use of instrumentation to detecting or estimating a particular physiological or psychological state in a human subject is a broad problem repeatedly encountered across various disciplines, not just transportation. Therefore, the ideas emanating from this work are readily applicable to a wealth of practical problems in other fields – including medicine where, for instance, one may seek to automatically monitor a patient for arrhythmias, or for other diverse medical conditions.

1.2 Challenges of Fatigue/Drowsiness State Estimation

Before proceeding any further, a discussion about the difficulty of detecting fatigue or drowsiness in humans in operational settings is in order. The National Aeronautics and Space Administration (NASA)'s Fatigue Resource Directory (1995) defines fatigue as the experience of being 'sleepy', 'tired' or 'exhausted'. The US Department of Transportation (DOT, 1999) Human Factors Coordinating Committee elaborates on this, calling fatigue "…a complex state characterized by a lack of alertness and reduced mental and physical performance, often

accompanied by drowsiness." Crawford (1961) outlines two main sources of fatigue in transportation, namely:

1) Operating the vehicle itself (particularly for extended periods of time)
2) Other sources including sleep loss, disruptions in the biological clock, and other physical or mental work accomplished before operating the vehicle.

However, defining fatigue has historically been a notoriously difficult task (Hancock & Desmond, 2001); the literature has many different definitions and there does not seem to be a consensus about the multi-dimensional nature of the phenomenon. One major difficulty is the absence of a simple marker to measure. The signs and symptoms, summarized in Table 1, are numerous and include physiological and psychological changes that may be assessed subjectively or objectively. Unfortunately, a number of them (e.g. preoccupation, irritability, depression, attitude changes) are not unique to fatigue, and most changes do not occur until severe levels of fatigue are encountered, thus making early detection significantly challenging. The ability of a device to predict the occurrence of fatigue will likely determine its viability and value. If a sensitive and reliable early detector of fatigue could be provided, operators could change their behavior or use certain countermeasures before the occurrence of severe fatigue events.

Table 1 – Effects of Fatigue (IMO, 2001)

PERFORMANCE IMPAIRMENT	SIGNS/SYMPTOMS
1) Inability to concentrate	• Unable to organize a series of activities • Preoccupied with a single task • Focuses on a trivial problem, neglecting more important ones • Reverts to old but ineffective habits • Less vigilant than usual.
2) Diminished decision-making ability	• Misjudges distance, speed, time, etc. • Fails to appreciate the gravity of the situation • Overlooks items that should be included • Chooses risky options • Difficulty with simple arithmetic, geometry, etc.
3) Poor memory	• Fails to remember the sequence of task or task elements • Difficulty remembering events or procedures • Forgets to complete a task or part of a task
4) Slow response	• Responds slowly (if at all) to normal, abnormal or emergency situations.

5) Loss of bodily control	• May appear to be drunk • Inability to stay awake, drooping eyelids • Affected speech e.g. it may be slurred, slowed or garbled • Feeling heaviness in the arms and legs • Decreased ability to exert force while lifting, pushing or pulling • Increased frequency of dropping objects like tools or parts
6) Mood change	• Quieter, less talkative than usual • Unusually irritable • Increased intolerance and anti-social behavior • Depression
7) Attitude change	• Fails to anticipate danger • Fails to observe and obey warning signs • Seems unaware of own poor performance • Too willing to take risks • Ignores normal checks and procedures • Displays a "don't care" attitude • Weakness in drive or dislike for work

Electroencephalography (EEG) has been extensively used in clinical sleep laboratories. Changes in EEG frequency patterns have been correlated with different sleep stages; however, with the current state of the art, the method is encumbering, invasive, sensitive to vibration and motion, and does not represent a practical solution for a field (non-laboratory) vehicle setting. In sum, the difficulty in detecting fatigue states stems from the fact that there exist many signs and symptoms that *collectively but not uniquely* characterize it. Individual effects may be easily confounded with other phenomena unrelated to fatigue. However, slow eyelid closure (i.e. droops rather than blinks) over a certain time period seems to be a promising metric for use within a larger systems approach and will be the focus of this thesis.

1.3 Fatigue Problems in Railroad Operations

Operating a locomotive is a task that is inherently different from driving a car or flying an aircraft. Much of the operating knowledge is acquired through lengthy training on a particular railroad line governed by an extensive set of operating rules. Such preparation is intended to promote the acquisition of a mental model (Roth, unpublished report) of the terrain and of its intrinsic characteristics, including grade, curvature, mileposts and other permanent physical landmarks, speed restrictions and a strategy or plan for driving the route that meets arrival time requirements. Mental models may be temporarily updated by means of written orders before any

given trip, or in real-time via knowledge gained from out-the-window, radio communication and in-cab signals. Essentially, given the composition of the consist[1] (weight and length), the driving task involves the use of the mental model to make appropriate adjustments to the train throttle, brake and other systems. However, a train operator – traditionally called a 'locomotive engineer'[2] in the US – must make control adjustments well before signals, signs, and landmarks actually become physically visible from the cab. This requirement stems from the large mass[3] and resulting inertia of the rail vehicle, which – along with the initial speed – often dictate large braking and acceleration distances. These distances are orders of magnitude larger than the ones required for an automotive vehicle to brake. For example, it takes about 2 miles to bring a commuter train to a complete stop from an operating speed of ~ 50 mph in flat terrain; the braking distance may be greater if the rails are slippery due to bad weather or leaf litter. Moreover, operating a train at high speeds while negotiating a curve or going down a grade is potentially extremely hazardous. Inappropriate control actions can lead to a derailment or breakage of the consist. Indeed, careful energy management is of prime importance to sustain safe railroad operations. To achieve proper energy management, the operator needs to attain and maintain a correct mental model of the environment at all times, i.e. the operator must continuously maintain a proper level of situational awareness. Inattention, confusion, incapacitation and, especially, being asleep or drowsy may lead to errors and serious accidents. If a signal is missed as a result of drowsiness, it may be physically impossible for an engineer to stop the train soon enough from high operating speeds to avoid a collision or signal past at danger (SPAD, where a train disobeys a stop signal).

1.3.1 Problem History

Actions addressing concerns about the effects of fatigue and drowsiness in the United States railroad operations date back to the early 1900's. The reader is referred to Gamst (in press) for further details about the working conditions and practices at the time. Congress enacted the 'Hours of Service Act' in 1907 (most recently revised in 1969) to increase railroad safety by

[1] Interconnected set of rail vehicles which make up the train

[2] For simplicity, in the rest of this document, 'engineer' is used interchangeably with 'locomotive engineer'

[3] The mass of typical freight vehicles may easily rise up to hundreds and sometimes more than a thousand tons

prescribing restrictions on the number of working hours and the duration of off-duty time for the operating crafts, including locomotive engineers. “The basic standard is 8 consecutive hours off duty in the preceding 24 hours, or 10 consecutive hours off duty after working 12 consecutive hours (codified as 49 U.S.C. 21101(a)). Such regulations are not scientifically supported and in light of present knowledge are incompatible with the attributes of the human’s biological clock, the cumulative effects of work on consecutive days, and the long-term health consequences of different work/rest schedules (Sussman & Coplen, 2000). Actually, it was 65 years after those rules were established that a study, first of its kind, was conducted to examine the effects of fatigue on train crewmembers (Grant, 1971). Kogi and Ohta (1975) followed with a landmark study that associated railroad shiftwork and nightwork to drowsiness-related human errors that led to accidents and near-accidents. Since then, diverse studies in railroad operations have identified sleep deprivation, caused by long hours on duty as well as irregular and unpredictable schedules as contributor to the buildup of fatigue in crewmembers (Pollard, 1991; Buck & Lamonde, 1992; Sherry, 1998; Boivin, 2000). Sleep disorders such as sleep apnea[4] have been identified as a potential cause of fatigue and are currently being investigated (Philip, 2005).

Schedule irregularity and unpredictability have continuously raised much concern among the affected crew, although these practices have financially benefited both the employer and the crews themselves. Most passenger locomotive engineers have a fixed schedule communicated to them months in advance. As a result, they know when to report to work and could hence schedule other aspects of their life, adjusting their sleep-wake schedule accordingly, even if they were to work during the night. Although the risk of being drowsy may occur in both passenger and freight operations, the problem is much more prevalent in freight[5] where night and extra-board[6] operations predominate and work schedules are generally uncertain. Those types of operators typically work on a first-in first-out (FIFO) rotation, in which trying to predict the next day’s schedule is extremely challenging, and the following day nearly impossible (Gamst, 2005). Starting times may occur at any time and have been characterized as "erratically variable",

[4] Sleep apnea is a disease causing patients to repeatedly stop breathing during their sleep. The brain briefly wakes up the patient at each apnea event to resume breathing. As a result, sleep is very fragmented and of poor quality.

[5] Transportation of goods (non-passenger).

[6] Substitute operator: they are called to work only when the regular operators go on vacation, report sick, or are absent for personal reasons.

sometime regressive – starting earlier each new day – and being totally out of synchronization with the human biological clock (Gamst, 2005). A sample schedule is illustrated in Figure 1. The dark grey rectangles span the hours spent at work, while the light grey rectangles span the hours spent in bed. The remaining white spaces include other personal daily activities. Starting on Day 2, a little after 1:00 A.M., after only ~ 5 hours of sleep, the engineer is called to work. There are no two consecutive days (except for Days 13 and 14) on which the engineer starts work at the same time. On working days, duty starts later than the previous days, thus not permitting for a regular sleeping pattern. On days 8 & 9, and 13 & 14 the working hours conflict with the previous days sleeping hours. As a result, the engineer's biological clock is not allowed to get into a steady rhythm and severe buildup of fatigue ensues.

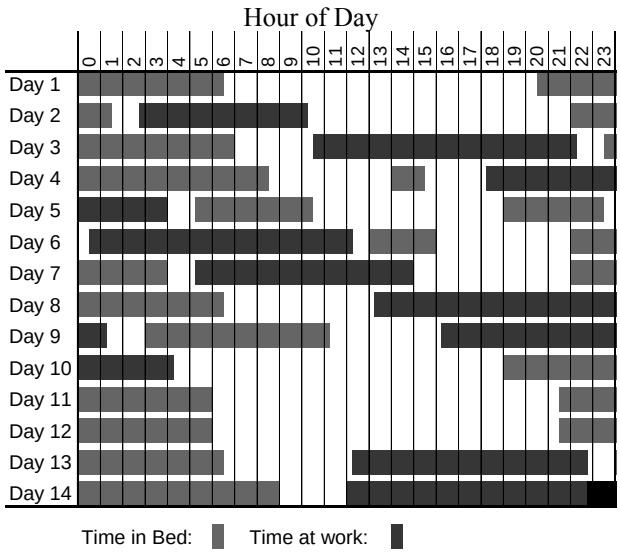

Figure 1 – Representative work schedule of a US freight locomotive engineer highlighting its irregularity and unpredictability (Karnali, 2004)

Drowsiness represents a serious problem in such irregular operations, but may arise in other types of operations, and not solely at night. The resulting impairment in alertness and vigilance prevents the crew from properly accomplishing its duties, and poses a serious threat to their own, their coworkers' and the public's safety. It is believed that these issues will remain present if no scientifically-based approaches (regulatory or otherwise) are implemented to measure and reduce operator fatigue. Meanwhile, the drowsiness issue must be addressed in parallel, since regulations alone are not expected to solve all facets of the problem.

Therefore, in its latest Strategic Research, Development and Demonstrations effort (Federal Railroad Administration, 2002), the Federal Railroad Administration (FRA) stated its intent to promote crew alertness and vigilance through its Human Factors program. This endeavor aims to use non-invasive real-time technologies to assess and feedback the operator's alertness level to allow him (or her) to change his behavior and therefore reduce the risk of substandard and hazardous performance. Such instrumentation is required since operators are often poor judges of the effects of fatigue on their performance; they are unaware of their performance decrements which become accentuated with increasing drowsiness levels (Horne & Reyner, 1995; Wylie et al, 1996; Mallis & Deroshia, 2005). Currently there are few countermeasures available to locomotive engineers, those being: changing operators, stopping to sleep and using alertness enhancing drugs such as caffeine. A device that properly monitors their alertness level may 1) provide the opportunity to correctly time available countermeasures, and 2) provide an additional countermeasure through a properly crafted feedback loop.

1.3.2 Existing Solutions

The alerting technology currently deployed on board locomotives is not infallible and leaves room for improvement. For instance, the Dead Man Pedal (also known as a Pneumatic Foot Valve) shown in Figure 2 requires the engineer to apply constant foot pressure onto a pedal mechanism. If the engineer releases the pedal, a warning whistle sounds. If the whistle is ignored for more than 7 seconds, the engine is shifted to idle and an automatic and irreversible full brake application is made until the train is brought to a complete stop – a condition called "penalty braking"; therefore, an engineer waking up during a penalty brake cannot cancel it. The consequential muscular fatigue makes it unpopular among engineers. Moreover, placing a brick or any heavy object on the pedal can easily disable the mechanism.

Figure 2 – Dead Man Pedal (Vossloh Kiepe, 2005)

A French adaptation of the Dead Man Pedal, the Veille Automatique par Contrôle du Maintien d'Appui (VACMA) system, requires the engineer to periodically press and depress the pedal; although it presents a more sensitive measure of drowsiness, its interference with the engineer's tasks increases workload and makes it unattractive as well.

A variation is a simple panel mounted timer switch that the engineer must press to reset the mechanism within a specific period of time. A failure to reset the timer leads to a flashing visual signal followed by an auditory signal, followed by penalty braking. In order to allow enough time for braking, the acknowledgement time for the visual signal had to be kept short. In early systems, the time period was fixed and easy to predict; however, the time was short enough so that the visual signal went off frequently and the timer task was viewed a nuisance.

In an attempt to reduce the false alarm rate, a new class of alerting systems called Electronic Alertness Controls was developed that attempted to utilize information from the engineer's normal activities to reset the timing mechanism. An early system monitored inductance changes produced by engineer movement in a copper mesh antenna mounted in the engineer's seat. Detected changes reset a 20 second timing mechanism. If no motion was detected during the timing period, an audible and/or visible signal was made. If no motion was detected during an additional period, nominally 10 seconds, the engine was reduced to idle and penalty braking was

automatically applied. Train Sentry I (Pulse Electronics, Inc/Wabtec Railway Electronics) utilizes activity detection switches on throttle, independent, train and dynamic brakes, horn, bell, sander (and in some systems, lights and windows) to reset the timer. Train Sentry II & III improved the design by adjusting the acknowledgement time window to be an inverse function of the train's speed, requiring about one reset per mile, and disabling it at very low speeds. In fact, Train Sentry III operates on a fixed interstimulus interval, subject to train speed.

$$\text{Time Interval (sec)} = \frac{K}{speed(mph)} \quad \ldots(1.1)$$

above speed X, where X may be set from 15 to 60 mph and was usually 30 mph.

For example, to obtain 1 reset per 1 mile traveled,

$$\text{K} = 1 \text{ mile x } 3600 \text{ sec/hr} \quad \ldots(1.2)$$

and thus,

$$\text{Time Interval} = \begin{cases} \dfrac{\text{K}}{30 \text{ mph}} = 120 \text{ sec}, & \text{speed} \le 30 \text{ mph} \quad ...(1.3a) \\ \dfrac{\text{K}}{\text{speed (mph)}}, & \text{speed} > 30 \text{ mph} \quad ...(1.3b) \end{cases}$$

Moreover, Train Sentry III visual and audio alerts ramp up in intensity. Currently, several manufacturers produce activity-based monitors similar to Train Sentry. Each might differ by a minor feature (number/type of switches to reset the timer, interstimulus duration and speed dependency, etc); nonetheless, the main alerting concept here described remains the same.

Figure 3 – Train Sentry III in an F-40 passenger cab

Because of concerns about sleepy drivers, particularly in single engineer/off circadian cycle operations, many US and European railroads have installed activity monitors. Activity based alerting systems functionally similar to Train Sentry are offered as original equipment by several locomotive manufacturers. However, engineers still see them as a nuisance. Despite the monitoring of various activities, in practice engineers need to hit the reset switch so regularly[7] that the response becomes automatic – so automatic that engineers still push the switch when drowsy or in a light sleep. There have even been anecdotal reports that engineers make button pushing movements while at home sleeping in bed (Haworth et al. 2000). Also, in practice, detection switches fail, and certain types can be defeated by mechanical means such as by simply hanging a set of keys on the reset switch (Pollard, personal communication).

Clearly, the existing alerting technologies are unsuitable. More reliable alertness monitors with a lower false alarm rate are needed or perhaps a superior approach for their overall design is required.

[7] Engineers must respond to the alerter's first level alarm once per mile, or about once every minute at track speed, if no other control adjustments are made.

1.3.3 Latest Efforts

Over the past few years, the John A. Volpe National Transportation Systems Center (Volpe Center) has been working with the Massachusetts Institute of Technology (MIT) as part of the FRA effort to evaluate novel vigilance monitoring techniques[8] (Oudonesom, 2001) and to study performance and alertness of one vs. two engineer crews (Karnali, 2004); on-going evaluation is also being conducted with the developers of several prototype alternative alerting devices. At present, testing (Oudonesom, 2001, Pollard et al. 2003, Karnali, 2004) revealed that these alerters are ineffective as standalone systems, partly because they were not specifically designed for use in a locomotive environment. To illustrate the situation, one may consider the following two examples.

A capacitive head movement monitor is shown in Figure 4. This system operates on the main principle that the change in sensor electrode plate mutual capacitance correlates with the head's distance from the plate. Therefore, by using three sensor plates, one may obtain the three-dimensional head position by triangulation. The use of some algorithms may reveal instances when the driver nods off (Kithil et al, 2001).

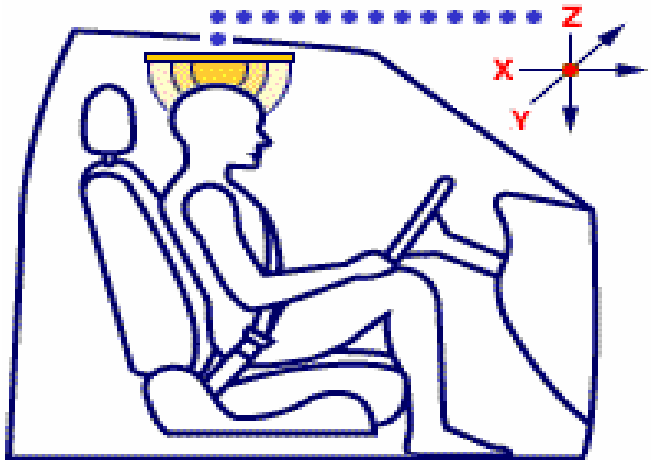

Figure 4 – Proximity Array Sensing System. Changes in capacitance are correlated with the head's distance from the sensor plate. Three-dimensional head position can be obtained by triangulation from the array of sensors fixed to the car's roof (Kithil et al, 2001).

[8] Reaction time, fine motion, gross motor activity, electro-dermal activity, pupilometry, ocular motion and other physiological and performance-based measures.

Although a system of this kind may possibly work in an automotive setting, it may not work at all in a locomotive environment, where the head-ceiling distance is considerably larger, and where the locomotive engineer, unlike a car driver, often gets up and moves around the cab, outside of the capacitor's correct operating range.

The second example is an eyelid closure monitor, initially developed for the automotive industry. Wierwille, et al (1994) showed that a degradation in simulated driving performance was associated with the percentage of time the driver's eyes were closed, measured over an appropriate time interval using a video camera; he coined this measure "PERCLOS" for percent eyelid closure. Dinges, et al (1998) evaluated PERCLOS and other measures, and showed that PERCLOS predicted fatigue-induced lapses in vigilance. As a result, the Federal Highway Administration and the National Highway Traffic Safety Administration (NHTSA) regarded PERCLOS as the most promising real-time measure of driver alertness for in-vehicle systems (Sherry, 2000). Grace, et al (1999) tested and marketed a PERCLOS monitor, CoPilot®, shown in Figure 5 for use in nighttime trucking operations. The device uses an infrared (IR) sensitive camera to image the driver's head, illuminated by IR light emitting diodes (LEDs). The LEDs flash alternately at two different wavelengths, 850 and 950 nm. The face reflects both wavelengths equally, but the interior of the eye absorbs more of the 950 nm wavelength, so if two successive camera images are subtracted from one another, facial features approximately cancel, but the pupils of the eye show up as bright spots. Thresholding techniques are then used to determine whether the eyes are open or closed depending on the pupil's image size. The resulting binary signal is then averaged over 1, 3, or 5 minutes to produce the PERCLOS value. If the PERCLOS level exceeds certain thresholds, a combination of auditory and visual alerts (green, yellow, and red color spectrum) is issued.

Figure 5 – CoPilot PERCLOS monitor for trucking operations

However, despite the face validity of the PERCLOS metric *per se*, several problems were encountered when attempting to use the CoPilot instrument in the locomotive cab environment. With the cooperation of the developer, Attention Technology Inc., preliminary simulator tests at Volpe (Karnali, 2004) showed that CoPilot was not ideally suited for use in the locomotive cab because of inherent differences between the locomotive engineer and the automobile/truck driver tasks and the physical operational environment. The instrument worked only in darkness, since broad spectrum sunlight IR overwhelmed the signal from the illuminators. Even at night, the operator had to be within about an arm length of the apparatus, due to the illuminator's power limitations. The operator's gaze had to be within approximately 50 degrees of the camera axis. The operator could not wear eyeglasses as the instrument saw reflections caused by lenses and frames as open eyes/pupils. Also, the device could not distinguish between "eyes closed" and "no-eyes/head-in view" conditions. Such design limitations and measurement errors limited the reliability and utility of the CoPilot alerter.

In summary, over the years, the consistent lack of success in the development of reliable fatigue management solutions for the locomotive environment indicates that a new line of attack needs to be developed.

1.4 Thesis Goals

The design of locomotive fatigue management technologies is addressed in this effort. To this end, a systems-level approach was used.

1.4.1 Specific Objectives: Thesis Outline

The development of the systems approach followed several steps that took the form of distinct chapters, each focused on a specific goal:

1) **Chapter 2:** Introduce a systems solution design framework.
2) **Chapter 3:** Characterize passenger and freight locomotive physical environment and engineer behavior, particularly look-away and go-away head movements pertinent to image-based eye closure detectors.
3) **Chapter 4:** Evaluate the performance of Attention Technology's most current infrared eye closure monitor prototype, Model DD-850, a version of CoPilot developed for locomotive cabs with DOT-FRA support. Verify whether its performance is a good match to the locomotive physical environment and engineer behavioral characteristics.
4) **Chapters 5 & 6:** Employ Signal Detection Theory (SDT) to develop a stochastic simulation tool to:
    a. Evaluate the effectiveness of various alternative fatigue management system architecture designs:
        i. **Activity Monitor standalone:** Quantitatively evaluate performance of a simulated generic "Train Sentry" class activity monitor.
        ii. **Eye Closure Monitor standalone:** Assess whether devices with performance comparable to the DD-850 are appropriate as a replacement for Train Sentry class detectors in a locomotive cab environment.
        iii. **Tandem Detector designs:** Propose the design of a tandem architecture utilizing both types of detectors working in parallel.
    b. Evaluate the system's component parameters to obtain acceptable or optimal overall drowsiness detection performance.

c. Support the design, analysis, and optimization of future alertness monitoring devices as individual functional components and as a part of a multi-detector system.

5) **Chapter 7:** Develop a set of requirements to guide future research and design of novel solutions for the locomotive cab environment.

Chapter 2

Systems Approach

The previous chapter outlined the inherent challenges faced in detecting drowsiness in railroad operations. The current chapter sets the stage for a conceptual solution, outlining a broad systems approach to the analysis and design of alertness monitors. This chapter also presents brief tutorials of systems engineering, quantitative model based design, hidden state estimation, signal detection theory, stochastic modeling, and Markov processes. After reading section 2.3.2, engineering oriented readers familiar with these concepts may choose to skip directly to Chapter 3.

2.1 Systems and Systems Engineering

First, it is valuable to define what a system is and what systems engineering signifies. Although these words might be widely and loosely employed in everyday parlance, a *system* may be defined as "a set or assemblage of things connected, associated, or interdependent, so as to form a complex unity; a whole composed of parts in orderly arrangement according to some scheme or plan" (OED, 2005). Often, the complexity of a system depends on the interactions and dependencies between its components, which may also be systems in their own right. Therefore, to produce a sound design, a system-wide view is required to understand how each component operates and interacts with other components as a part of one whole body aiming to accomplish a common purpose; to meet the various requirements, it is crucial that trade-offs and design decisions be evaluated at the overall system level. One may term such an approach *systems engineering* (SE), but the definitions of SE are many and varied. For example, Griffin & French (2004) define it as "the art and science of developing an operable system capable of meeting [the] requirements within imposed constraints…" For a more detailed definition, one may turn to

NASA, an organization where systems engineering is its core activity. The NASA Systems Engineering Handbook (1995) draws from the experience of engineers who have had considerable experience working on large and complex engineering problems and offers the following definition: "systems engineering is a robust approach to the design, creation, and operation of systems." This discipline entails the following:

- Identifying and quantifying system goals
- Creating alternative system design concepts
- Performing design trades
- Selecting and implementing the best design (balanced and robust)
- Verifying that the design is actually built and properly integrated according to specifications
- Assessing how well the system meets the goals

Perhaps the central points of SE are that it is a broad and multidisciplinary engineering discipline requiring an understanding of the different components and how they operate in concert as one entity. Moreover, one must employ quantitative metrics of overall system performance, and then conduct trade studies aimed at optimizing the value of the whole system, and not the subsystem alone as is typically done.

2.2 Systems Solution

From an SE standpoint, the design of an alerting system can be considered as a signal detection or stochastic estimation problem. The main concept presented here is the use of an estimation theory framework to combine information across relatively noisy detectors operating in tandem – each exploiting a different modality, for example slow eyelid closure and engineer activity (often called "motor activity" by physiologists) – can be used to isolate the common drowsiness component and obtain an improved estimate of the operator's state. This approach is analogous to investment portfolio optimization using statistically based asset allocation strategies. For instance, risk may be quantified as the variance of a security's expected return (Markowitz, 1991). Therefore, by selecting securities whose performance is uncorrelated with each other (or

by filtering out the correlated components) one can reduce the overall portfolio variance to levels that are smaller than the variances of the individual investments.

2.2.1 System Architecture

Figure 6 illustrates the overall alerting system (AS) architecture. The simple block diagram is intended to provide a high-level cartoon view of the proposed system solution and illustrate the interaction between the main components. The system designer seeks to estimate the operator's internal drowsiness state to assess the operator's ability to drive safely. The operator's drowsiness state is assumed to be "hidden" – i.e. there is no way to directly measure it. Instead the designer estimates the hidden drowsiness state using e.g. a DD-850 eye closure monitor and a Train Sentry class activity monitor to respectively measure physically observable states that depend in part on the hidden drowsiness state. The designer then judiciously combines the measurements using state estimation logic, here referred to as an arbiter, and thereby estimates the engineer's drowsiness state. In general, alerting systems could employ additional or different types of sensors, but the main design approach essentially remains the same.

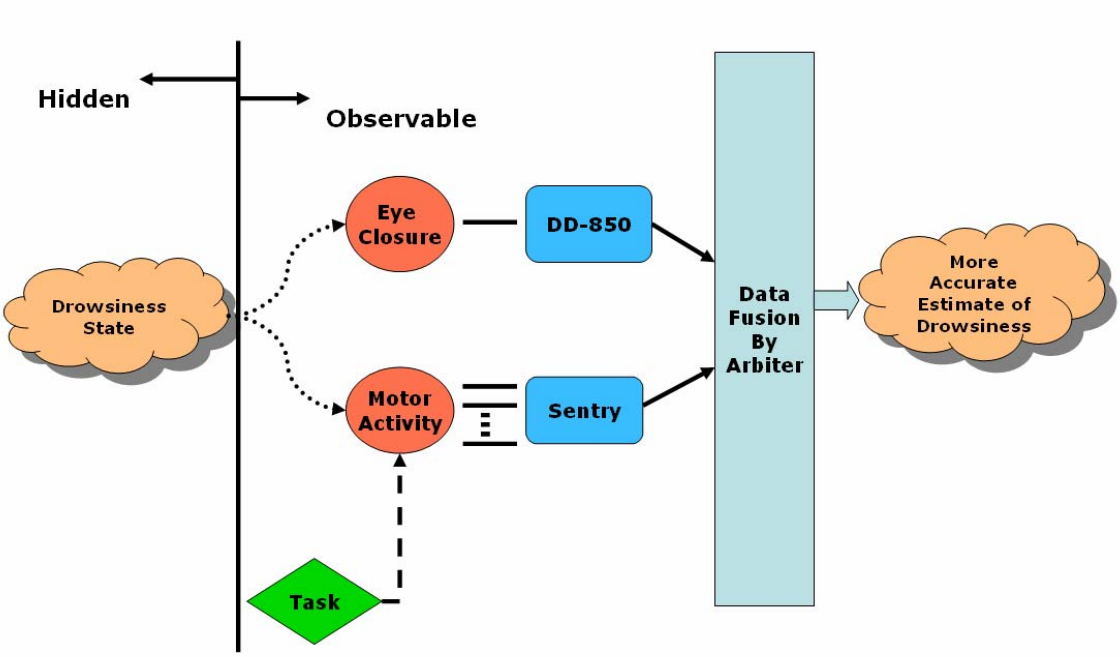

Figure 6 – Architecture of an Alerting System (AS) solution

2.3 Approach Outline

2.3.1 Environmental Characterization

A key step in the design procedure is to first characterize the environment in which the AS is expected to operate. Specifically, one seeks to understand the interaction between the AS and its environment, identify the important relevant variables that affect detection performance, and quantify those. As shown in Figure 7, the AS's environment as such is not limited to the physical setting but also includes the human operator. Therefore, in this case, one must characterize the pertinent physical attributes of the operator and of the locomotive cab, as well as the operator's tasks and behaviors, which are the result of the human-vehicle interaction.

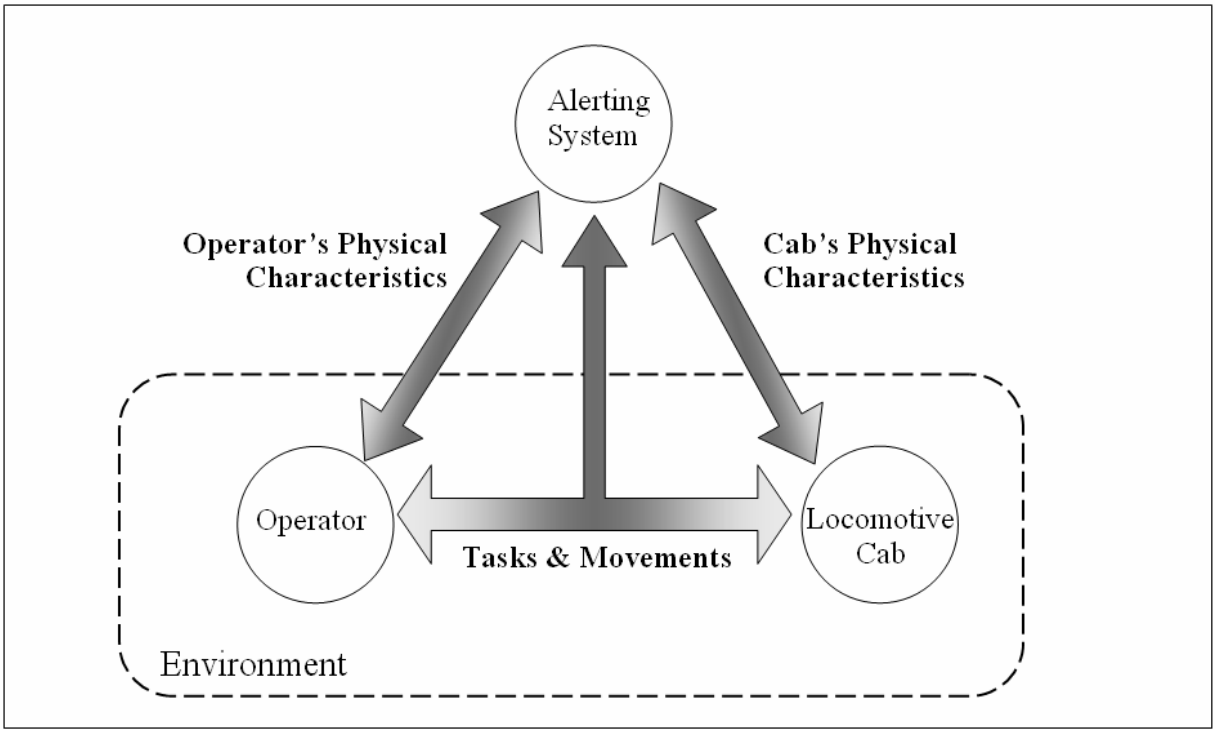

Figure 7 – The interaction between the alerting system and its environment, which encompasses the operator, the locomotive cab and the interaction between those two entities.

However, in this example, the AS is a 'system of systems', where each is composed of a different type of sensor (an image based engineer eye closure monitor, such as the DD-850 and the Train Sentry class activity monitor), software and hardware components (i.e., the arbiter and its logic). Therefore, one must appreciate the specific factors that affect each component's

behavior and performance, and include those in the environmental characterization study as well. The subsystems and final system can in the end be developed in accordance to the specifications derived from the collected environmental data.

2.3.2 Subsystem Characterization

Adopting a top-down approach, the next step involves identifying the strengths and limitations of the available system components and understanding how they operate at the individual level. The main purpose is to assess the subsystems' empirical performance characteristics and determine whether they can ultimately be standalone elements and/or whether they fit as elements of the overall system. Specifically, the performance of Attention Technology's DD-850 prototype shown in Figure 8 was evaluated (Chapter 4) as an example of an image based eye movement monitor.

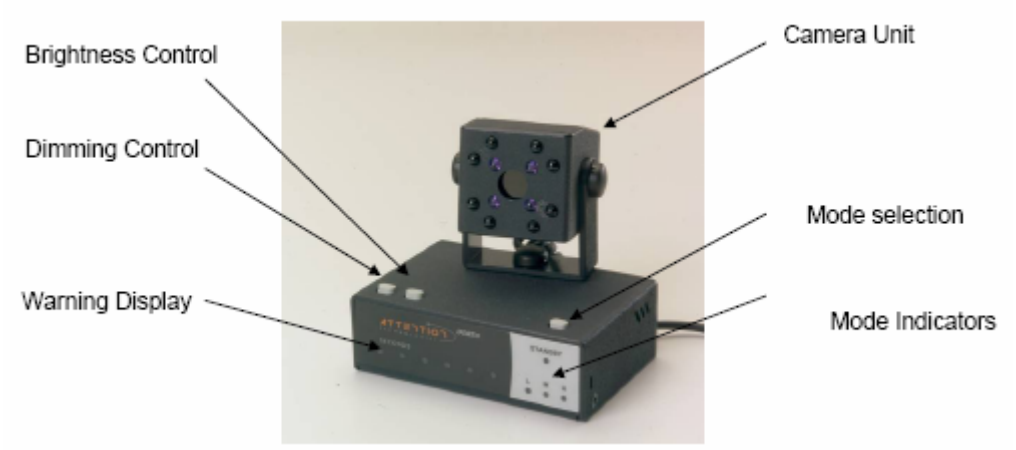

Figure 8 – DD-850: Attention Technology's latest infrared PERCLOS monitor prototype.

There are several different kinds of Electronic Alertness Controls that monitor engineer physical actions made by different locomotive and after-market manufacturers. A variety of different types are currently in use in the field, and it is impractical to test them all. Therefore, for purposes of this study, a generic "Train Sentry" class activity monitor was defined and used in the simulation. Eventually a more elaborate or model specific simulation could be built. Capturing the first-level characteristics of a generic activity monitor similar to Train Sentry was deemed sufficient for present purposes, as described in section 2.3.4 below.

As far as we have been able to determine, quantitative techniques are not currently being employed to assess performance of either image or activity based locomotive alertness monitors. Existing devices are being typically designed and developed apparently in relatively impromptu fashion with no formal testing of sensitivity and reliability. The problem of a detector system deciding in presence of uncertainty as to whether an operator is awake or asleep may be approached from a Signal Detection Theory (SDT) viewpoint. An alerter's performance may be conveniently characterized using such a mathematical framework; the advantages of employing an SDT modeling approach are the precise language and the pictorial notation for analysis, as well as the quantitative basis that permit assessment of the performance of drowsiness monitors and issue of guidelines and recommendations (Kuchar, 1995). Nevertheless, the alertness detection problem in this thesis is too complex to be *solely* addressed with classical mathematical SDT concepts. The detectors are nonlinear, the noise characteristics are non-Gaussian, and the monitors, whose dynamics are time-dependent, have multiple alerting levels. Therefore, using additional mathematical tools, an SDT approach was taken to assess the performance of a generic Train Sentry class activity monitor, of the DD-850 and then that of the whole system. Under ideal controlled bench test conditions, the DD-850 operating envelope was characterized in terms of its significant dependency on eye color, and presence or absence of head motion. Its performance was evaluated based on engineer behavior metrics determined in a field study, and deficiencies were reported to the manufacturer for evaluation.

A summary of the most important concepts of SDT relevant to this thesis are presented below.

2.3.3 Signal Detection Theory Framework

There are numerous applications in which a system must make a decision based on noisy observations. For example, in radar detection, the return signal is measured or 'observed' and the system tries to decide whether a target is present. The solution requires making decisions based on observations which are of random nature and often corrupted by noise. The theory underlying solutions to such problems has been termed decision theory, hypothesis testing or signal detection theory (SDT) (Shanmugan & Breipohl, 1988). The SDT model was developed in World War II to tackle radar problems and was then further applied to human decision-making in

the 1950s by Green & Swets, and continues to enjoy success in numerous present day applications.

In the context of this study, a DD-850 instrument deciding whether an operator has the eyes open or closed may be seen as a simple binary detector that makes one observation and one decision at a given time. For the current purposes, the DD-850 is considered an eye closure "detection" system, (although the current system is unable to distinguish an 'eyes closed' condition from a 'head absent' condition).

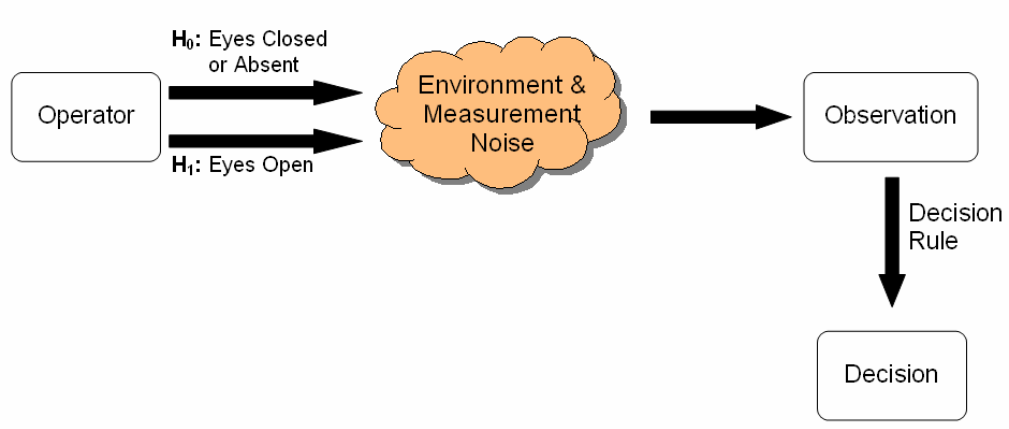

Figure 9 – Components of an SDT problem. Adapted from Van Trees (1968)

One may thus hypothesize that at any instant of time the operator has either

> ***H₀*:** Eyes closed or out of view[9]: suggesting that the operator is asleep
> ***H₁*:** One or both eyes open and in view: suggesting that the operator is awake

The observations corresponding to each hypothesis may be corrupted by noise[10] originating from the environment or from the instrument design limitations. For example, reflections caused by lenses, frames or objects in the background may be treated as eyes open; high IR levels may overwhelm the signal from the illuminators and eye reflections may not be detected. Therefore, the inherent uncertainties in the detection process results in four possible outcomes:

[9] due to the camera's limited field of view

[10] Basic SDT models employ a linear Gaussian noise probability distribution; however, in this case, such an assumption may not hold due to the complexity and nonlinearity of the instrument. Numerical simulation methods were resorted to as detailed further below.

H_0 true; choose ***H_0*****:** Correct Detection **(CD)**
H_0 true; choose ***H_1*****:** Missed Detection **(MD)**
H_1 true; choose ***H_1*****:** Correct Rejection **(CR)**
H_1 true; choose ***H_0*****:** False Alarm **(FA)**

Each response is assumed to have an associated quantitative cost or benefit associated with it in relevant units, e.g. dollars, human lives, etc. These four values are sometimes known as the "payoff matrix".

B_{00}**:** Benefit of a Correct Detection **(CD)**
C_{10}**:** Cost of a Missed Detection **(MD)**
B_{11}**:** Benefit of a Correct Rejection **(CR)**
C_{01}**:** Cost of a False Alarm **(FA)**

Using some basic definitions of probability, one may express $P(CD|H_0) = 1- P(MD|H_0)$, and $P(FA|H_1)= 1- P(CR|H_1)$. Therefore, given $P(H_0)$ or $P(H_1)$, system performance can be *fully* specified by *two* independent probabilities $P(FA|H_1)$ and $P(CD|H_0)$.

The results of an example experiment may be tabulated as follows:

Table 2 – Results of an SDT detection paradigm of a signal in presence of uncertainty

		True State of the Operator	
		H_0	H_1
State Reported by Device	H_0	**Correct Detection** $P(CD\|H_0) = 0.8$ (80% of the eye closures were correctly detected)	**False Alarm** $P(FA\|H_1) = 0.6$ (60% of the eye open were cases incorrectly reported as eyes closed)
	H_1	**Missed Detection** $P(MD\|H_0) = 0.2$ (20% of the eye closures were incorrectly detected as eyes open)	**Correct Rejection** $P(CR\|H_1) = 0.4$ (40% of the eye open cases were correctly detected)

The FAs and MDs are undesirable for they correspond to faulty responses. FAs may reduce the operator's confidence and trust in the system; MDs, on the other hand, may lead to accidents, property damage, financial costs, injuries, and loss of life. An ideal behavior would only produce the correct responses, CDs and CRs. However, in practice this is rarely realized, due to the

presence of noise in the detector; the ultimate result is the presence of FAs and MDs. One desires to maintain a balance between a low level of FAs and a high level of safety (low level of MDs). On the other hand, MDs and FAs cannot be concurrently eliminated; one must consider their respective costs and must trade off between their probabilities of incidence when setting a value for an alerting threshold[11]. Optimality of the threshold or decision rule will depend on the operation's end goal: whether one seeks to reduce costs, risks, or other variables of interest.

The performance of a detector can also be represented graphically as the plot of P(CD) against P(FA). Such curves are traditionally referred to as "Receiver Operating Characteristic" (ROC) curves and can be used to assess the changes of performance due to the joint effect of threshold location and sensitivity (d'). At constant sensitivity, a change in the threshold location amounts to moving the operating point (changing detection performance) while remaining on the same curve. As sensitivity is increased, the curve moves to the upper left corner region as shown in Figure 10. Therefore, an alerter desirably has its operating region in the top left corner of in ROC space, displaying a high CD with a low FA.

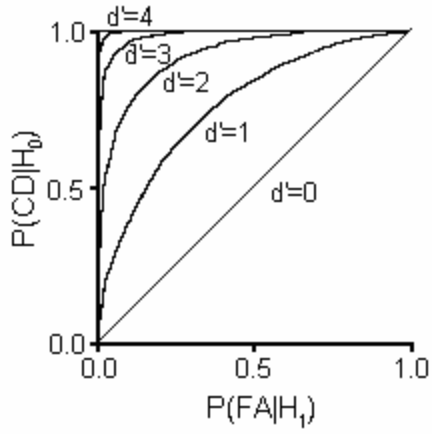

Figure 10 – Receiver Operating Characteristic (ROC) curves with increasing sensitivity, d' (Heeger, 2003).

The slope, β, of the ROC curve at a given point is a "conservativeness" measure of the decision making process; a steeper slope (operating point located in the lower left region) indicates a

[11] A particular level that the measurement of the operator's state must exceed for the device to produce an alert

conservative response while a shallower slope (operating point located in the upper right region) indicates a risky response. For instance, a conservative response would not readily produce alerts thus making few false alarms, though likely missing many of the signals. A liberal response would tend to produce more alerts hence detecting most of the signals but making many false alarms. Selecting an optimal strategy will depend on the probability of the signal's presence, P(H_0), and on the costs and benefits of the possible outcomes. If a signal is not frequently present, one must set the operating point to be more conservative, and vice-versa. If it is crucial that a signal does not go undetected, then high benefits are assigned to CRs and high costs to MDs, resulting in a low β, and the system operates further to the right on the ROC curve. Therefore, one does not only seek to maximize the number of correct responses, but also maximize the total expected costs and benefits associated with the decisions.

The expected value of a decision-making process may be expressed in terms of the following cost function (Wickens & Hollands, 2000).

$$J = \underset{\text{averaged over all outcomes}}{(\text{outcome probability}) \times (\text{outcome value})} \quad \ldots(2.1)$$

$$J = P(FA \mid H_1)(1 - P(H_0))C_{01} + P(CR \mid H_1)(1 - P(H_0))B_{11} + P(MD \mid H_0)P(H_0)C_{10} + P(CD \mid H_0)P(H_0)B_{00} \quad \ldots(2.2)$$

To optimize a detector system design, one can choose the operating point on the ROC curve so it minimizes or maximizes the latter cost function. Given linear cost and benefit functions, it can be easily mathematically shown that the optimal value of β is:

$$\beta_{optimal} = \frac{P(H_1)}{P(H_0)} \times \frac{B_{11} + C_{01}}{B_{00} + C_{10}} \quad \ldots(2.3)$$

For alerting system design, it is important to know the probability of drowsiness events, the probability that a drowsiness event causes a serious accident, and to determine what costs and benefits to assign to the four different possible outcomes. To our knowledge, no estimates have

yet been made for in-cab, real time alerting systems. However, Raslear and Coplen (2004) provide some rough estimates for costs and benefits associated with a locomotive engineer fatigue prediction tool that provide an illustrative framework for discussion. Values are shown in Table 3.

Table 3 – Payoff Matrix. Estimates of the costs and benefits of the different possible decisions of a fatigue model that would guide the decision as to whether an operator is fit or unfit to drive.

State	Diagnosis	
	Unsafe	Safe
Fatigued	B_{00} = \$220,000	C_{10} = -\$740,000
Not Fatigued	C_{01} = -\$100,000	B_{11} = \$10,000

Based on Pollard (1996), Raslear and Coplen estimated that P(experiencing fatigue or drowsiness event on a given trip) = 0.16. Pollard (personal communication), noted that this is an overall average. The probability may vary between approximately 0.08 for regular night operations and 0.33 for extra-board operations.

Raslear and Coplen estimated that for typical operations, a fatigue model must display a performance measure that is near the following

$$\beta_{optimal} = \frac{1-0.16}{0.16} \times \frac{10{,}000 - 100{,}000}{220{,}000 - 740{,}000} \approx 0.91 \quad \ldots(2.4)$$

However we believe Raslear and Coplen's (2004) payoff matrix estimates – though arguably valid for the case they considered - may not apply to the case of an in-cab alerting system, particularly in terms of the relative costs assigned to false alarms. This point will be further discussed in Chapter 6.

In summary, when using a signal detection framework it becomes possible to quantitatively assess alerting performance. For example, one may investigate whether a DD-850 class PERCLOS monitor is a viable standalone drowsiness monitoring system. If it is an inadequate

standalone detector, e.g. due to a high false alarm rate, it may be possible to employ the monitor in tandem with the activity monitor to increase the overall system's reliability. If it is possible to model both the DD-850 and a train sentry class activity monitor and make credible estimates of the payoff matrix, it should ultimately be possible to answer such design questions as:

- Can one maximize the strengths of multiple detectors while minimizing their weaknesses by judiciously combining them together? For example, reducing the false alarm rate by requiring that both the image based eye closure detector AND the activity detector system responses be positive.
- What types of detector arbiter logic is satisficing? Optimal?
- How does one design a system and its components to obtain the best overall drowsiness detection performance?
    - What design parameters make the overall system satisficing? Optimal?
    - How can future image based eye closure detector and activity detector designs be changed to obtain improved standalone and tandem performance?

At the system design stage, before actually building and assembling the overall system, one should be able to estimate the appropriate performance parameters and assess whether the given designs are likely to satisfy the top-level system performance requirements: whether the system can correctly detect drowsy operators with a certain probability.

2.3.4 The Case for Modeling and Simulation

In order to take this SDT based systems approach, the designer needs a quantitative simulation tool to emulate system performance. MATLAB and its system modeling subset, SIMULINK, (Version 7.0.4, Release 14, Mathworks, Inc., Natick, MA) were selected. Such a general-purpose software environment is readily available and allows rapid building and assembly of mathematical models and visualization of the overall system behavior.

A model is a simplified representation of a system (Wastney et al, 1999). Modeling a system is quantitatively describing its behavior to predict its performance over a range of operating conditions (Moir and Seabridge, 2004). At low cost, modeling enables rapid system analysis

under different circumstances that would otherwise be time consuming and sometimes impractical to emulate in real-life hardware. Modeling techniques are widely used in engineering to assess trade-offs between different potential solutions; one may quickly and effectively examine complex design options before committing time and monetary resources to developing the actual product. Most relevant to this case, models allow evaluation and prediction of system performance, hence substantiating a system's qualification, long before its final operating version becomes available. Moreover, at any stage of the development, bench tests and prototypes may aid in providing further evidence about the operation (Moir and Seabridge, 2004).

The first step in design involves developing engineering models of the DD-850, Train Sentry and arbiter logic in MATLAB/SIMULINK. The models are intended to be as simple as possible while capturing the essential characteristics of subsystem behavior. Much of the model operation depends on the effects of stochastic processes, including the roadbed, tasks, head and eye position, as well as measurement imperfections. The values of many model variables, which are certainly not deterministic, are not totally random either. Thus, one could describe them with probabilistic models. Although a characterization of this sort may not be accurate for all types of locomotives, operators, and times of day, it can usefully represent more limited situations of interest for purposes of system conceptual design and analysis.

2.3.5 Stochastic Modeling

The simulation developed in this thesis required the use of stochastic modeling techniques. For example, Markov stochastic models were used to represent the variability in engineer head movements, and in the roadbed the train passes over on a given trip. The main rationale for using stochastic components in a system model is that rather than conceding one's ignorance about some aspects of subsystem behavior, one characterizes the lack of knowledge using stochastic methods. Uncertainty or randomness is a ubiquitous element of all phenomena occurring in the physical world, including economic, engineering, and biological systems. Earlier scientists and physicists noted that repeating an experiment under seemingly identical conditions yielded results that were not totally reproducible (Shanmugan & Breipohl, 1988); examples include the

motion of planets, transmission and reception of radio signals in communication systems, and measuring biological processes. Probabilistic models have thus been devised to describe and understand the random nature of such experimental results and of a broad range of other general phenomena. In particular, Markov stochastic models are used in this alerter system design application.

2.3.5.1 Markov Processes

It is relevant to review a few basic notions about Markov processes. These basic introductory concepts will help the reader understand the analysis and modeling efforts in this work. The material presented here is synthesized from Papoulis & Pillai (2002), Van Trees video lectures (1971), Howard (1971), and Rabiner (1989).

Markov processes are named after Andrey Andreyevich Markov, who first developed this concept in 1907. Since then, they have gradually been widely applied in various fields, ranging from communications to ecology, mainly because they lend themselves to a mathematical model that is both tractable and easy to analyze. They are able to model a large number of physical situations in which the state of a process at a given moment of time is solely dependent on the state that immediately precedes it; this property is called the *Markovian condition*. At any time, a discrete-state Markov process can be described as occupying one of a set of N states $\{S_1, S_2, \ldots, S_N\}$, which may be finite or countably infinite. The process may move from a state S_i to a state S_j (where j may equal i) based on a certain probability, a_{ij}, termed *transition probability*. If q_t denotes the process state at time instants $t = 1, 2, \ldots$, the Markovian condition may be formally expressed mathematically as follows:

$$P[q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots] = P[q_t = S_j \mid q_{t-1} = S_i] \quad \ldots(2.5)$$

Where $P[q_t = S_i]$ indicates the probability of being in state S_i at an instant of time t. Thus, for this N state process, one may form the *state transition matrix,* $A_{N \times N}$, which contains the transition probability coefficients, a_{ij}, with the following properties:

$$a_{ij} = P[q_t = S_j \mid q_{t-1} = S_i],\ 1 \le i, j \le N \quad \ldots(2.6)$$

The row number i indicates the current state and the column number j indicates the future state. A state transition may or may not occur at a given instant of time, so

$$a_{ij} \ge 0 \quad \ldots(2.7)$$

and as a result, all rows of A must sum to 1.

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \ldots(2.8)$$

Finally, to fully characterize a Markov process, one must not only characterize the state transition matrix, A, but also the *initial state probability matrix*, Π, in which the process commences.

$$\begin{aligned} \Pi &= [\pi_1 \quad \pi_2 \quad \ldots \quad \pi_N] \\ \pi_i &= P[q_1 = S_i], \qquad 1 \le i \le N \end{aligned} \quad \ldots(2.9)$$

Example

To illustrate the main concept, an example adapted from Howard (1971) is used. Let a simple 3-state Markov model describe the behavior of a frog jumping on a set of water lily pads, with the frog's position being the stochastic process and the lily pads being the Markov states:

State 1: water lily with mosquitoes (food)

State 2: water lily with a crocodile nearby (predator)

State 3: plain water lily

Let the frog start in any state with an equal probability, i.e.

$$\Pi = [1/3 \quad 1/3 \quad 1/3] \quad \ldots(2.10)$$

Then, at every second, t, the frog has the option of jumping to another pad or remaining on the same pad with certain probabilities as described by the transition matrix

$$A = \begin{bmatrix} 0.8 & 0.05 & 0.15 \\ 0.6 & 0 & 0.4 \\ 0.55 & 0.05 & 0.4 \end{bmatrix} \quad \ldots(2.11)$$

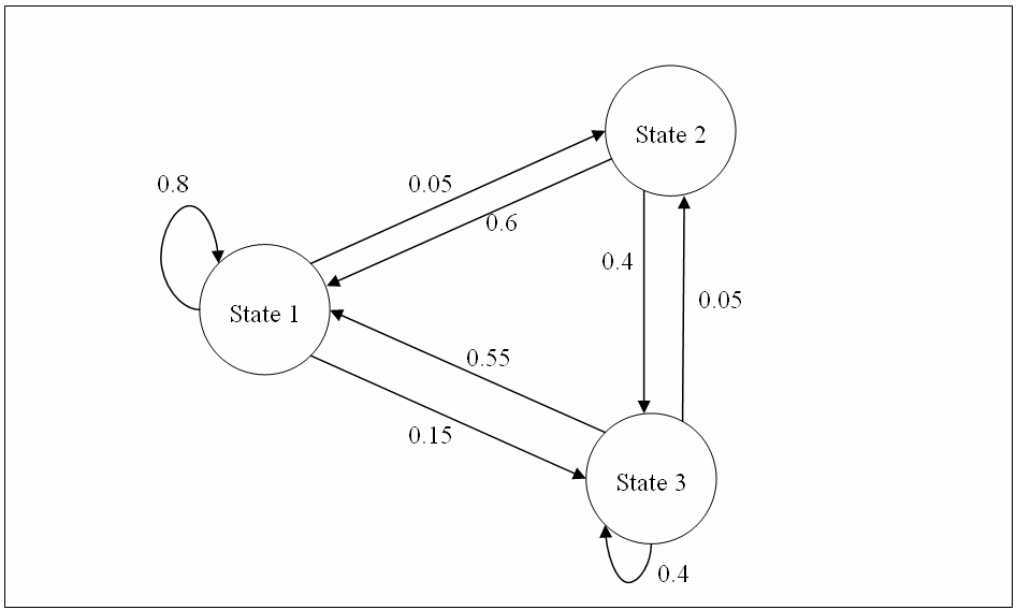

Figure 11 – Illustration of the use of a Markov chain to model the behavior of a frog jumping on a set of lily pads.

Assume that the frog moves (executes a state transition) depending on how hungry or how safe he feels. For example, if the frog starts in state 1, then there is a high probability that he will remain there and eat the mosquitoes ($a_{11} = 0.8$) until he satisfies his hunger. At a later moment, he may decide to move to either state 2 ($a_{12} = 0.05$) or to state 3 ($a_{13} = 0.15$). If on the other hand the frog finds himself in state 2 (row 2) facing the crocodile, he will not stay there ($a_{22} = 0$) but will jump to either state 1 (where food is) or to state 3 (the plain lily pad), with a higher probability to the former state.

Of course, complex and more basic examples of this sort may be described using a Markov process in a straightforward manner; however, in practice, such a modeling approach has several limitations. First, determining the exact transition probabilities is often not possible and one may need to resort to estimation techniques. Moreover, the number of states, which may not always be observable, may grow in a nonlinear fashion with the number of variables being modeled. Since the dimension of the state transition matrix increases quadratically with the number of states, the processing speed and memory requirements of an implemented computer simulation become negatively affected. Memory size and speed are always improving but these concerns

can limit the model's practical size (Chung, 2004). Further challenges specific to the problem at hand, are discussed in Chapter 5.

The basic approach taken was to create Markov models of different processes that drive the alerters, based on data collected in a cab environment study, and a set of associated assumptions. For example, the roadbed was modeled as a Markov stochastic process displaying certain statistical properties, which in turn drove the operator's required tasks as inputs to a Train Sentry class monitor. This technique was also used to model the DD-850 operation as well as the operator's eye visibility to the instrument. It is important to understand that Markov model simulations generate Markov series that are simply samples from the Markov process that generated them. If used individually, they do not fully represent the average statistical parameters of the underlying Markov process. Therefore, to capture the overall stochastic behavior of a given Markov process, and because many of the detector components being modeled are nonlinear, it is necessary to repeat the simulation using Monte Carlo methods, and use descriptive statistical methods to characterize the results. These ideas will be extended in further detail in the final model implementation in Chapter 6.

Stochastic models were used in this thesis to address the lack of extensive field or simulator data needed to drive engineering models for eye closure and activity. However, one may argue that driving the engineering models with authentic eyelid closure and activity data from a specific trip may lead to an analysis that is too restrictive (i.e., not generalizable) and not quite insightful, although extremely realistic. Moreover, the memory and processing duration of such a simulation may be exceedingly prohibitive. A more ideal and elegant method would require extracting the statistical properties of an extensive set of field data in order to produce tractable stochastic models to drive the simulation.

2.3.6 System Optimization

The final step in the approach requires overall system performance optimization. As in most engineering problems, the requirements and aims of a particular system may be at odds, and it may be challenging to meet all of them. In this application for instance, one ideally seeks a high CD rate and a low FA rate. However, as reviewed earlier, exactly how convex the system ROC curve can be (how far into the upper left corner of the ROC plot the curve passes) depends on the system architecture. Where the system should operate on the ROC curve depends on the probability of a drowsiness event, and the costs and benefits of CD and FA.

2.4 Concluding Remarks

The Alerting System design approach demonstrated in this thesis starts with the characterization of the AS environment; it continues with the evaluation of the subsystem's performance and concludes with a simulation tool that is expected to aid in obtaining an optimal or satisficing AS design. The model is specifically intended to answer the following questions, which otherwise may not be answered quantitatively:

- Is the presence of a single alertness monitor (either DD-850 class or Train Sentry class) "sufficient" to provide an acceptable level of performance?
- Do multiple alerters of different types, arranged in tandem, increase the robustness of the detection solution by circumventing the limitations of individual detectors?
- What are the required characteristics for the devices to display an improved overall system performance?

The results of the simulation and the answers to these questions must be interpreted with care. When (and if) implementing the overall optimal solution, one must not immediately expect optimal performance. Indeed, no computer model can fully replicate reality and totally replace the role of the physical prototype. However, effective use of such a simulation tool can produce a design that is satisficing, while considerably decreasing testing error rate.

Chapter 3

Quantitative Assessment of the Locomotive Cab Environment and Engineer Head Movement

This chapter[12] summarizes progress on the effort aimed at characterizing locomotive cab environmental and engineer behavioral factors. The focus of this study is to quantify the parameters that affect the performance of image-based eye closure (e.g. PERCLOS) detectors, as a guide for designers and evaluators of such systems.

[12] This chapter is an edited version of the conference paper:

Aboukhalil A., Oman C.M., Popkin S., Pollard J.K., Howarth H. *"Quantitative assessment of locomotive cab environment and engineer head movement for development of an alertness monitor"*, Proceedings of the 2005 International Conference on Fatigue Management in Transport Operations, Seattle WA, Sept 11-15, 2005.

3.1 Background and Objectives:

Chapter 1 outlined the main limitations associated with the existing locomotive cab alerters and the ones under development. Failing to consider the inherent differences between the locomotive engineer and automobile/truck drivers tasks and cab environment was a major concern. Therefore, the next logical step would be to quantify the important physical aspects of the cab and engineer behavior. This survey was not intended to be exhaustive. Its scope was limited to the factors pertinent to the evaluation of image-based fatigue detectors.

Based on the earlier Volpe evaluations by Oudonesom and Karnali, principal concerns about using PERCLOS information from IR image-based sensors as the primary or sole drowsiness detection criterion were:

a) Both daylight and nighttime monitoring capability is desirable. The automotive CoPilot works only in darkness. In railroad operations, extra-board freight crews working out of phase with their normal sleep-wake cycle can experience drowsiness problems during daylight hours. Unfortunately, no data was available to designers on the range of IR and visible light intensity in locomotive cabs.

b) Locomotive cab geometry is different and somewhat more variable than with automobiles and trucks. Cameras and illuminators can be mounted in front of the engineer, atop the display panels in modern locomotives (e.g. Bombardier Acela, GE/AC4400 and 6000, GM/EMD SD-70), but must be mounted elsewhere in older cabs utilizing a side mounted control stand (e.g. GM/EMD F-40). Due to the geometry of the cab, locomotives will generally require a larger camera working distance than automobiles and trucks.

c) Automobile and truck drivers must keep their eyes almost continuously on the road ahead. Locomotive engineers must look outside to check signal status, and watch for intruders and problems with the roadbed. However, engineers have detailed knowledge of the roadbed, landmarks and signal locations, Although no quantitative data was previously available, locomotive engineers were believed to devote somewhat more time to looking in the rearview mirror, and at objects inside the cab such as their radios, brake pressure displays, and console paperwork than the

corresponding truck driver. Locomotive engineers are also able to leave their seats (and hence the camera field of view) briefly e.g. to check displays or set controls elsewhere in the cab, or obtain documents from their luggage. Unfortunately, no quantitative data on engineer head movements and eye visibility to a camera were available to designers.

With these concerns in mind, between the summer of 2004 and the spring of 2005, Volpe and MIT conducted a survey of typical passenger and freight locomotive environments. Specifically, answers to the following questions were sought:

- How variable is the locomotive engineer's head position in three dimensions?
- How much of the time do the engineer's eyes – or for that matter, the engineer's head - remain in view of a panel-mounted camera?
- What is the range of IR and visible illumination in typical locomotive cabs?
- What range of cab temperature and vibration levels must designs accommodate?

3.2 Methods:

The five locomotive types studied included the F-40, AEM-7, Acela, GE AC-4400, and EMD SD-70. Eight locomotive engineers participated; one was a trainee. Experimenters avoided initiating conversation with subjects during the experiment. Procedures were approved by the MIT Committee on the Use of Humans as Experimental Subjects. Subjects were compensated for their participation.

As shown in Table 4, physical measurements were collected during 8 passenger and 5 freight trips made between Boston, MA and Providence, RI or Worcester, MA mostly during daytime hours, except for the freight trips that all started westbound, just before sunset and ended in darkness. Instrumentation was portable and battery powered, since AC power was not always available in the cab. The investigators typically had to only 5-10 minutes to set up the equipment prior to departure.

Table 4 – Summary of physical measurements classified per trip and engine type

	Passenger								**Freight**				
Engine	F-40	F-40	F-40	F-40	F-40	F-40 & AEM 7	Acela & AEM 7	Acela	GE AC4400	EMD SD-70	EMD SD-70	EMD SD-70	GE AC4400
3D Head Position		X	X	X	X					X	X	X	X
Video Camera						X	X	X					
IR Video Camera									X	X	X	X	X
Light Level (Lux)	X	X	X	X	X						X	X	X
Irradiance (mW/cm^2)											X	X	X
Sound Level (dBA)			X		X		X						
Temperature (°F)	X	X	X	X	X								
3-D Vibration (m/s^2)	X		X	X	X				X	X	X	X	X
GPS	X	X	X	X	X	X			X	X	X	X	

Primary measurements included:

a) Head position: Three-dimensional position of an adhesive retro-reflector mounted on the operator's forehead was tracked using an IR motion tracker (Dynasight, Origin Instruments, Inc., Grand Prairie, TX). Data was sampled at 36 Hz by a portable

computer (Model Latitude D505, Dell, Inc. Austin, TX). Absolute accuracy was 2 mm in the cab lateral and vertical directions, and 8 mm in the fore-aft direction.

b) Eye and head visibility: A visible and infrared sensitive video camera with a 40 deg. field of view (FOV) (Model DD850 CoPilot, Attention Technology, Inc., Pittsburgh, PA) was mounted on the engineer's instrument panel directly in front of the engineer, and 5-10 deg below eye level. IR illumination was used during night freight trips. Due to differences in cab configuration the distance from the camera to the subject's eyes varied between 24"-38". Images (black and white, 700x450 pixels) were sampled at .3 Hz by a frame grabber (VCE-Pro, Imperx, Inc., Boca Raton, FL) equipped portable computer.

Secondary measurements included:

a) Illumination at the engineer's head level, measured using either a visible range light meter (Model 401036 datalogger, Extech, Inc. Waltham, MA, sampled once/min.) or a radiometer/photometer (Model IL1400, International Light Inc., Newburyport, MA, sampled approximately every 20 min.).

b) Cab sound level, hanging from the cab ceiling near the engineer's head. (Model 407764 sound meter, Extech, Inc., Waltham, MA).

c) Cab temperature near the windshield, near the video camera mount location, sampled using a stowaway temperature data logger (Model XTI32-37+46, Onset Computer Corporation, Pocasset, MA).

d) Floor or console vibration, measured using an adhesive-mounted 3-axis linear accelerometer (Model CXL02LF3 Crossbow Technology, Inc., San Jose, CA, 0-50 Hz freq. response), oriented approximately along the major cab axes.

e) Locomotive speed and position were measured by a GPS system (GPS-35 USB, Garmin International, Inc, Olathe, Kansas) using an externally mounted antenna. Data was sampled at 1 Hz via portable computer.

3.3 Analysis:

3.3.1 Head Position:

Data segments each lasting 40-70 minutes were pooled and statistically analyzed (Systat v. 11, SAS Institute, Cary, NC) for each separate trip. Data distributions were approximately normal as shown in the Figure 12 example. The collected data excluded epochs when the engineer left his seat or fully turned away (>180^{o} to either side of straight ahead), so that the retro-reflector mounted on the engineer's forehead was no longer visible to the Dynasight tracker.

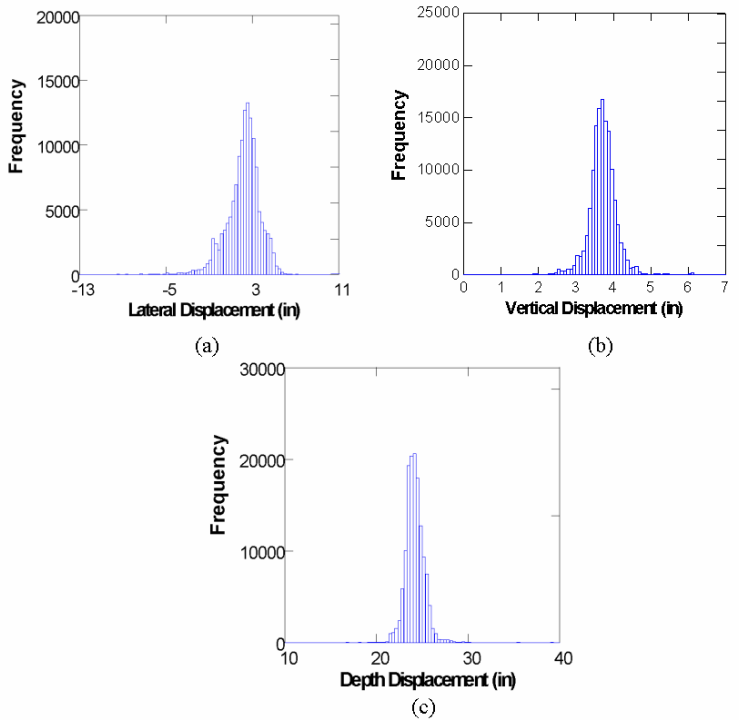

Figure 12 – Sample of head position distributions along the a) horizontal b) vertical and c) fore-aft axes from 1 trip (data duration 69 minutes)

The 95% limits of the head movement distribution on each axis were estimated by taking the mean of the +/- 2 sample standard deviation histogram width for each trip, as illustrated in Figure 13, and then averaging for each trip type. This defined the dimensions of an arbitrary 3-D volume or "head box" within which the operator's head was located. To the extent that the lateral, vertical and fore-aft head movements are statistically independent, the overall probability of the head being within the head box is 0.95^{3} or at least 86% of the time. To the extent that movement on the various axes was correlated, the percentage increases. Cross-axis correlation

coefficients varied between 0.07 – 0.7 on different runs. When the head box is defined this way, it should contain approximately 90% of all valid head position samples.

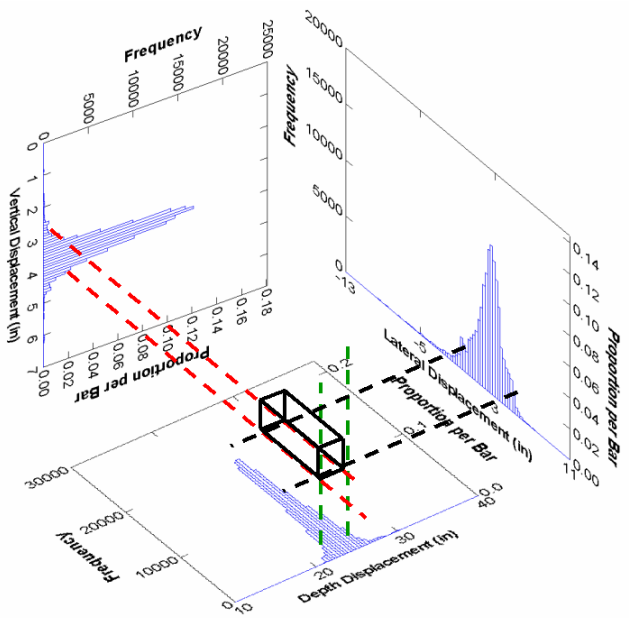

Figure 13 – Generating dimensions of three-dimensional head box. The +/-2 SD of the head position histograms on each axis were used to define the dimensions of a three-dimensional head box which contains the engineer's forehead ~ 90% of the time (see text for details.)

3.3.2 Eye and Head Visibility:

Data segment duration ranged between 16 minutes and 50 minutes, with all but one greater than 30 minutes. Camera images were classified by a human observer into one of four categories, based on whether part or all of the iris was visible, using the following criteria: 0: No eyes visible, 1: One eye visible, 2: Both eyes visible, and 3: Head not present in the image. Condition 0 – which we refer to as a look-away event - typically occurred when the operator turned to look at the rearview mirror or a side-mounted console, or to speak to another person in the cab. Condition 3 – which we refer to as a go-away event - resulted when the operator left his seat to attend to other tasks.

Secondary data measurements were analyzed using Systat's 'Descriptive Statistics' package.

3.4 Results:

3.4.1 Head Position:

The +/-2 standard deviation range (95% interval) of head position along each axis on each trip were computed and then averaged by type. The dimensions for each trip type are shown in Table 5:

Table 5 – Mean head box dimensions +/- SD.

	Lateral (in)	**Vertical (in)**	**Fore-aft (in)**
Passenger (4 trips)	9.3 +/- 1.0	6.7 +/- 6.4	9.7 +/- 3.4
Freight (4 trips)	9.5 +/- 2.9	2.4 +/- 0.6	7.1 +/-2.1
Overall (8 trips)	9.4 +/- 2.0	4.6 +/- 4.8	8.4 +/-3.0

3.4.2 Eye and Head Visibility:

Passenger and freight trips exhibited very similar eye and head visibility characteristics, and thus the percentage values in each state were averaged over all trips (Figure 14). Results are summarized by trip type in Table 6.

Table 6 – Percentage of eye visibility +/- SD.

	Look-Aways (%)	**One Eye (%)**	**Two Eyes (%)**	**Go-Aways (%)**
Passenger (4 trips)	5.8 +/- 4.8	9.2 +/- 3.4	82.2 +/- 7.5	2.8 +/- 2.4
Freight (5 trips)	6.0 +/- 3.9	5.8 +/- 3.11	86.8 +/- 6.5	1.4 +/- 2.1
Overall (9 trips)	5.9 +/- 4.0	7.3 +/- 3.5	84.8 +/- 6.9	2.0 +/- 2.2

Thus, a 40˚ FOV camera placed 24-38” in front of the engineer had both eyes in view approximately 85% of the time, one eye 7% of time and both eyes were out of view (a look-away event) 6% of the time, e.g. when looking briefly in a side mirror, or when engaging in conversation. The entire head was outside of the camera view (a go-away event) about 2% of the time.

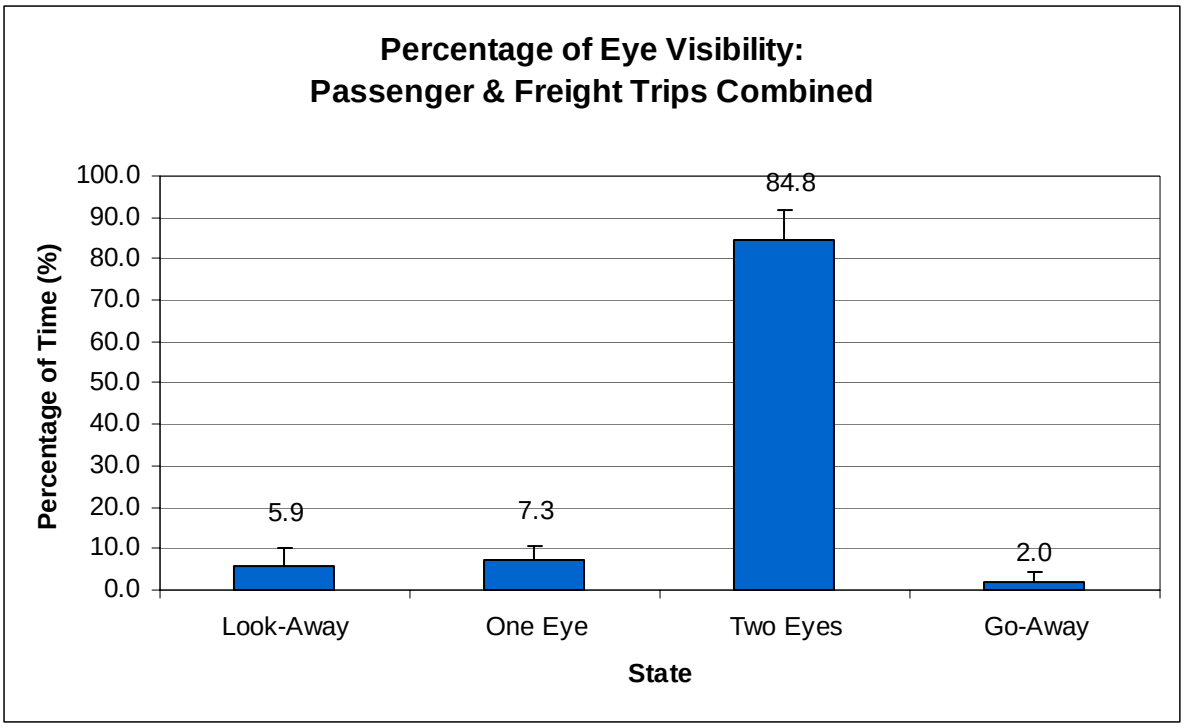

Figure 14 – Eye visibility percentages for freight and passenger trips combined. Error bars: +/-SD

Dwell time duration histograms for each of the four states (a-d) for the combined data set are shown in Figure 15. Distributions of freight and passenger dwell times were similar. Median look-away and one eye event durations (Fig 4a, b) were 1.6 and 1.5 sec, respectively. Median time that both eyes were continuously in view (Fig. 4c) was 11.7 seconds, though some intervals lasted up to 240 sec. Go-away events (d) - though relatively rare (2% of the time) - had a median duration of 7.0 sec. One 93 sec. go-away event was recorded.

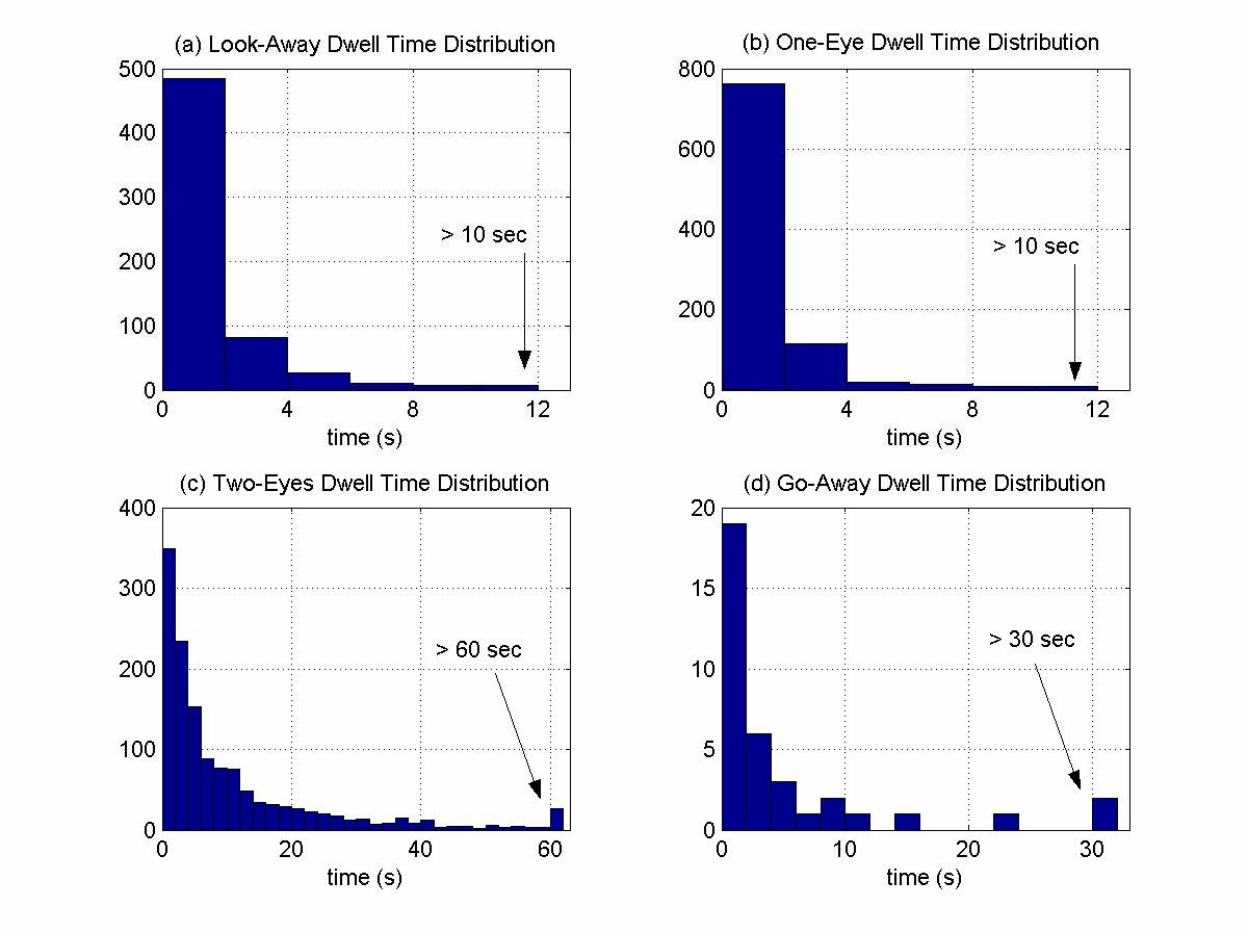

Figure 15 - Dwell time duration histograms for all the trips combined for each different state. The final bin of each histogram represents the frequency of all durations greater than the axis limit

3.4.3 Secondary Measurements:

Cab environmental characteristics are summarized below. Daylight illumination values are an average across all passenger trips, while sunset illumination values represent a range obtained from a sample collected on freight trips. Irradiance values correspond to representative values and ranges sampled during freight trips. “Direct” values mean the sun was directly in view of the radiometer. Acceleration range based on 4 passenger and 5 freight trips.

Table 7 – Summary of cab environmental characteristics based on samples obtained (Table 4.)

Measurement Type	Value
Illumination (visible, Lux)	342+/-202 Lux (in cab daylight) peak 1250 Lux (in cab daylight) 7-60 Lux (in cab near sunset)
Irradiance (IR+visible, mW/cm^2)	30 (in direct sunlight) ~ 2 (in direct setting sun) 0.6-0.9 (in cab in daylight) 0.1-0.3 (in cab near sunset) <0.01 (incoming train headlights)
Sound Levels (dBA)	F-40 79 dB +/-4 dB (SD), 95 dB peak Acela 66 dB AEM-7 82 dB
Temperature (F), max and min.	78.5 - 82.0
Acceleration range (g) +/- 2SD	+/-0.11 fore-aft +/-0.10 lateral +/- 0.10 vertical

IR and visible light intensity at the operator's eye level vary over several orders of magnitude; quantifying those levels is of particular significance in the design of image-based alerting systems. In the case of 'visible light'-based detectors, sufficient illumination intensity is required to identify the eye, or pupil. However, high levels of visible light constrict the pupil, and make it more difficult to detect with an IR image-based alerter. Incoming light and reflections off eyeglass lenses may conceal the eyes in different regions of the spectrum.

Sound levels matched passenger cabs certificates and were below the 90 dB/8 hr OSHA limit. However, the 66-82 dB typical levels warrant the hearing protection devices used by some – but not all – of our subjects.

The reported temperatures were collected on 5 passenger trips in the late morning, just before noon, mostly on sunny summer days. However, the weather conditions were not extreme, and the all of the surveyed cabs had air conditioning capability. Therefore, the reported range (78.5-

82.0F) does not reflect worst-case scenario operating conditions, where the weather conditions exceed the regulation limits of the cab's air conditioning system.

The 95% range of cab accelerations was approximately +/- 0.1g on each axis. Camera mounts must be robust to vibration and high-cycle fatigue, yet adjustable to accommodate differences in engineer and seat height and camera placement.

3.5 Conclusions:

Both freight and passenger results yielded head position box and eye/head visibility behaviors of similar magnitude despite the cab geometry differences. A console-mounted camera – a DD-850 prototype - with a 40 deg. FOV placed at least 13 inches away from the subject can see the entire (9.4" lateral, 4.6" vertical, 8.4" fore/aft) head box, provided the subject is initially properly centered. However, eye/head visibility data show that such a camera cannot image the eye about 8% of the time – either because the engineer is working elsewhere in cab (2%), or his head is turned away (6%). Some of these look-aways and go-aways are brief, and can be dealt with by temporal averaging of the eye visibility information. For example the CoPilot device incorporates a one-minute running averager. However, go-aways or long or successive look-aways, e.g. when conversing, can potentially defeat such temporal averaging logic. The value of adding a second camera to the system depends on the ability of either camera to reliably detect eye closure when the subject is looking off axis. IR and visible light levels vary over many orders of magnitude, which complicates the design of image-based detectors. Sound levels (79dB) were as anticipated; these levels are below the 90 dB/8 hr OSHA standard, but high enough to justify the hearing protection measures employed by some crews. Cab accelerations varied over approximately +/- 0.1g on each axis, which should be considered when designing camera mounts.

The results of this study raised questions regarding the performance of the Attention Technologies DD-850 in such an environment and motivated the evaluation in the next chapter.

Chapter 4

Performance Evaluation of the DD-850 Alertness Monitor

This section summarizes a performance evaluation report[13] of an Attention Technology's DD-850 prototype. A signal detection approach was taken to quantify the instrument's performance under relatively ideal laboratory conditions. The main goal was to estimate the performance of the DD-850 detector in ROC (CD vs FA) probability space, and determine whether performance remained in the same area as the operating conditions change. Some of the main factors affecting performance were evaluated. It is important to note that we concentrated on the running performance of the DD-850 "front end", not the downstream temporal averaging and threshold processes, since the parameters of those elements were considered part of the overall alerting system design that was to be optimized. Data obtained on look-aways in Chapter 3 were combined from front end performance data on an a DD-850 engineering model described in Chapter 4 to develop a MATLAB/SIMULINK simulation of the DD-850 described in Chapters 5 and 6.

[13] Aboukhalil A., and Oman C.M., Technical Note. Attention Technology DD-850 Eye Closure Monitor: Effects on Head Motion and Eye Color on Detection Performance. 07/06/2005.

4.1 Objective

To evaluate the performance of 04/27/2005 version of an Attention Technology DD-850 prototype under relatively ideal laboratory conditions from a signal detection perspective, and document an anticipated dependence on head movement and eye color.

4.2 Methods

Three subjects (A, B, C) were recruited for this experiment. One was Japanese with dark brown eyes and black hair (A, age 31), and the other two were Caucasian with light blue eyes (B, age 28 & C, age 61) and light brown hair. Subjects were seated in a totally dark room 30" away from the camera looking straight ahead at an angle of about 20 degrees above the camera as shown in the figure below. Subjects did not wear eyeglasses or contact lenses. Compared to conditions anticipated in actual railroad applications, these are relatively ideal conditions.

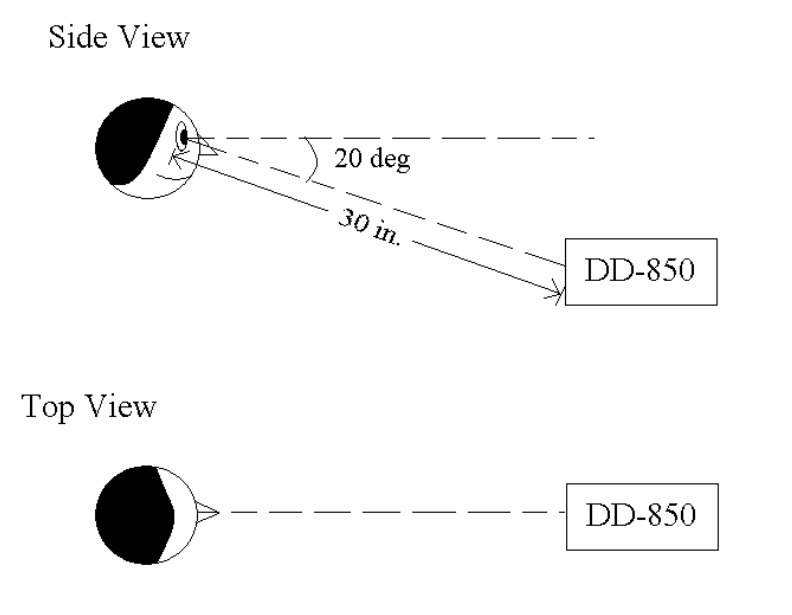

Figure 16 – Subject's position with respect to the DD-850 camera

The experiment was repeated under 2 conditions while seated:

A) Subjects remained motionless

B) Subjects moved ("bobbed") their head and body vertically ~ +/- 1cm at a frequency of ~1.5 Hz.

In each condition, subjects changed successively between the following 4 states:

i) Both eyes open

ii) Right eye open

iii) Both eyes closed

iv) Left eye open

Subjects A & B remained in every state for 15 seconds, resulting in about 45 samples collected per state (at 3Hz sampling rate). Subject C remained in every state for 30 seconds, resulting in about 90 samples per state. Subject C repeated the experiment 3 times to assess consistency of performance within subjects.

From a signal detection point of view, the DD-850 can be considered an eye *closure* detection system, (although strictly speaking the system cannot distinguish an 'eyes closed' condition from a 'head absent' condition). However, for present purposes, the head is assumed present so the eye closure detector perspective is appropriate. A microprocessor in the DD-850 analyzes every captured image frame and outputs the number of eyes detected as a 3Hz serial bit stream – where 0 = both eyes closed (or head absent), 1 = one eye open, and 2 = both eyes open. This 0-1-2 time series is then converted to a 0-1 binary 3Hz stream – where 0 = both eyes closed and 1 = one or two eyes open. Finally, the 0-1 binary stream is fed to an internal running averager with a square weighting function to compute the 1, 3, and 5-minute PERCLOS average as per Wierwille's definition (percentage of time the eyes are closed beyond a certain criterion) (Wierwille et al., 1994). Since only the 0-1-2 time series was available, it was recorded under each of the four conditions described above and performance statistics were calculated. For the evaluation, a signal detection perspective was employed. Therefore, we defined:

Correct Detection (CD): Reporting 0 eyes when both eyes are closed.

Missed Detection (MD): Reporting 1 or 2 eyes when both eyes are closed.

False Alert (FA): Reporting 0 eyes when one or both eyes are open.

Correct Rejection (CR): Reporting 1 or 2 eyes when one or both eyes are open.

Thus, one can characterize the detector's performance using 2 probabilities, namely P(CD) and P(FA) (with P(MD) = 1-P(CD) and P(CR) = 1-P(FA)).

4.3 Results

4.3.1 Effect of Head Motion

The DD-850's performance data for all subjects was pooled and tabulated below. Although there is variability in the data for each individual subject shown in Table 8, the mean provides an overall view of the effect of motion on the detector's performance.

Table 8 – DD-850 eye closure sensitivity and reliability data by subject

	Subject A		Subject B		Subject C Trial 1		Subject C Trial 2		Subject C Trial 3	
	No motion	Motion	No motion	Motion	No motion	Motion	No motion	Motion	No motion	Motion
CD	63.04%	91.30%	100.00%	92.47%	100.00%	72.83%	78.02%	73.91%	100.00%	85.87%
MD	36.96%	8.70%	0.00%	7.53%	0.00%	27.17%	21.98%	26.09%	0.00%	14.13%
FP	30.81%	6.42%	90.21%	36.52%	89.85%	5.44%	62.68%	1.09%	53.17%	1.09%
CR	69.19%	93.58%	9.79%	63.48%	10.15%	94.56%	37.32%	98.91%	46.83%	98.91%

Table 9 – Mean Detection Performance of the DD-850 for a) No Motion and b) Motion conditions. Each entry represents the average of subjects A, B, C (trials 1, 2 & 3).

	No Motion	Motion
Correct Detection	88.21%	83.28%
False Alarm	65.34%	10.11%

The receiver operating characteristic (ROC) is plotted in Figure 17. A curve (arbitrarily, logarithmic) was fit to the data of all subjects and trials was fit to emphasize the general character. As shown in the figure below, the false alarm probability of the device is very high (40-90%, average 65%) in the no motion cases. The device was signaling eye closure when in fact one or both eyes were open. When subjects made bobbing head motions, the false alarm probability was dramatically reduced down to the 10% level. The Correct Detection probability

(signaling eye closure when both eyes were closed) decreased only slightly (about 5%). Such a dramatic change in the ROC contingent on subject movement is undesirable from an alerting/detection system perspective. Presumably the sensitivity to motion originates in the infrared image processing software, but the manufacturer could not provide an explanation of its origin. Hopefully the performance of the instrument can be improved so that the no motion cases have SDT performance at least resembling the motion case, i.e. a much lower false alarm probability.

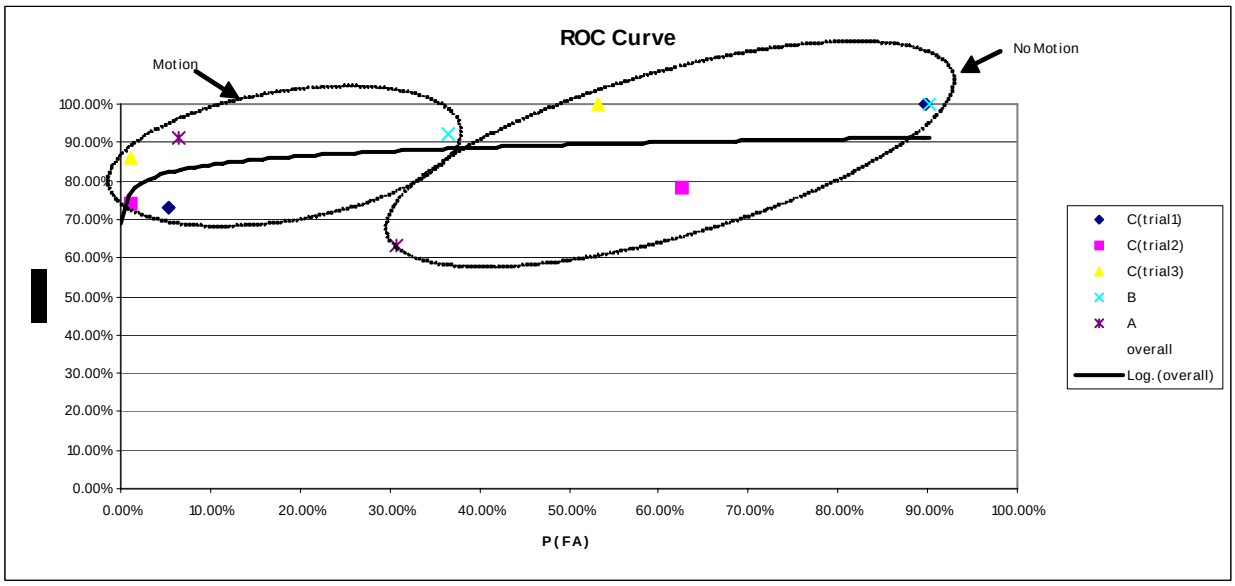

Figure 17 – Receiver Operating Characteristic for the DD-850

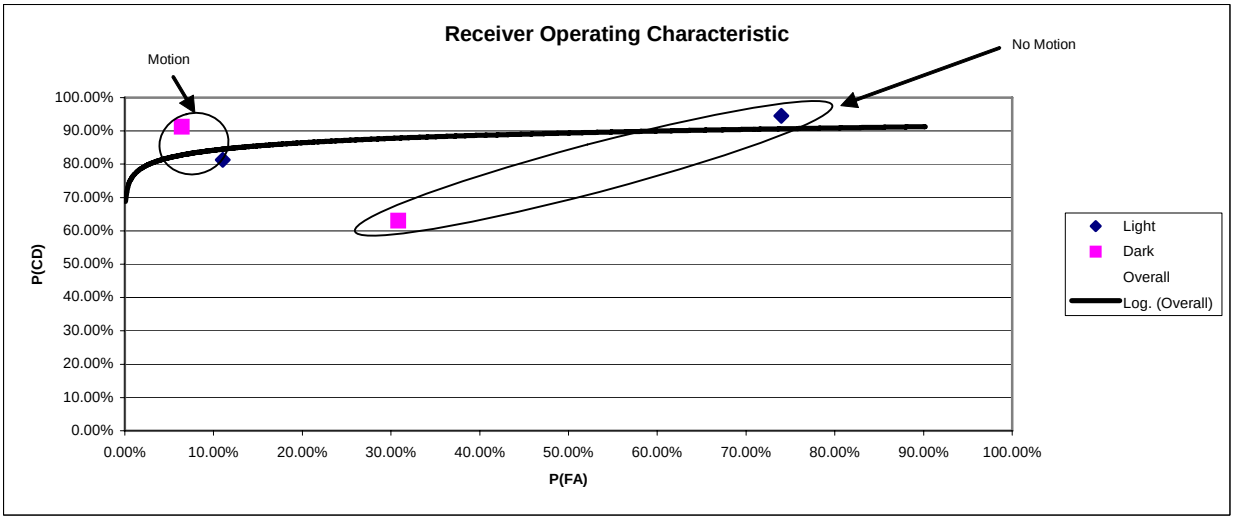

Figure 18 - Receiver Operating Characteristic by Eye Color

4.3.2 Effect of Eye Color

As shown in Table 10 and Figure 18, eye color also has a major impact on detection system performance, albeit not as large as motion. The FA is about 40% higher for the light eyed subjects than the dark eyed subject in the no motion condition. The sensitivity of performance to eye color is much smaller in the no motion condition, showing the importance of reducing the sensitivity of the DD-850 to head movement.

Table 10 – Mean Detection Performance of the DD-850 for a) Dark-eyed subject A and b) Light-eyed subjects B, & C (trials 1, 2 & 3).

	(a) Dark	
	No motion	Motion
Correct Detection	63.04%	91.3%
False Alarm	30.81%	6.42%

	(b) Light	
	No motion	Motion
Correct Detection	94.51%	81.27%
False Alarm	73.98%	11.03%

4.4 Consistency of Detection Performance

Subject C's data show that detector performance probabilities vary from one trial to the next. The cause of this variability could not be directly determined, but is unlikely to be due simply to undersampling, since 90 samples were taken for each trial. More likely it reflects an inherent sensitivity of detector performance to small variations in subject gaze angle, blink rate, etc., that undoubtedly occurred within and between trials. However, this variability is dramatically improved when the subject bobs his head: as shown in Table 11, for subject C, bobbing decreases the standard deviation from 12.69% (CD & MD) and 19.04% (FP & CR) to 7.24% and 2.51% respectively. Thus eliminating the sensitivity of the DD-850 to subject head motion will also likely improve the consistency of front end detector performance from a SDT perspective.

Table 11 – Reports the effect of motion on the variability of the repeated trials data for Subject C. Entries = 1 SD.

	No motion	Motion
CD & MD	12.69%	7.24%
FP & CR	19.04%	2.51%

4.5 Conclusion

Even under ideal laboratory conditions, the current version of the DD-850 front end image analysis software exhibits large sensitivity in detector ROC performance and consistency due to sensitivity to head movement and eye color. Subject's eye color can't be standardized, but eliminating or substantially reducing the dependence of performance on subject head movement is a highly desirable goal. Based on this limited data set, if the dependency on head movement can be reduced, detector P(CD) = 70-90% and P(FA) = 1-5% can hopefully be achieved, regardless of eye color. A low P(FA) and high P(CD) are particularly important if the detector is to be used in real world railroad applications.

Dependence of performance on visible and IR ambient light levels was not evaluated in these tests, but ultimately should be. The manufacturer specifies that the instrument is for nighttime use only. However, numerous accidents occur in the early morning, when the circadian phase – and hence alertness – is at its nadir.

Chapter 5

Stochastic Simulation Tool Development

The rationale for the development of a stochastic simulation Alerting System (AS) design tool was explained in Chapter 2. This chapter details the structure of the model as implemented in MATLAB v.7.0.4 and SIMULINK 6.2.1 (using the MATLAB Statistics and the Signal Processing Toolboxes). The overall system architecture is presented and the modeling approaches and strategies behind each module are explained. Finally, to control the flow of the simulation, a MATLAB Monte Carlo simulation script (in which the simulation model is embedded) is described.

5.1 Model Assumptions & Simplifications

Modeling a train trip at a realistic complexity level, while accounting for all possible variables is a nearly impossible task. Therefore, a number of assumptions and simplifications were made to make the AS simulation tool manageable, and still yield useful results.

a) Jewett and Kronauer's (1999) Cognitive Throughput Model (CTM) output was used to predict when the engineer will be susceptible to falling asleep. The original CTM itself predicts Cognitive Throughput but does not specifically predict sleep susceptibility. Also, it does not yet describe the variability in cognitive throughput. Therefore, to model the onset of sleep for purposes of design, an ad-hoc stochastic sleep threshold was added to the CTM model. If the operator's alertness level drops below that threshold value, the operator was assumed to be asleep or experiencing a drowsiness event. This modification/extrapolation of the CTM has some face validity, but was not formally validated as part of this thesis.
b) A drowsiness event was defined as an extended period of time where the operator does not respond to the controls due to an alertness level below the defined sleep threshold.

Therefore, without available data regarding the incidence of drowsiness or sleepiness events, it is assumed that all drowsiness events ultimately lead to sleep events. See section 5.3.2 for a more detailed and complete definition.

c) Any possible effects of physical activity and monotony on alertness were not modeled, since they are not yet quantitatively well understood. Engineers anecdotally report that driving up a long slope or on any dull and unengaging terrain – not requiring many control adjustments or not providing any sensory stimulation – often leads to decrements in alertness level. Conversely, physical activity may reduce drowsiness level, at least temporarily. Once such effects are more quantitatively defined, they could be incorporated into the model.

d) The operator was assumed to be competent and well-trained, i.e. not a student engineer.

e) The operator was assumed not to make any faulty inputs when controlling the train. If the operator falls asleep, no tasks were assumed accomplished; when the operator is awake, drowsiness level fluctuations were assumed not to induce operating errors. (A second generation model could incorporate a non-binary relationship between drowsiness and task performance.)

f) The operator tasks were assumed to be primarily driven by the roadbed's topographical characteristics. The roadbed was represented as a series of serial segments, each of which defined the roadbed local topographical state and required a control response from the operator. The operator was assumed to make all necessary train control adjustments in one temporal cluster of activities immediately after entering a particular roadbed segment (station, crossing, curve, grade, etc). The operator was assumed to have good knowledge of the roadbed and consist, and properly anticipate the response of the train, so not more than one set of adjustments was needed in every roadbed segment.

g) The train dynamics *(e.g. acceleration and deceleration, car interactions and coupler slack)* were not specifically modeled. Although the engineer was assumed to initiate speed changes appropriate for each segment, train acceleration and deceleration behavior per se were not considered to be major determinants of the alerters' performance. All speed changes depend on the roadbed's topographical type, with each state prescribing a particular nominal speed. A more detailed model for manual control could be incorporated in the future.

h) The simulation was designed to run by 1-second steps, and all events occurred at multiples of 1 second intervals.

i) Operator delays associated with perception, decision, reaction, and execution times and their dependence on alertness level were not modeled. Therefore, if awake, the model's operator acknowledged Train Sentry's visual alert 1 second after it was issued; if asleep, the operator will not be responsive and will remain asleep.

Other assumptions specifically relating to terrain/track roadbed characteristics, tasks, train speeds, alerter performance, and operator eye visibility are explained in the respective module descriptions that follow. In general, a combination of forward and inverse modeling approaches was taken. Starting from first principles, blocks were defined and the simulation was built to describe the behavior of a system. At other times, when data was available, statistical methods were used to work backwards and generate the model's parameters.

5.2 Overall Architecture

A top-level graphical representation of the AS design tool structure, logic, inputs and outputs is shown in Figure 19. The general structure of the model can be best explained by describing the flow of information from inputs to outputs. Details of individual blocks are described in subsequent sections.

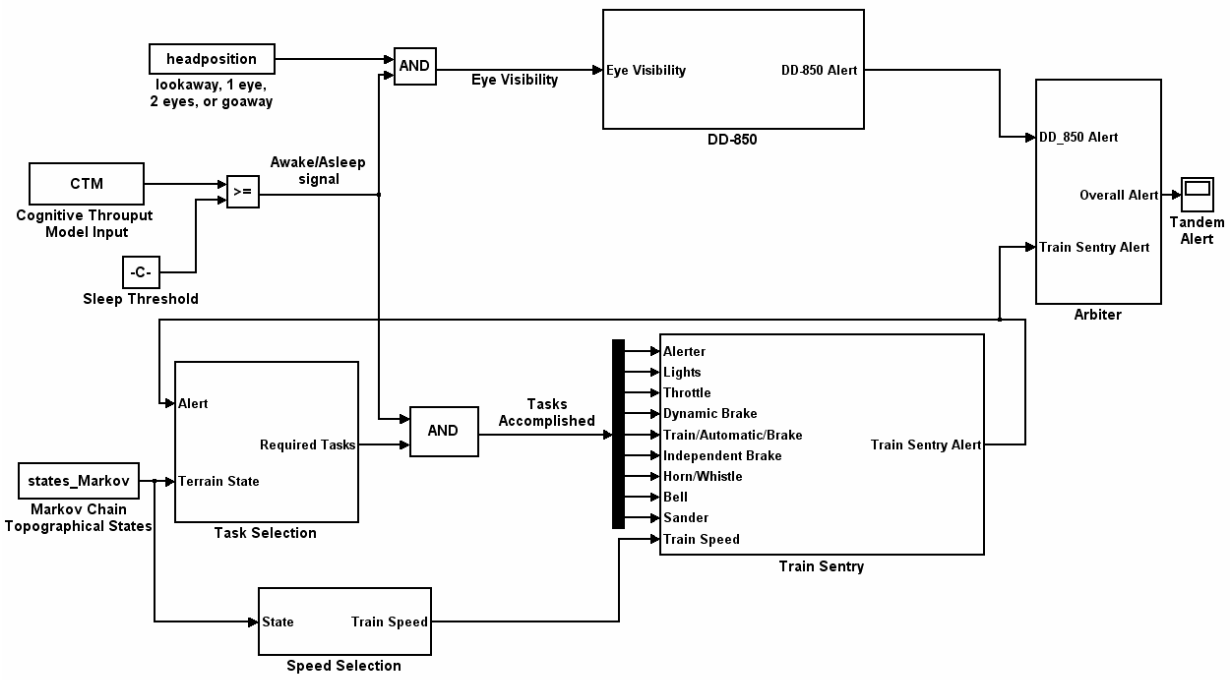

Figure 19– Top-Level Model Schematic

The 'states_Markov' input signal describes the roadbed's terrain characteristics and is used to specify a) the train's speed selection as well as b) the required control task subset the operator must accomplish through the Task Selection block. Whether those tasks are carried out or not depends whether the operator is awake or asleep. Jewett and Kronauer's (1999) CTM was used to model the operator's level of alertness as a function of time, given the past history of sleep and wakeup times and the ambient light intensity. A threshold, 'Sleep Threshold', was set below which the operator was assumed to be asleep. The binary 'Awake/Asleep signal' determined whether activities were executed and whether the operator's eyes were open. The 'Tasks Accomplished' subset was sensed by the Train Sentry class monitor portion of the model and

alerts were issued if too few operator activities were detected. The simulation stops if the trip duration is expired or until there is automatic train brake application after an operator falls asleep.

When the subject fell asleep, the eyes were assumed to close. At other times, although the eyes may be open, they may not be in view of the DD-850 monitor. Therefore, in the model the 'headposition' input determined whether the operator was "looking away", has "gone away", or has 1 or 2 eyes in view. The resulting 'Eye Visibility' signal was detected by the DD-850 portion of the model and the alerts were issued correspondingly.

The alerts produced by the Train Sentry and the DD-850 were sent to an *Arbiter* module that combined both inputs to issue an alert of the tandem system. Finally, there were additional blocks, only shown (shaded) in Figure 20 for simplicity, that were used to compute the SDT statistics of each alerter and of their tandem configuration. Figure 20 is a more detailed schematic of the model discussed in the remaining of the chapter. One major difference between the two figures is the presence of two arbiter blocks, one for each of the instruments' two-level alerts (first and second-level).

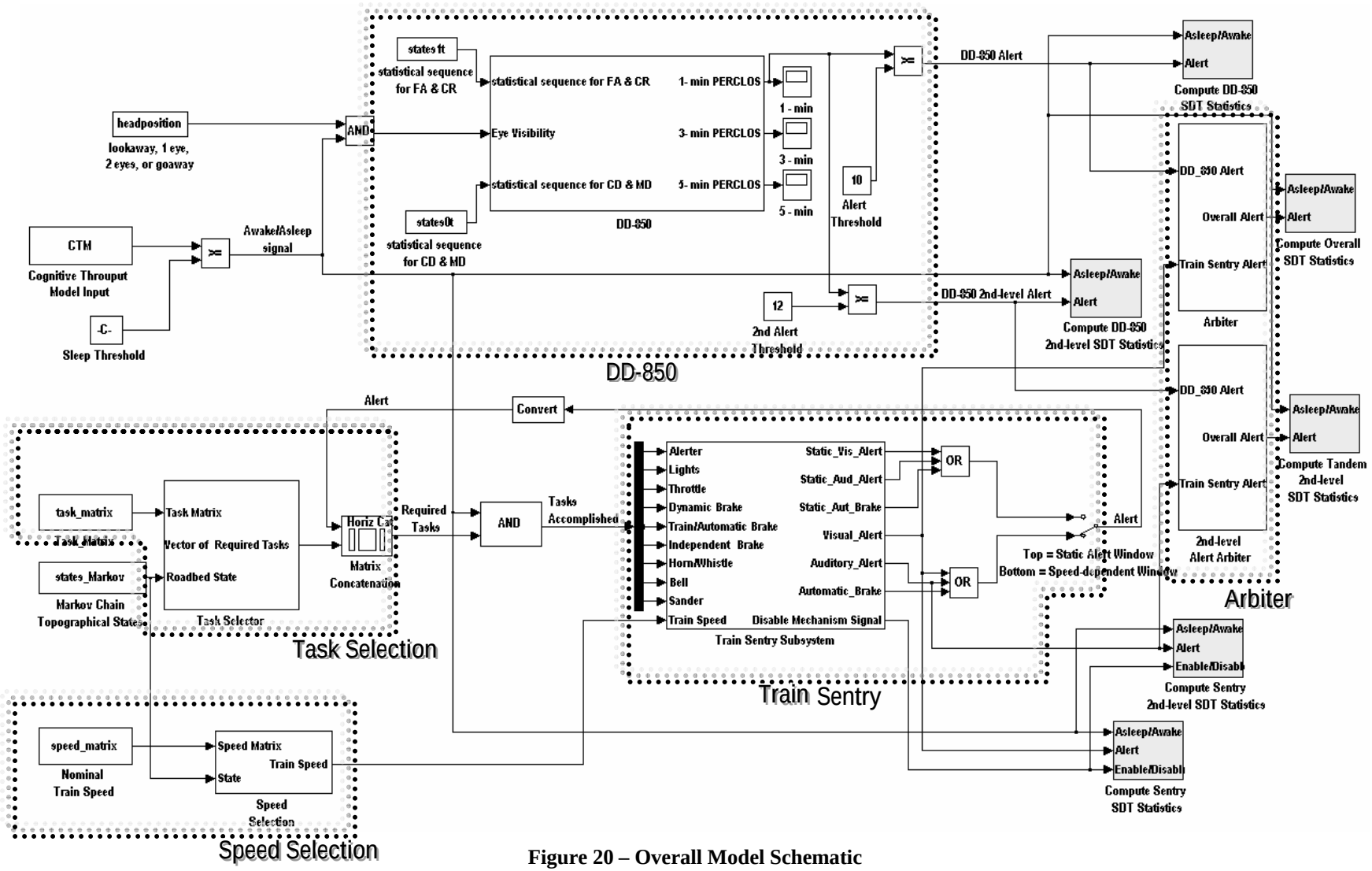

Figure 20 – Overall Model Schematic

5.3 Model Implementation

The specifics of the model's implementation in MATLAB and SIMULINK will be discussed.

5.3.1 Cognitive Throughput Model: Alertness Level

An introduction to the physiology of circadian rhythms is useful to understand this portion of the model. In humans the sleep-wake cycle, as well as many other physiological and psychological mechanisms, follow a rhythmic behavior with a near 24-hour period. Franz Halberg termed those cycles 'circadian rhythms' from the Latin words *circa* (about) and *dies* (day) (Brown et al., 2000). The control center of this biological timekeeping mechanism is formed of a group of 10,000 hypothalamic cells (Klein, Moore & Reppert, 1991) called the suprachiasmatic nuclei (SCN), symmetrically located in both sides of the brain. This master oscillator is believed to control other oscillators in peripheral tissues and organs, thereby regulating their local activities (Schibler et al., 2003). The operation of the master pacemaker is affected by external stimuli; the SCN can receive inputs through a direct route from special retinal cells. In particular, light information (intensity and wavelength) has been found to be the most effective means of delaying the circadian phase, as measured indirectly by levels of core body temperature, melatonin and cortisol[14]. Lately, it has been shown that shorter wavelengths (blue/green) are the most effective light agents in advancing the phase of the circadian rhythm (Wright, Lack & Kennaway, 2004). In sum, characterizing the behavior of the pacemaker is of high scientific interest and of high operational applicability to performance assessment and fatigue countermeasure development in transportation and shift work industries. Although mathematical models have been developed to predict circadian effects, many questions remain unanswered as to the underlying mechanisms of the human biological clock.

Mathematical models have been developed as tools to describe circadian-driven responses and behaviors, to probe the system and to propose further investigations. Kronauer et al. (1982) introduced a dual, coupled Van der Pol oscillator model to describe the variation in core body temperature and levels of cortisol in the blood. However, this approach could not fully account

[14] Assessing the activity of such physiological processes – that are driven by the oscillator and whose behavior is coupled to it – is one paradigm to study the oscillator's operation.

for the experimental sleep-wake behavior data. The two-process model (Borbély, 1982), which subsequently evolved into the three-process model (Folkard & Åkerstedt, 1987), appears to capture sleep-wake performance more faithfully. The model implemented in this simulation is Jewett and Kronauer's (1999) Cognitive Throughput Model (CTM) as shown in Figure 21, and is an example of a three process model. Cognitive Throughput (CT) represents the number of two-digit arithmetic additions correctly performed in 4 minutes, scaled between 0 and 1 measured under controlled laboratory conditions involving sleep deprivation, 28-h forced desynchrony, and sleep inertia investigation. For present purposes CT was preferred to Subjective Alertness measures (e.g Jewett & Kronauer, 1999) since it predicted experimental performance in an objective task, rather than subjective alertness. Humans are poor judges of their experimental performance abilities, especially when drowsy.

In summary, as shown in Figure 21, CTM is a three-process model, with H representing a homeostatic drive process, C is the circadian process, originating in two Van der Pol oscillators (modeling the circadian pacemaker), and W is a sleep inertia process. CT is assumed related to the sum of these three components. The homeostatic process models the drive for sleep that increases exponentially with time awake and decreases with time asleep. If H rises above an upper theoretical threshold, sleep occurs (unless of course the subject decides by his/her own will to remain awake), and if it declines below a lower threshold, awakening occurs.

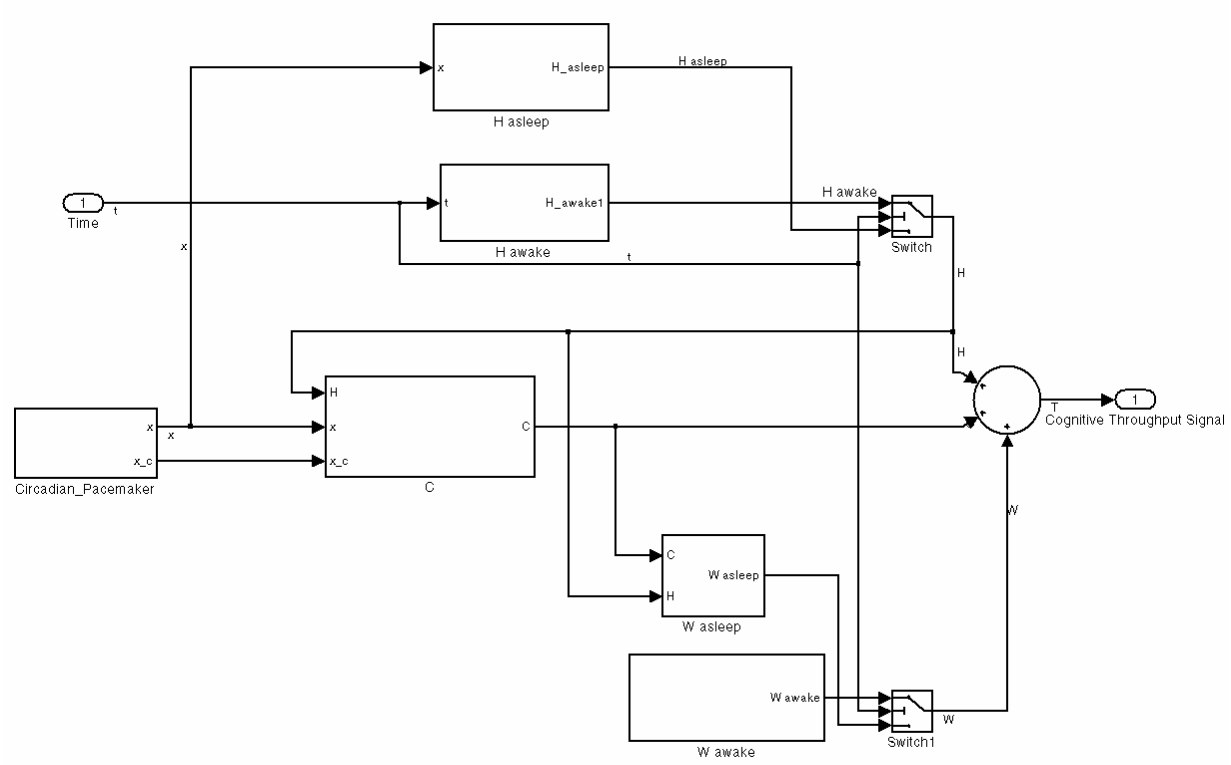

Figure 21 – CTM Block: Implementation of Jewett and Kronauer's (1999) Cognitive Throughput Model

The circadian component represents the drive for sleep or wakefulness as a function of the biological clock. For example it models the enhanced alertness in the morning despite a sleepless night, or the diminished alertness in the afternoon despite a well-rested night. The amplitude of the circadian component depends in a nonlinear fashion on H, and its phase depends on the phase of the 'Circadian Pacemaker', which is affected in turn by ambient light. Sleep inertia represents the decrease in alertness and grogginess that humans experience when they wake up, which gradually fades away.

The CTM was developed to predict how CT depends on the past history of sleep/wake states and ambient illumination. The threshold of the switches shown in Figure 21 is the waking time; it permits the user to select which processes are active at a given moment, namely, asleep (W_asleep and H_asleep) or awake (W_awake and H_awake). Therefore, the user can readily change the wakeup times and lighting intensity exposure and study how CT varies during subsequent periods of wakefulness, such as during a trip in a locomotive cab. Figure 22 shows

the CT model's output for a subject who slept for 8 hours, and then remained awake in 10 lux (dim light) for 50 hours.

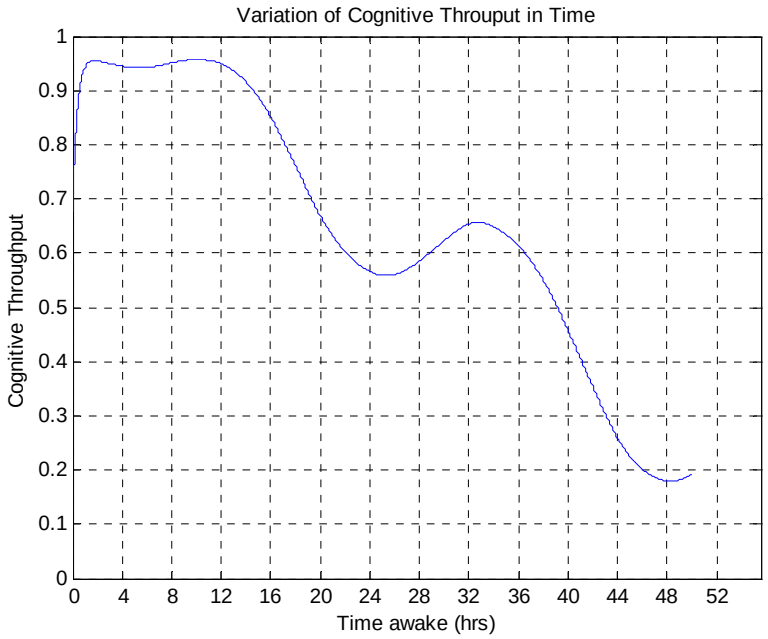

Figure 22 – Cognitive Throughput (CT) model output as a function of time. Represents the CT of a subject going to sleep at midnight, waking up at 8am, remaining awake for 50 hours.

The CTM model structure and parameters were chosen to fit data collected from one hundred fifty 30 to 50-hour sleep deprivation studies. However, the validity of the CTM outside of controlled laboratory studies is certainly still open to discussion. The main problem in applying the model to predicting sleep/wake states in railroad operations is that it predicts CT, not sleep/wake state. Also there are many real-world factors not present in the lab that affect sleep, including food, caffeine, uncontrolled light exposure, physical activity, and perhaps other unknown or misunderstood elements, accounting for such factors in a model may be extremely challenging. Also, the model only projects the mean CT as a function of time awake. It does not predict the confidence interval in this metric. However, given the limited availability of open models in the literature, CTM was chosen for this simulation as a method to project the sleep/wake state of engineers leaving on a trip, given their circadian cycle and past sleep pattern.

5.3.2 Sleep Threshold

Since experimental data was lacking to describe the relationship between Cognitive Throughput and the occurrence of a sleep event, an ad-hoc stochastic sleep threshold was used to create an 'Awake/Asleep signal' from the CT signal. If the CT level was above the 'Sleep Threshold', the value of the 'Awake/Asleep signal' was set to 1 and the operator was assumed to have eyes open and to accomplish all required tasks. Conversely, when the 'Awake/Asleep signal' was 0, the eyes were assumed closed ('Eye Visibility' signal was 0) and the operator accomplished no tasks ('Tasks Accomplished' vector signal was 0). Without any data with regards to the temporary effects and alertness dynamics of the alarm waking up the operator, the operator was assumed to remain asleep until the end of the particular trip simulation, despite an auditory alert was issued.

When the simulation is run, the user specifies the operator's sleep history and circadian phase, which then defines the range of the value of CT throughout that type of trip (e.g. daytime passenger; extraboard freight, etc.) An example is shown in Figure 23 below. However, the CTM model does not predict the probability of falling asleep on a given trip, or exactly when it will happen. To introduce this type of variability into the model for Monte-Carlo simulation purposes, the Sleep Threshold was assumed to vary for each trip. The model assumes that once out of *n* trips, the operator falls asleep. The probability of falling asleep for a given trip type was inferred from personal communication with the engineers and from the limited literature (Raslear & Coplen, 2004; Pollard, 1996). By using the maximum and minimum CT values on a given trip type predicted by the CTM model, one may compute the trip CT range, max(CT) – min(CT), and set a lower bound of sleep threshold *k,* as illustrated in Figure 23, as follows

$$\frac{\max(CT)-\min(CT)}{\max(CT)-k}=\frac{1}{n}=\textit{P(1 drowsiness event per n trips)} \quad \ldots(5.1a)$$

or

$$k=\max(CT)-n\left(\max(CT)-\min(CT)\right) \quad \ldots(5.1b)$$

The idea is to judiciously set *k* so that a uniform random number generator built into the model will produce a sleep threshold value for each Monte-Carlo simulated trip between *k* and

max(CT). The lower bound *k* value is chosen, given the trip type, so that on the average 1 sleep event per *n* trips occurs.

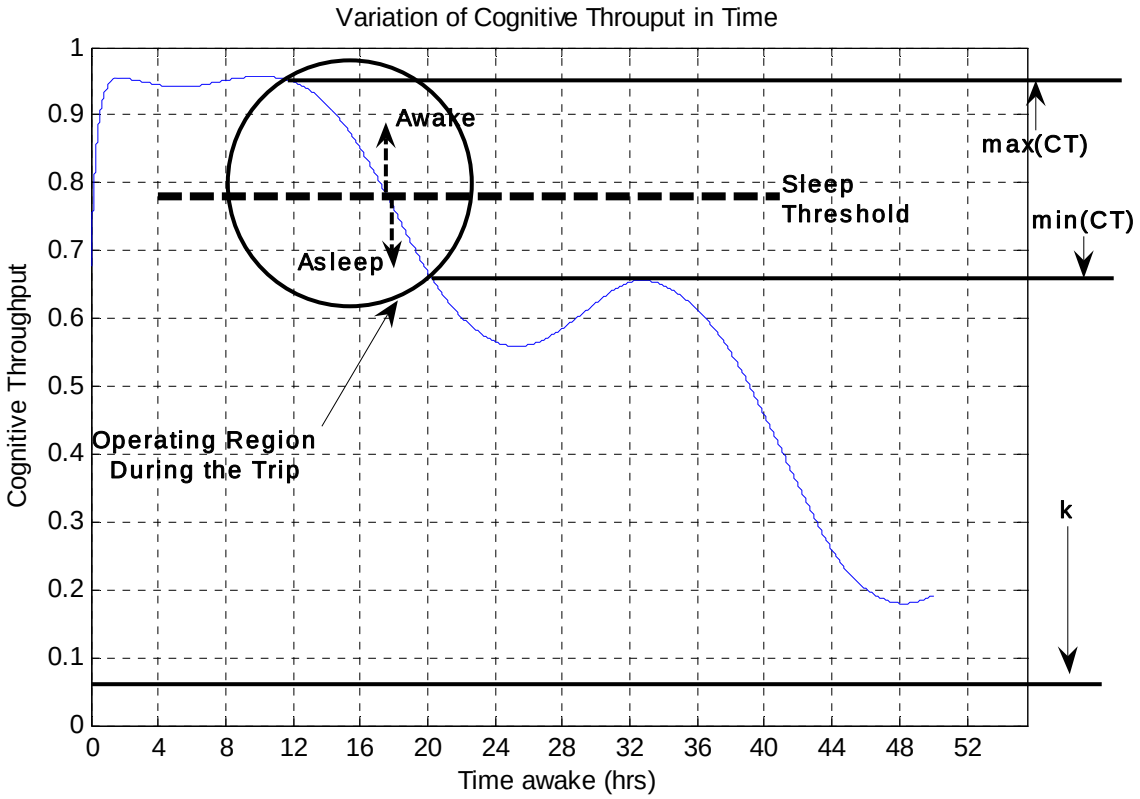

Figure 23 – Threshold setting procedure. The value *k* is set so that a uniform random number generator would produce a number between *k* and max(CT) that occurs during the trip (i.e. between max(CT) and min(CT) based on the required frequency of 1 drowsiness event per *n* trips.

Chapter 6 will detail the rationale for the drowsiness event probability depending on the type of rail operation.

Note that since the goal was to evaluate simulated detector performance, the onset of a sleep event during a trip was assumed not to be affected by the occurrence of a first or second level alert by the Train Sentry, PERCLOS monitor, or arbiter, i.e. the possibility that a first level alert would "wake up" the operator, and influence his/her response to a subsequent second level alert was considered an unnecessary complication.

5.3.3 Task Selection

The main objective of the Task Selection elements is to use a Roadbed Model and a Task Matrix to provide a plausible representation of the frequency of control inputs to the Train Sentry portion of the model.

As stated earlier, a number of permanent and temporary factors influence the engineer's train control activities. For this simulation, the main assumption was that it is the roadbed's topographical state (in terms of grade and curvature) that largely drives the engineer's activities. Although weather conditions, unexpected obstacles, signal states, and switch states considerably influence the control inputs, they are not here considered for the sake of simplicity. Therefore, to construct the simulated roadbed, the use of the following 9 states along with a series of associated rules was adopted:

Roadbed topographical states:

- State 1: Flat and straight
- State 2: Flat and curvy[15]
- State 3: Upward slope and straight
- State 4: Upward slope and curvy
- State 5: Downward slope and straight
- State 6: Downward slope and curvy
- State 7: Station
- State 8: Crossing
- State 9: Siding

Roadbed building rules (constraints on numerical probability values in the Roadbed Markov state transition matrix)

- The train enters and leaves stations and crossings (states 7 and 8) from a flat and straight terrain (state 1).
- The train enters and leaves sidings (state 9) from a flat curve (state 2).
- A grade cannot change its slope from upwards (states 3 and 4) to downwards (states 5 and 6) or vice-versa without going through flat terrain (states 1 or 2). However, there are no transitions between a flat curve (state 2) and an upward or downward curvy terrain (states 4 and 6) and vice-versa; the curve is considered part of the latter segment.

[15] A curve may equivalently be to the right or left; the direction of the curve does not influence the fact that a control adjustment is made.

It was assumed that each roadbed state has a fixed subset of required operator tasks associated with it, as defined by a Task Matrix. The task inputs determined by this matrix are the Train Sentry's inputs. The model represented eight tasks: lights, throttle, dynamic, train and independent brakes, horn/whistle, bell and sander. As defined by the Task Matrix, in each of the nine roadbed states the engineer had to perform tasks in the following combinations:

- State 1: Throttle
- State 2: Throttle and train brake
- State 3: Throttle
- State 4: Throttle and train brake
- State 5: Throttle, dynamic brake, train brake, and independent brake
- State 6: Throttle, dynamic brake, train brake, and independent brake
- State 7: Lights, throttle, train brake, and bells
- State 8: Lights, throttle, train brake, horn/whistle, and bells
- State 9: Throttle, train brake, and independent brake

By modifying the "task_matrix, the user can easily implement additional or fewer tasks on each type of roadbed. Again, the goal of the task matrix component is to provide a binary input vector to the Train Sentry that is monitoring all types of activity in train controls. Therefore, the nature of the control change is not crucial to the performance evaluation of the Train Sentry; the following cases are all equivalent from such perspective: the dynamic brake is used rather than the train brake, the throttle is adjusted from 3 to 5 rather from 3 to 4, the horn is sounded once rather than twice. The main point is that the engineer makes an activity that is detected by the Train Sentry.

A 9x8 matrix 'task_matrix' was defined with each row containing a binary vector with 1s for the required command adjustments and 0s for the rest. Figure 24 shows how the value of the 9-state 'states_Markov' stochastic roadbed signal (described below) is used to select the corresponding vector or row of the 'task_matrix'. Since the simulation is run on a 1-second clock, and one does not intend the operator to make adjustments every second; only one set of inputs per roadbed state was assumed. As a result, the 'Vector of Tasks Required' is a null vector (i.e. no tasks required) if the current and past states are the same.

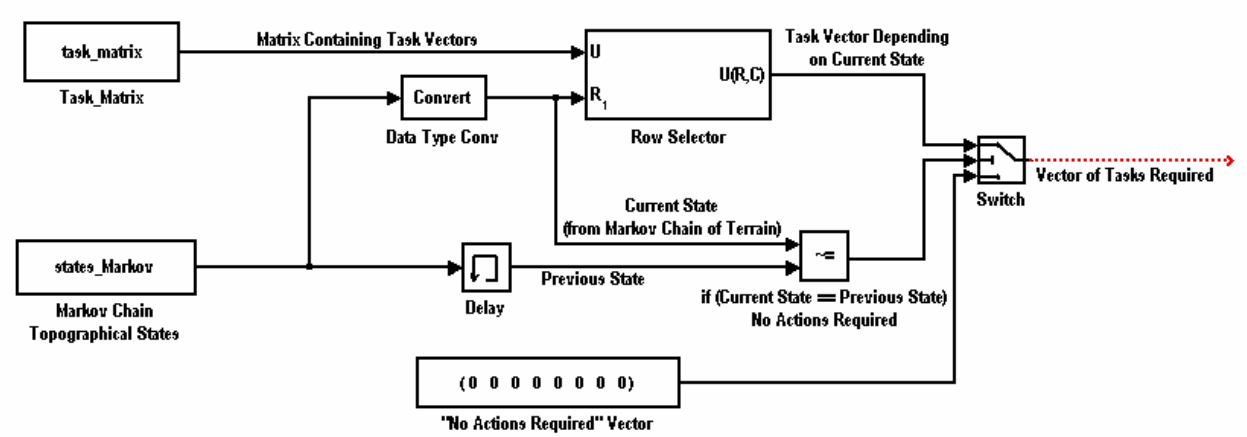

Figure 24 – Task Selection Block: Schematic describing how the vector of required tasks is generated

One task that the engineer must perform that does not depend directly on the roadbed state or task matrix is the acknowledgement of the Train Sentry timer. This task is triggered by the Sentry's internal timer mechanism described in section 5.3.8. To model this, the alerter acknowledgement task is added to the vector of required tasks using the 'Matrix Concatenation' block as shown in Figure 20.

For testing purposes, the simulation was implemented so that the performance of all three configurations (DD-850 standalone, Train Sentry standalone and tandem) may be assessed simultaneously. Therefore, the Train Sentry alerter light is not disabled and replaced by the arbiter or the standalone DD-850. Although the Train Sentry alerter light remains on, the computed SDT statistics of the three configurations are not affected (ex. FA rates or probabilities) due to the way they are implemented, as described in section 5.3.10.

5.3.4 Roadbed

With no available roadbed terrain data, Markov modeling techniques were used to statistically describe the variation in the roadbed. The roadbed can be considered as a sample realization path of a stochastic 9-state Markov process, 'states_Markov', with the train moving between those states over time. The description of the terrain is governed by the structure of the transition matrix. The transition matrix diagonal entries, p_{ii}, provide the probability of remaining in the

same state from one moment in time to another. Since the simulation is run at 1 second intervals, the following equation relates the diagonal probabilities to the mean duration in each state $\overline{d}_i$

$$\overline{d_i} = \frac{1}{(1 - p_{ii})} \ \ (sec), \ \ i = 1,2,...,9 \quad \ldots(5.2)$$

One may rewrite the latter equation to select the transition probabilities based on the desired mean state durations

$$p_{ii} = 1 - \frac{1}{\overline{d}}, \quad i = 1,2,...,9 \quad \ldots(5.3)$$

Thus, for this simulation the following mean state durations were arbitrarily chosen:

- State 1: Flat and straight: ~33 seconds
- State 2: Flat and curvy: 4 seconds
- State 3: Upward slope and straight: 20 seconds
- State 4: Upward slope and curvy: 7 seconds
- State 5: Downward slope and straight: 20 seconds
- State 6: Downward slope and curvy: 7 seconds
- State 7: Station:
    - Passenger: 50 seconds
    - Freight: 5 minutes
- State 8: Crossing: 10 seconds
- State 9: Siding: 7 minutes (420 seconds)

Next, given the starting state '*i'*, one can select the frequency of occurrence of the states '*j'* by simply tuning the matrix entries p_{ij} as desired. For example, if $p_{12} = 0.02$ and $p_{13} = 0.002$, one is 10 times more likely to encounter a curve than an upward slope when operating on a flat and straight terrain. Therefore, the user may construct any stochastic terrain desired by selecting the mean time for each state and the relative frequency of state occurrence.

A roadbed state transition matrix was qualitatively tuned to yield sample terrain functions that drove task input incidence at a frequency that seemed qualitatively reasonable based on the author's first hand experience riding in freight cabs. The following transition matrix was built and used in the simulation:

$$T = \begin{bmatrix} 0.9700 & 0.0200 & 0.0020 & 0.0020 & 0.0020 & 0.0020 & 0.0010 & 0.0010 & 0.0000 \\ 0.2350 & 0.7500 & 0.0100 & 0.0000 & 0.0049 & 0.0000 & 0.0000 & 0.0000 & 0.0001 \\ 0.0200 & 0.0100 & 0.9500 & 0.0200 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0100 & 0.0000 & 0.1329 & 0.8571 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0200 & 0.0100 & 0.0000 & 0.0000 & 0.9500 & 0.0200 & 0.0000 & 0.0000 & 0.0000 \\ 0.0100 & 0.0000 & 0.0000 & 0.0000 & 0.1329 & 0.8571 & 0.0000 & 0.0000 & 0.0000 \\ 0.0200 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9800 & 0.0000 & 0.0000 \\ 0.1000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9000 & 0.0000 \\ 0.0000 & 0.0024 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.9976 \end{bmatrix} \quad \ldots(5.4)$$

Example terrain profiles generated by this Markov process are shown in Figure 25 and Figure 26. Of course, although the transition matrix T has fixed entries, the history of state transitions is different every time the model is run. As one repeats the experiment (runs the model several times) one obtains a collection of stochastic samples whose overall statistical properties are described by the fixed transition probabilities.

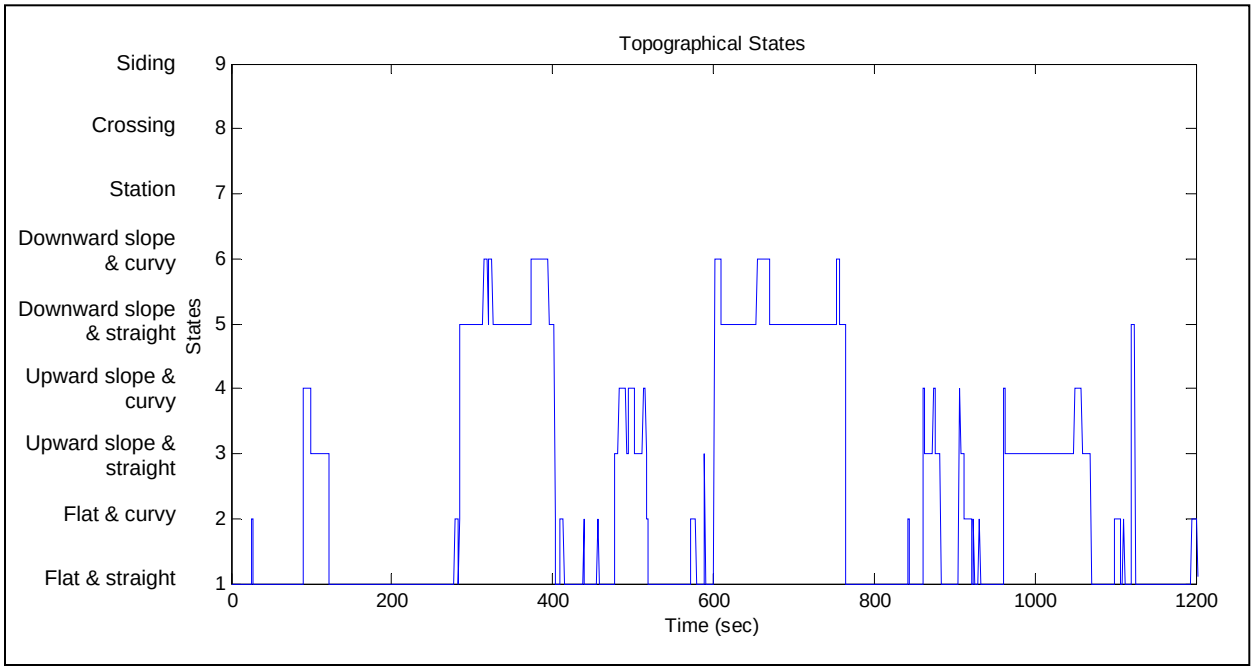

Figure 25 – Sample Terrain Profile covering a 20-minute period

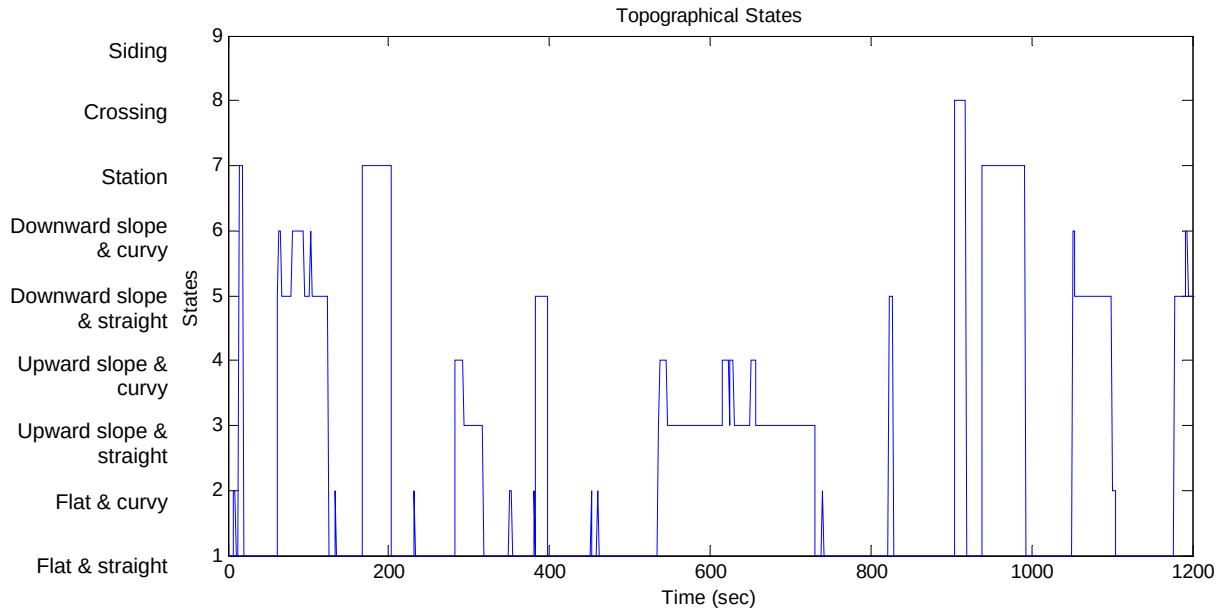

Figure 26 – Another sample terrain profile covering a 20-minute period

A more rigorous approach, had time and resources been available, would have been to conduct a field study to collect video data of the operator's inputs. A human observer could then observe and classify the time sequence of the control adjustments for representative routes. These in turn could be studied, and used to further tune and validate the model. Based on those characteristics, one can build a more accurate model to drive the Train Sentry's operation.

5.3.5 Speed Selection

Train Sentry's alerting time window and detection performance depend on the vehicle's speed. Train operating speed is based on the signals encountered, the train's weight and composition, the physical limitations of the roadbed, as well as company operating rules. For convenience, the simulation used a set of constant nominal train speeds depending on the particular topographical state.

- State 1: Flat and straight: 50 mph
- State 2: Flat and curvy: 45 mph
- State 3: Upward slope and straight: 40 mph
- State 4: Upward slope and curvy: 35 mph
- State 5: Downward slope and straight: 50 mph
- State 6: Downward slope and curvy: 45 mph
- State 7: Station: 0mph
- State 8: Crossing: 25 mph
- State 9: Siding: 3 mph

Those speeds were assumed to be adopted immediately upon state transition. The acceleration and deceleration profiles were not considered, since they do not greatly impact the Train Sentry's performance. Thus, as shown in Figure 27, the terrain 'State' input selects the appropriate velocity from the 9 x 1 'speed_matrix' input.

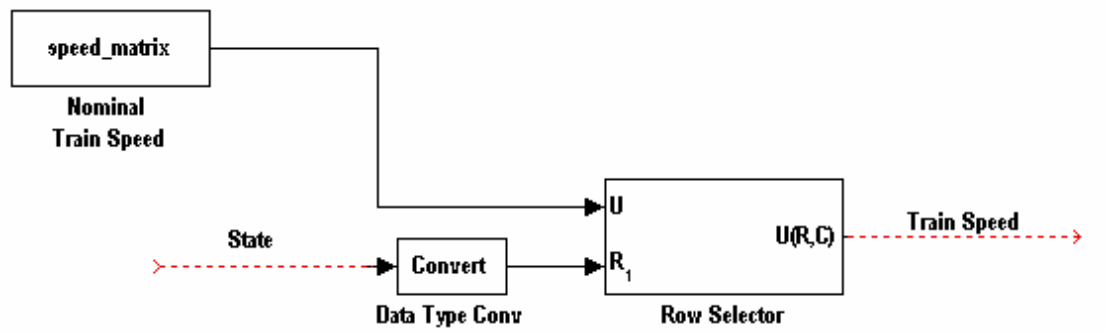

Figure 27 – Speed Selection Block: Schematic describing how the train speed is generated

Varying the train speed during simulated trips adds to the realism of the alerter's evaluation. The aim was to capture the Train Sentry device's performance across its operating range, especially at instances where the mechanism may fail, i.e. at very low speeds, around 3 mph, where a drowsiness event would be detected in at least 10 minutes. Although the difference in performance may not initially seem substantial at cruise speeds between 35 and 50 mph, one must note that if there is no activity, the first-level alerts are respectively issued at 51 and 36 seconds. Examining the effect of such time window differences on Train Sentry's performance in terms of the Time to True Alert (TTA) and the probability of correct detection and false alarm was considered necessary.

5.3.6 Eye Visibility

Before describing the operation of the DD-850, this section details the generation of the 'Eye Visibility' signal – which is subsequently fed to the DD-850 – illustrated in Figure 28. As previously stated, the built in DD-850 image analysis software cannot differentiate between 'eyes closed' and 'eyes absent'. Therefore, it is not sufficient to only model eye closure; one must also model the eyes being out of the instrument's view, depending on the operator's head

movements. To represent this in the model, the 'Awake/Asleep signal' (0 = asleep, 1 = awake) was ANDed with the binary stochastic 'headposition' signal (0 = 'look-away or go-away', 1 = '1 or 2 eyes in view'). The resulting 'Eye Visibility' signal was never 1 when the modeled operator was asleep, but could be 0 when the operator looked away or left the alerter's detection area.

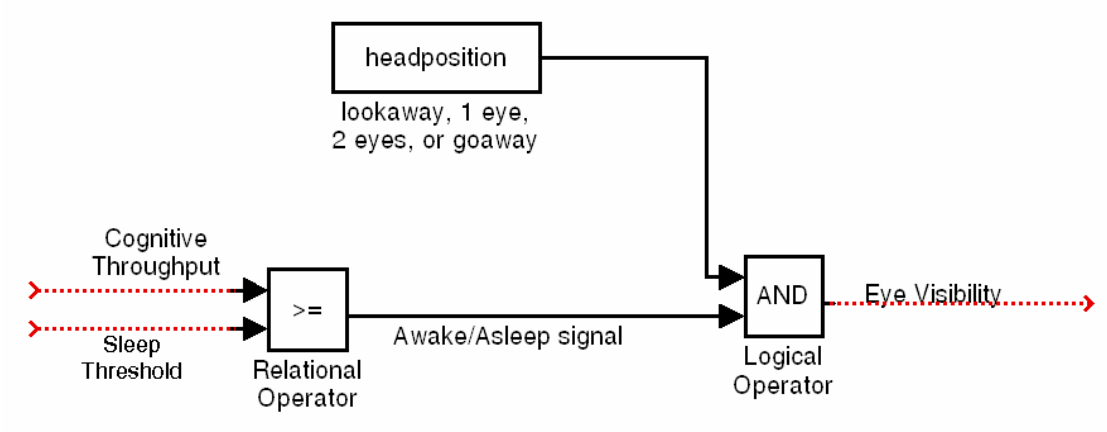

Figure 28 – Generating the 'Eye Visibility' signal to be fed to the DD-850

The 'headposition' signal was generated based on the statistics of the data collected during the Cab Characterization study described in Chapter 3. The sequences of look-aways, go-aways, and 1 or 2 eyes in view for each trip may be viewed as a sample function of a stochastic process with specific underlying probability distribution, statistical parameters, or transition probabilities. To estimate the transition probabilities, all the results of the video analysis states were first sampled at 1Hz. The transitions between the various states 0 to 0, 0 to 1, 0 to 2, 0 to 3, 1 to 0,…. etc, (0 = look-away, 1 = 1 eye, 2 = 2 eyes, 3 = go-away) were then summed and averaged, at first for both freight and passenger trips separately. The resulting matrices provided an estimate of the relative likelihood of the different transitions for each class of operation:

$$A_{passenger} = \begin{bmatrix} 0.5762 & 0.0831 & 0.3213 & 0.0194 \\ 0.0648 & 0.4738 & 0.4537 & 0.0077 \\ 0.0182 & 0.0519 & 0.9287 & 0.0012 \\ 0.0234 & 0.0175 & 0.0702 & 0.8889 \end{bmatrix} \quad \ldots(5.5)$$

$$A_{freight} = \begin{bmatrix} 0.5379 & 0.0829 & 0.3704 & 0.0088 \\ 0.0652 & 0.4894 & 0.4426 & 0.0028 \\ 0.0219 & 0.0316 & 0.9458 & 0.0007 \\ 0.0253 & 0.0380 & 0.1266 & 0.8101 \end{bmatrix} \quad \ldots(5.6)$$

Since the entries of the passenger and freight transition matrices were of the same order of magnitude, they were averaged and the following Markov Process generating matrix was used in the simulation:

$$A_{transition} = \begin{bmatrix} 0.5528 & 0.0830 & 0.3513 & 0.0129 \\ 0.0650 & 0.4819 & 0.4479 & 0.0052 \\ 0.0205 & 0.0392 & 0.9394 & 0.0009 \\ 0.0240 & 0.0240 & 0.0880 & 0.8640 \end{bmatrix} \quad \ldots(5.7)$$

To produce the 'headposition' binary signal, the resulting 4-state Markov process was transformed to one with 2 states by replacing states 0 and 3 by 0s and states 1 and 2 by 1s.

Note that the statistics of the 'headposition' binary signal were assumed independent – to first approximation – of Roadbed State and Task. In a second generation model, such dependencies could be considered.

5.3.7 DD-850

The model of the DD-850 subsystem is shown in Figure 29. It employs the definition of Wierwille et al. (1994) to output a PERCLOS value. The prototype instrument provides PERCLOS outputs averaged at the user's option over 1, 3, and 5 minutes. The current prototype implementation provides adjustable first and second level threshold values.

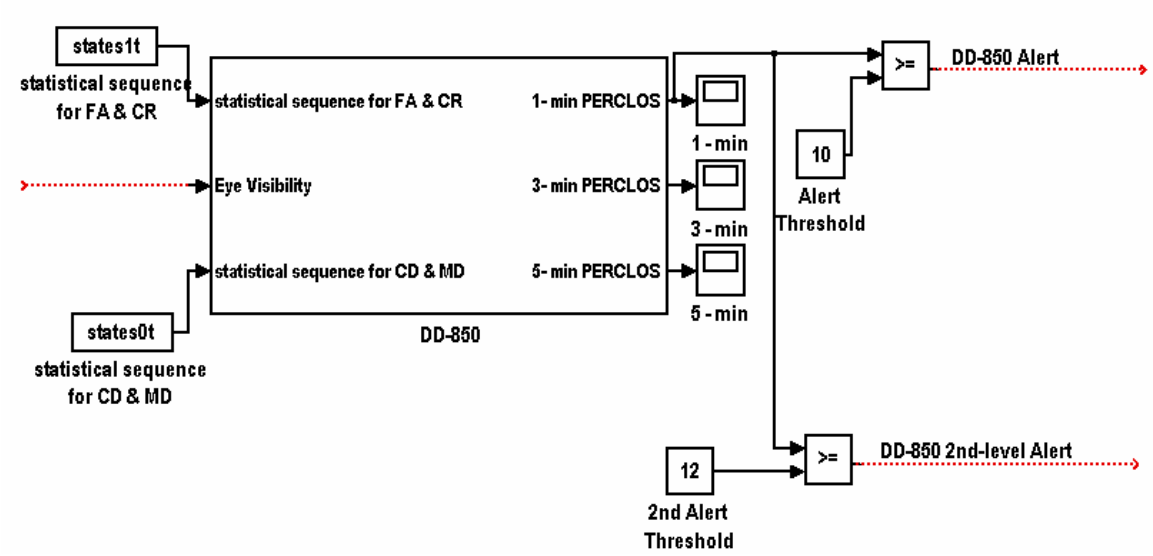

Figure 29 – DD-850 Module

The DD-850 SIMULINK model used, as shown schematically in Figure 30 and detailed below, in essence consists of two stages: 1) Front end image analysis software that estimates the presence or absence of eyes, based on the "Detected Signal" (DS). The DS is determined by an 'Eye Visibility Signal' (EVS) computed from a head position signal (look-aways, etc), the awake/asleep signal from the CTM model, and a stochastic process designed to mimic DD-850 front end false alarms and missed detection characteristics (Chapter 4). 2) An averager that computes PERCLOS over 1, 3 and 5 minute running intervals.

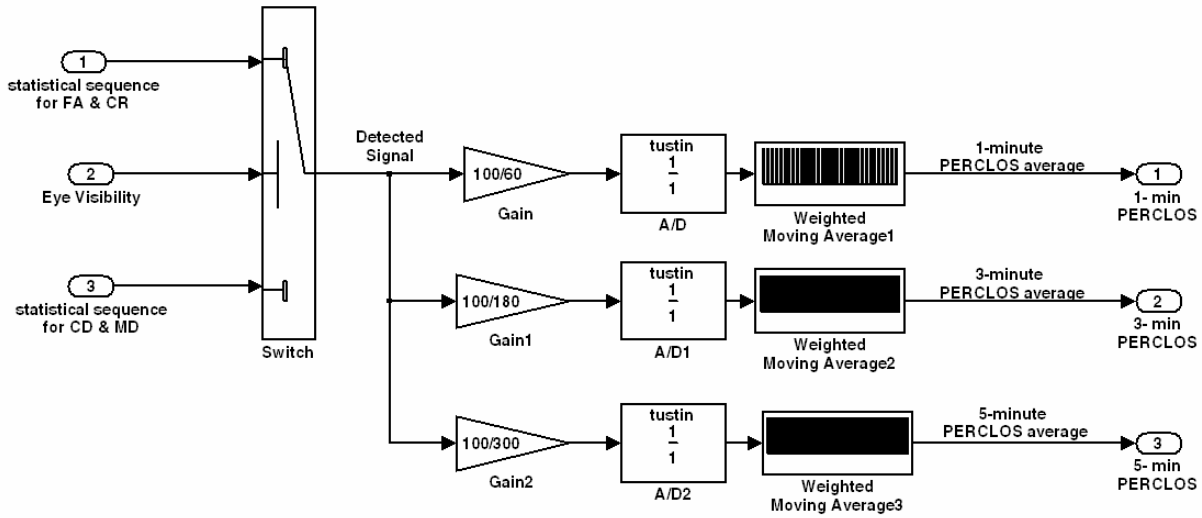

Figure 30 – DD-850 Underlying Mechanism

To describe the device's simulated FA and MD performance, and to account for the imperfections in the detection process, a stochastic approach was taken. As reviewed in Chapter 2, in terms of SDT parameters the following four conditions may occur[16]:

- **MD:** DS = 0 (1 or 2 eyes in view), when EVS = 0 (Eyes closed/absent)
- **CD:** DS = 1 (Eyes closed/absent), when EVS = 0 (Eyes closed/absent)
- **CR:** DS = 0 (1 or 2 eyes in view), when EVS = 1 (1 or 2 eyes in view)
- **FA:** DS = 1 (Eyes closed/absent), when EVS = 1 (1 or 2 eyes in view)

Based on the bench tests described in Chapter 4, one aims to construct a process that successfully detects an 'Eyes closed/absent' event with a probability of 83% (or equivalently has an MD rate of 17%), and successfully detects a '1 or 2 eyes in view' event with a probability of 89% (or has an FA rate of 11%). To accomplish this, the model incorporates a switch driven by two binary sequences, 'Statistical sequence for FA & CR' (SSFA&CR) and 'Statistical sequence for CD & MD' (SSCD&MD), as shown in Figure 30. Both stochastic sequences are generated using MATLAB statistics toolbox's *binornd* function. The SSFA&CR sequence has 1s for 83% of its samples (the rest are 0s), and SSCD&MD has zeros for 89% of its samples (and the rest are 1s). The input EVS gates the switch and controls which of the 2 signals passes at a given moment in time. If EVS = 0 then SSCD&MD passes; otherwise, SSFA&CR passes; the resulting output formed from all those segments is the 'detected signal' DS. To illustrate the operation, Figure 31 shows an example using some sample inputs.

[16] Because of the way PERCLOS is defined (it increases when no eyes are detected and decreases otherwise), the same sign designation convention was adopted for the model variable DS.

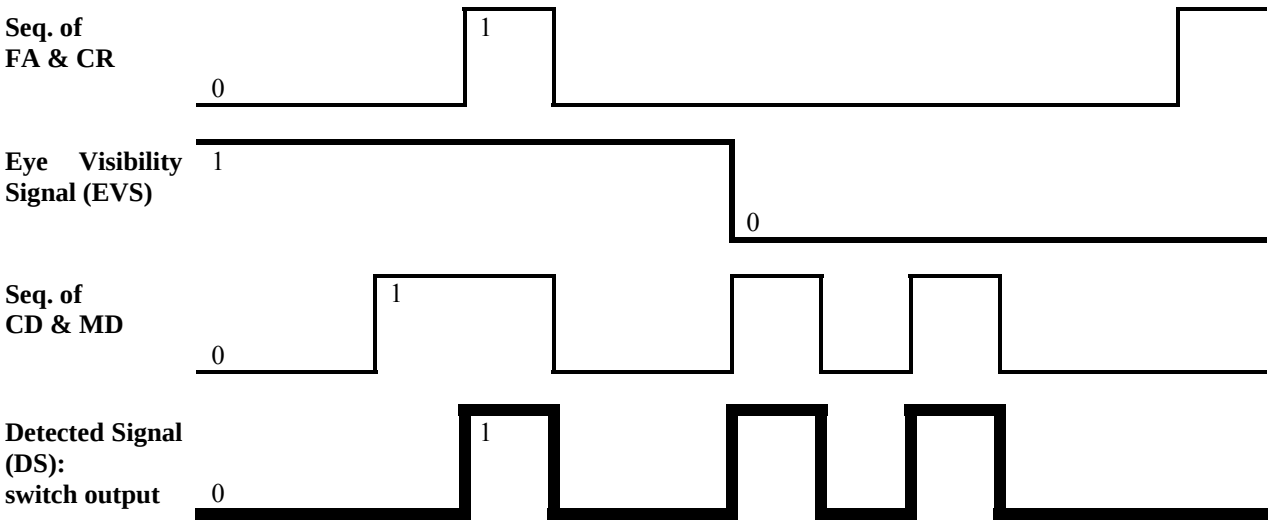

Figure 31 – Example illustrating the concept behind the switch detection process

When EVS = 1, then SSFA&CR passes, representing a stochastic process capable of providing an 89% success in correctly determining an eyes open signal. Similarly, if EVS = 0, SSCD&MD passes, corresponding to a stochastic process capable of providing an 83% in correctly identifying an eye closed/absent event.

The resulting DS output from the switch is then averaged over 1, 3, or 5 minutes. (In the simulations described in Chapter 6, only the 1-minute average was used). The gains shown in Figure 30 serve to scale the output between 0 and 100. The A/D block is used to convert the incoming continuous signals into digital signals, allowing them to be fed to the discrete-time moving averager.

To quantitatively validate the DD-850 MATLAB model's averaging architecture and determine whether it is similar to the actual instrument, a 3.5-minute sequence of DD-850's 'Detected Signal' input signal (0, 1, or 2) and 1-minute PERCLOS average output values were collected. The same 'Detected Signal' input signal was fed to the model's averager mechanism – after converting the '0s to 1s' and the '1s & 2s to 0s' – and the model's output is directly compared to the actual instrument's output in Figure 32.

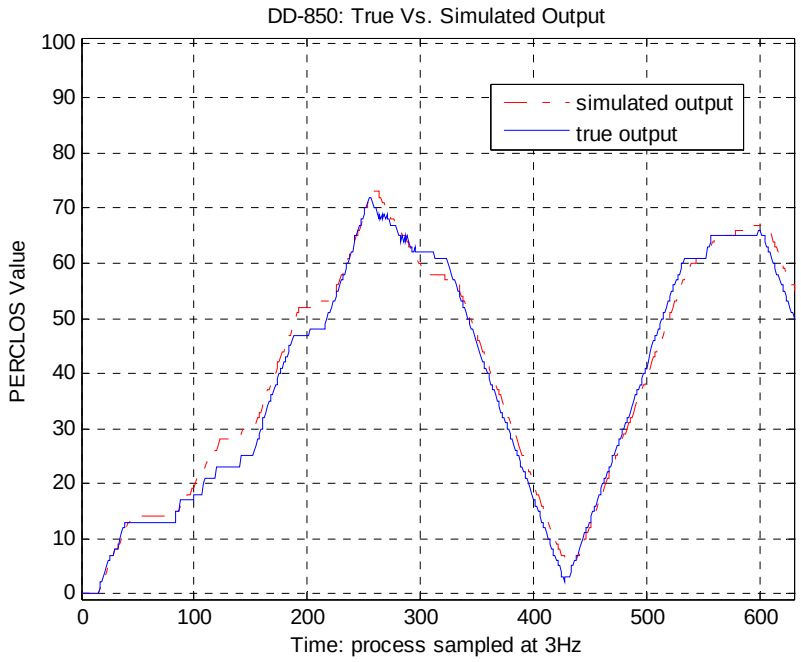

Figure 32 – Comparison of the averaging mechanism of the DD-850 device and simulation model

In sum, the simulated averaging output is consistent with the instrument's averager behavior. It reasonably fits the available data (R^2 = 0.994). Whether the model captures the instrument's *overall* behavior under operational conditions was not formally tested. Similarity between the true and simulated output depends on how well the DS parameters statistically characterize the 0-1-2 input signal to the DD-850 averager. Since DD-850's detection performance (the production of the 0-1-2 input signal) was found to considerably vary with motion and eye color, in this case the DD-850 *average* SDT statistics were used to model the DS behavior.

5.3.8 Activity Monitor

A generic Train Sentry class activity monitor, shown in Figure 33, was implemented with the representative features currently found in the field.

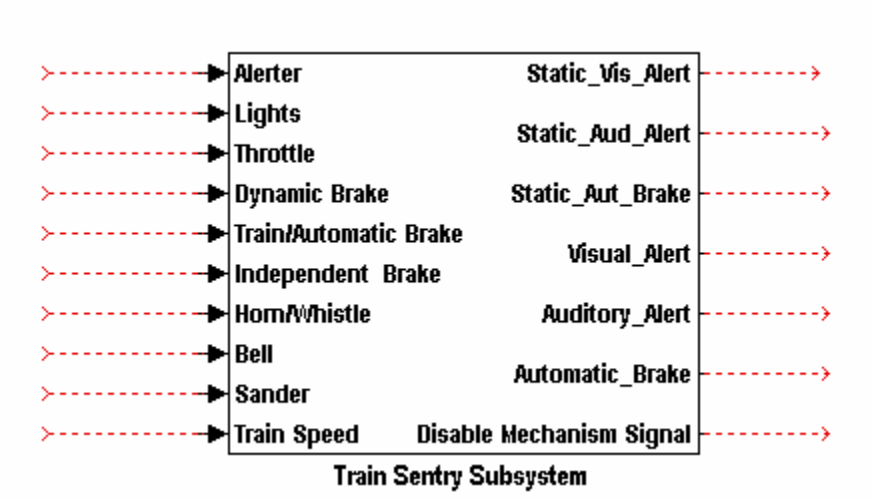

Figure 33 – Train Sentry Class Activity Monitor

As described earlier, this system monitors the engineer's adjustments to the train controls (throttle and brakes) and associated devices (horn/whistle, bell, sander, lights). If no activity is sensed after a certain time window, a visual alert is issued; if this alert is not acknowledged by depressing a switch (alerter) or other inputs, an auditory alert follows, and similarly, if no response is sensed, the train is brought to a complete stop through an automatic brake application. The internal mechanism of the generic Train Sentry model is illustrated in Figure 34.

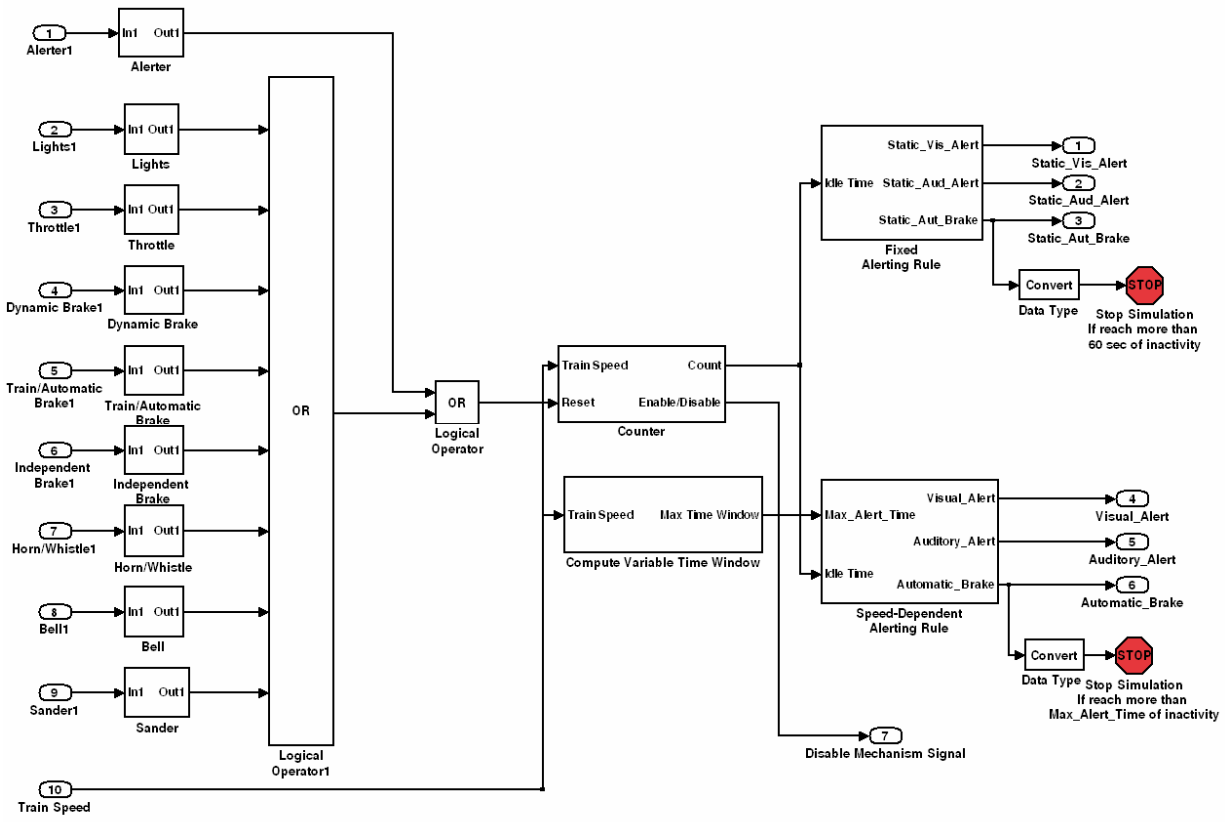

Figure 34 – Internal Mechanism of Train Sentry Class Activity Monitor

At the heart of the system lies the 'Counter' that computes the idle time. Any input to the train controls, related devices, or alerter switch can reset the counter to zero through the OR gate. To decide when to issue an alert, two alternative rules were implemented: a fixed alerting rule, and a speed dependent rule. The former uses a fixed time window (as found in the earlier versions of Train Sentry), and the latter uses a speed-dependent time window, requiring one input per mile traveled (as found in the most recent models). Both rules run in parallel, with both outputs simultaneously available to the user. As shown in Figure 20, by adjusting the switch at the output of the system, the user chooses which alerting rule is modeled as fed back to the operator's required tasks. In the speed-independent case, after 30 seconds of inactivity, a continuous visual alert, 'Static_Vis_Alert', is issued for 15 seconds, followed by a 15-second auditory alert, 'Static_Aud_Alert', ending in an automatic train brake application, 'Static_Aut_Brake', that stops the simulation if no activity is sensed in the past 60 seconds (indicating the simulated operator has fallen asleep).

The speed-dependent mechanism looks for at least one response per mile travelled by using the following relationship to determine the time windows for the alerting stages:

$$\text{Max Alert Time} = \frac{3600\text{sec/hr} \times 1\text{mile}}{\text{speed (mph)}} \quad \ldots(5.8)$$

For instance, at 50mph, the operator must make an input at least once per 72 seconds. The alerting stages were implemented based on the following plausible although arbitrary scale, with a 'Visual_Alert' at ½ Max Alert Time (36 seconds), an 'Auditory_Alert' at $^2/_3$ Max Alert Time (48 seconds), and an 'Automatic_Brake' at Max Alert Time. For the simulation, this speed-dependent alerting rule will always be used unless noted otherwise.

The actual train sentry is disabled at very low speeds to avoid unnecessary alerts. In the simulation the entire counter mechanism was disabled at speeds below 3 mph, when stopping at a station for example. The 'Disable Mechanism Signal' was used to disable further counter mechanisms in other modules that compute the alerter's SDT statistics. The imperfections in the detection process caused by switch contact wear due to excessive use are not modeled; all control movements produce the appropriate responses.

5.3.9 Arbiter

The 'Arbiter' block in Figure 35 combines two independent decisions inputs form the DD-850 and Train Sentry. This preliminary simulation used a simple internal AND gate to generate an overall tandem alert. The goal was to see if tandem detectors, arbitrated using even this relatively straightforward strategy, could reduce FAs.

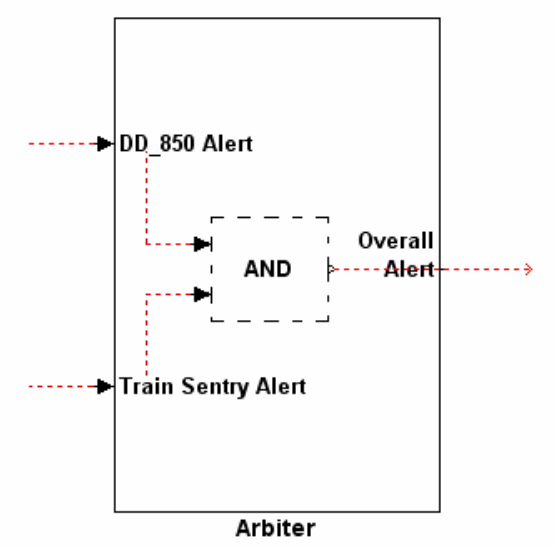

Figure 35 – Arbiter subsystem showing internal AND mechanism

The use of other potential fusion strategies will be discussed in the next chapter. For example, other rules may be implemented by weighting the inputs judiciously, giving more importance to one of the two alerts depending on their particular strengths and weaknesses in different situations. Also, rather than simply combining DD-850 and Train Sentry decisions at the subsystem output level, one could merge *raw* measurements of the PERCLOS average values and of the idle time counter to make decisions using certain algorithms. In short, this simulation platform permits exploring many different design options very quickly by making minor alterations to the arbiter's internal mechanisms.

5.3.10 SDT Statistics Computation

As shown earlier in Figure 20, each standalone alerter and the arbiter has a block associated with it to calculate its SDT parameters for first and second level alerts. Figure 36 shows the block for the Train Sentry class activity monitor. It is designed to estimate the SDT performance statistics as if the Train Sentry were operating on a standalone basis during the trip (i.e. the inputs to the Train Sentry are not influenced by the DD-850 or Arbiter portions of the trip simulation).

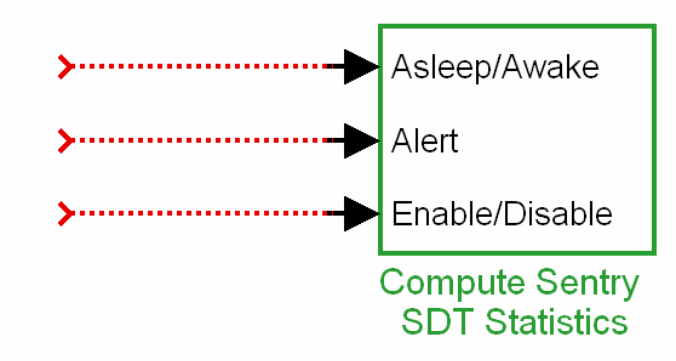

Figure 36 – Block to calculate the SDT parameters of the Train Sentry class monitor

The internal mechanism of the Train Sentry SDT statistics block is shown in Figure 37 below. The blocks for the DD-850 and Arbiter are conceptually similar with the exception that statistics on the Enable/Disable signal are recorded for the Train Sentry but not the DD-850 or Arbiter.

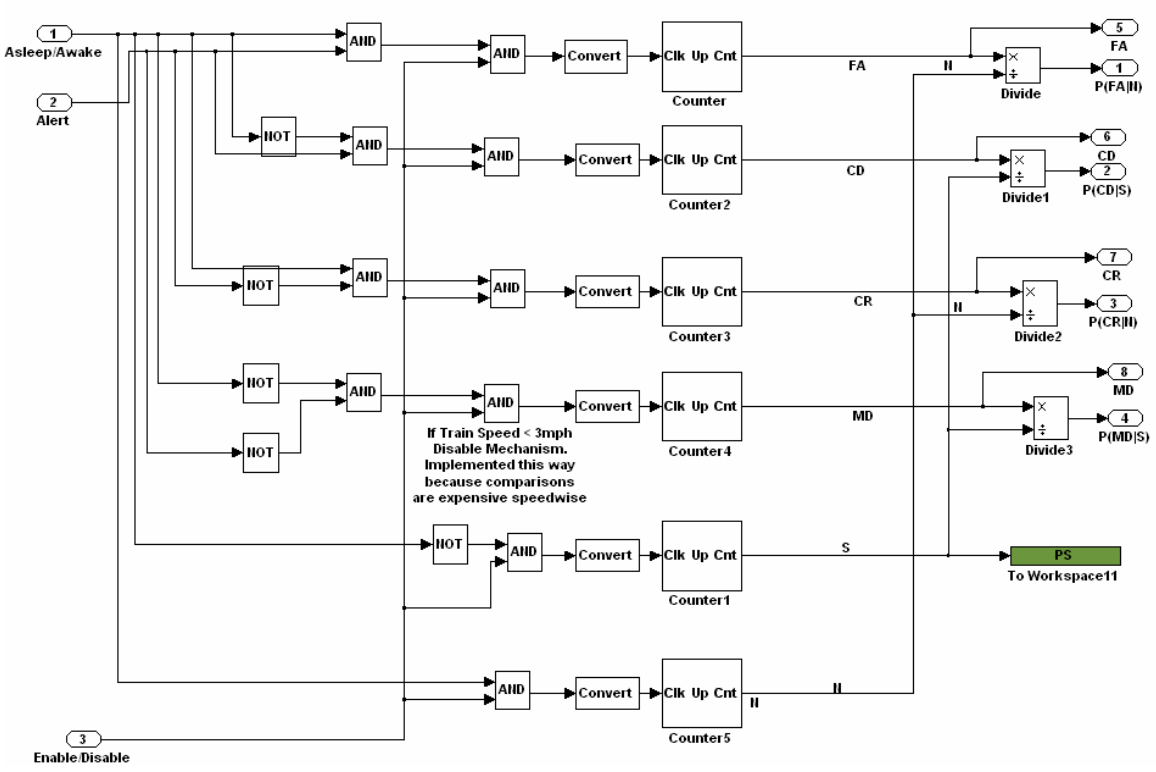

Figure 37 – Internal mechanism to compute the 4 SDT parameters: output count and probabilities. The rates are computed offline by a MATLAB program.

During every second of the simulated trip, the logic architecture of ANDs and NOTs uses the 2 input signals'Asleep/Awake' and 'Alert' to assess the presence of any one of the following 4 conditions:

- **MD:** Asleep/Awake = 0 (i.e. Asleep) and Alert = 0 (no alert issued)
- **CD:** Asleep/Awake = 0 (i.e. Asleep) and Alert = 1 (alert issued)
- **CR:** Asleep/Awake = 1 (i.e. Awake) and Alert = 0 (no alert issued)
- **FA:** Asleep/Awake = 1 (i.e. Awake) and Alert = 1 (alert issued)

The S (for signal or H_0) and N (for noise or H_1) are simply the number of seconds in which the operator is respectively asleep or awake. The FA, CD, CR and MD counters simply add up the number of seconds for which each criterion is satisfied during the simulated trip. Those counts are then used to compute the respective rates (events/hour) by the outer shell MATLAB Monte Carlo control script, MonteCarloSim.m. The 4 SDT conditional probabilities P(MD|S), P(CD|S), P(CR|N), and P(FA|N) for each simulated trip were calculated by simply dividing the count of each parameter by the count of instances where the operator is either awake or asleep. P(CD|S) thus measures the percentage of time during the trip when the operator was asleep with the associated alerter triggered. Hence P(MD|S) = 1- P(CD|S). Similarly, P(FA|N)= 1 - P(CR|N). Note that the Train Sentry was not designed to operate when the train speed drops below 3 mph. Also note that separate statistics modules were used to calculate trip detection percentages for both first and second level alerts. To avoid biasing the standalone Train Sentry statistics, the total awake and asleep time counters were not enabled when the train speed dropped below 3 mph. This adjustment was not required for DD-850 or Arbiter statistics, however.

Also, it should be noted that the trip simulations were designed assuming that the PERCLOS and/or Arbiter based detectors were "overlay" systems, i.e. it was assumed that a Train Sentry alerter system was always onboard, and so the trip simulation always proceeded until the Train Sentry reached its third level of alert, and stopped the train.

It is important to state that the SDT probabilities calculated by the model statistics modules estimate the *percentage of time* during each simulated trip that the alerter was on (or off) when the operator was awake (or asleep). However, since the detectors have response dynamics, they do not respond immediately. For instance, when the operator first falls asleep, the number of

MDs is continuously accumulated until the alert is finally issued (taking several seconds or several tens of seconds). This means that the MD probability, computed on an "asleep time averaged basis", tends to overestimate the MD probability computed on an event basis (and thus underestimate the CD probability computed on an event basis). Therefore, the CD rate depends on the detectors speed in issuing a correct alert. The faster the instrument responds, i.e. the smaller the time to alert, the greater the CD rate. By the same argument, the "awake time averaged" FA rate may slightly overestimate the event based FA rate in situations where the operator subsequently fell asleep, and thus slightly underestimate the event based CR rate. These caveats should be considered when evaluating the probabilities resulting from the simulation. An alternative approach would have been to estimate the MD probability on an event basis, by examining the total number of times a sleep event occurred when the detector subsequently failed to respond at all, ignoring detector latency. However this would have required many more Monte-Carlo iterations.

Finally, although both are calculated, it is important to note that CD and FA probabilities are not the same as CD and FA rates, which depend on the incidence of sleep episodes. The CD and FA rates establish the trip's cost and the system's acceptability as experienced by the operator.

For Alerting Systems with significant temporal dynamics, measures of response time are as important as SDT probabilities and rates. The Time to a True Alert (TTA) was measured in the simulation as another index of performance. From the operator's perspective it is the temporal dimension of $P(CD|H_0)$. For example, at low train speeds, an activity monitor operating on a speed-dependent rule may take several minutes to detect a nonreponsive individual. Similarly, it may take several tens of seconds for the DD-850 to issue an alert if the threshold is set too high. Quantifying the TTA permits a richer evaluation of the alerting strategy and from a wider perspective. In the simulation, when the operator falls asleep, the TTA was measured by counting the number of seconds elapsed between the incidence of the drowsiness event and the alert; this value was directly obtained from the MD count. MD count gives the number of seconds during which the operator is asleep and no alert is issued, i.e. it provides the number of seconds between the occurrence of a sleep event and of an alert event, which is precisely TTA.

Use of "awake/asleep" time averaged SDT parameters is unconventional, but necessary. However, since the same definition was consistently applied throughout the simulation and analysis phases, and since all comparisons are made on the same set of rules, the comparisons between alternative system architectures should to be valid.

5.3.11 Monte Carlo Simulation Control Program

To regulate the flow of the simulation, a MATLAB control script, *MonteCarloSim.m*, was developed. The details of the Monte Carlo simulation will be described in more detail in Chapter 6. The program sequentially performs the following four major steps::

1) Initialize variables: type of operation, trip duration, number of Monte Carlo iterations
2) Generate model inputs (that remain constant across all trips)
    a) Load the Cognitive Throughput Model and run the Simulink Model for 38 hours (just to cover all necessary input cases, see Chapter 6 for the illustrative case studies).
    b) Extract the Cognitive Throughput data for the specific duration of the trip and use it as an input to the simulation (*CTM* block shown in Figure 20)
    c) Initialize the SDT parameters
    d) Generate the Sleep Thresholds for the overall set of trips (one threshold will be used for every single trip in each iteration below)

3) Iterate (each iteration represents a trip) {
    a) Generate trip input: terrain (*states_Markov*), tasks (*task_matrix*), speed (*speed_matrix*) engineer head position (*headposition*), DD-850 front-end SDT statistics (*states0t, states1t*).
    b) Select the sleep threshold for the present run.
    c) Run the Simulink model shown in Figure 20.
    d) Update SDT parameters.
}
4) Compute and output the average SDT parameters of all runs

5.4 Conclusion

In sum, the development of the overall simulation tool required making several tradeoffs between modeling reality and answering the study's proposed key questions. Thus, after describing the overall model structure, the next chapter illustrates its use to examine different case scenarios and to assess the performance of the alerters both individually and in tandem.

Chapter 6

Simulation Case Studies: Alerter Performance Evaluation

Chapter 2 motivated the use of a systems engineering, SDT approach to tackle the problem of Alerter System (AS) design. Briefly, in such framework, the value of the cost function J (in dollars) is used as the common index of comparison between different design options. The cost depends not only on the $P(CD|H_0)$ and $P(FA|H_1)$ but also on the probability of the signal's presence, $P(H_0)$, which is believed to vary depending on the type of railroad operation. Three generic types of operations were modeled: regular nighttime freight, night extra board freight, and regular daytime passenger. This chapter illustrates the use of the simulation model developed in Chapter 5 to explore the AS design space for these three types of operations. One major goal was to examine how the performance of the proposed tandem AS approach compared to the performance of the individual alerters operating on a standalone basis as originally designed.

6.1 Monte Carlo Simulation

To capture the statistical properties of the simulation signals, the MATLAB trip simulation was imbedded in a Monte Carlo simulation control program, which iterated the trip many times to obtain mean values of the SDT probabilities, rates, and alerter response times. The variability in the resulting data sets determines the number of iterations required in a Monte Carlo simulation. Some applications in nuclear physics may require 10,000 iterations, whereas some environmental studies may find 1,000 iterations to be sufficiently large to ensure convergence of the data and reduce the parameters confidence intervals to acceptable values.

To empirically determine the appropriate number of iterations for this experiment, the simulation was initially repeated 500 times. The sleep threshold was set to zero. Since there were no drowsiness events to stop the simulation, the CD and MD rates were all zero. Thus, the first-level FA rate was chosen as the representative numerical performance index (the CR rate could have equally been used). Initially the DD-850 alerting thresholds were set to 10 and 12%, the values recommended by Attention Technology.

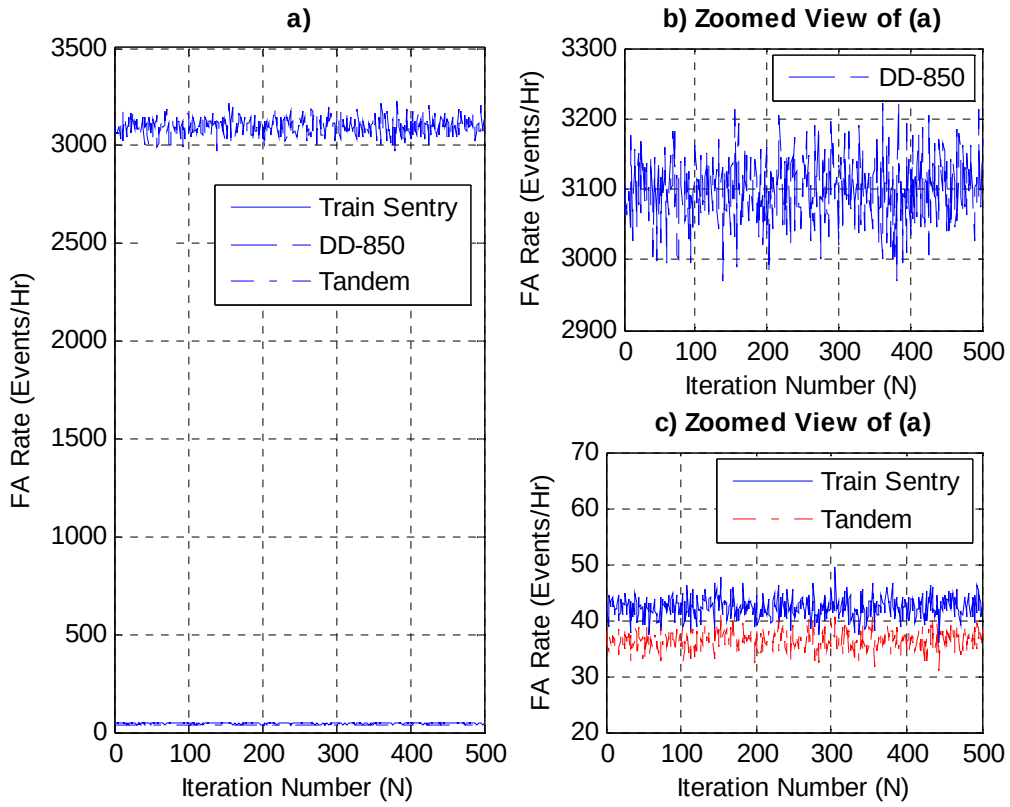

Figure 38 – a) False alarm rate (events/hour) for all three design options after each of the 500 experiment iterations. b) Zoomed view of the DD-850 FA rate performance displayed in (a). c) Zoomed view of the Train Sentry and the Tandem configuration performance displayed in (a).

Figure 38a) presents the FA rates for the Train Sentry, DD-850 and Tandem configurations in each of the N = 500 simulation runs. Due to the very high DD-850 FA rate with the manufacturer recommended thresholds (which will be addressed in the next section), one must zoom in to observe the respective behavior of the DD-850 in Figure 38b), and the Train Sentry and Tandem performance in Figure 38c). All rates stabilize around certain mean values, with the DD-850 results displaying the greatest variability with about +/-100 events/hour. To determine the

number of repetitions beyond which these rates converge to stable values, the respective running averages were computed and plotted in Figure 39.

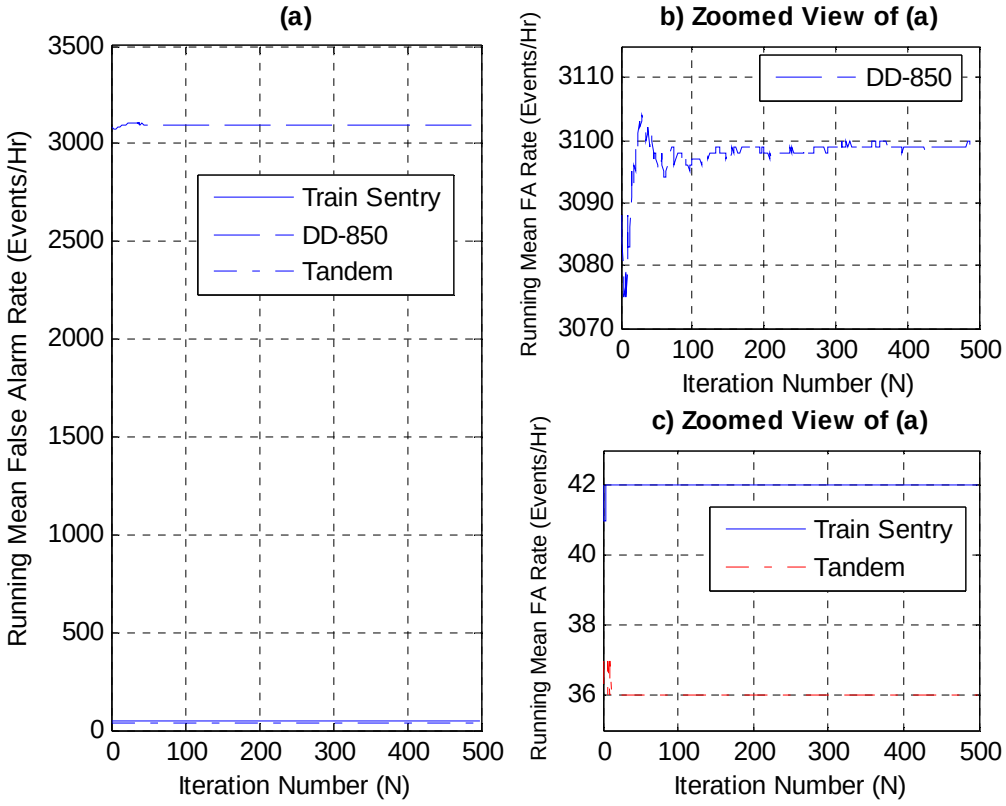

Figure 39 – a)Mean of the FA rates (events/hour) as a function of the number of iterations for all three design options after each of the 500 experiment iterations. b) Zoomed view of the DD-850 running mean FA rate displayed in (a). c) Zoomed view of the Train Sentry and the Tandem configuration mean FA rate displayed in (a).

Performance for the 3 cases is shown in Figure 39a). The mean FA rates vary for small numbers of iterations and become stable as the iteration number increases. A better view of the mean behavior is depicted in Figure 39b) & c). In the former subplot, one notices that it takes about 300 iterations for the mean to reach 3099 +/- 1 FAs/hr, while in the latter subplot, the means of the Train Sentry and Tandem configuration converge very early below N = 50 iterations. To determine the least number of iterations to obtain adequate convergence, the mean rates in Figure 39b) & c) were respectively plotted in Figure 40a) and b).

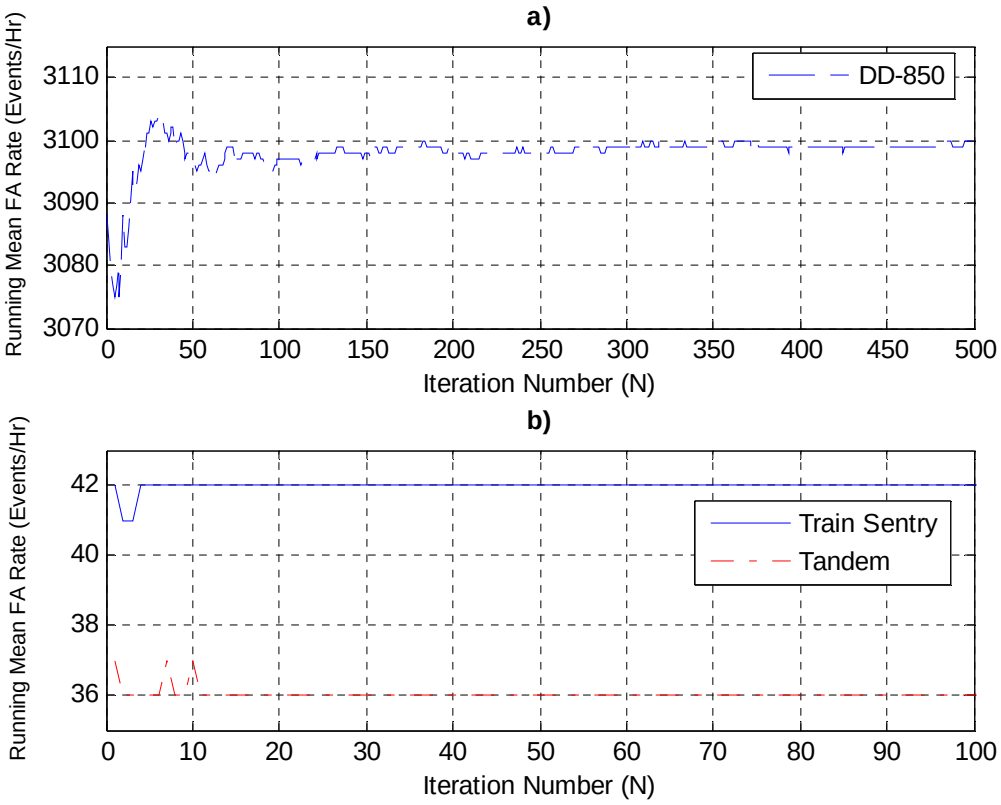

Figure 40 – Zoom view of Figure 39b) and c). In a) the mean value requires about 300 iterations to reach a mean of 3099 FAs/hr and remain within 1 FA/hr of that value. Around N = 11, the means of the FA rates converge.

In Figure 40a), after 300 iterations, the DD-850 mean value seems to converge to 3099 FAs/hr and varies thereafter by 1 FA/hr. However, it takes about 100 iterations for the mean to reach a level of 3098 FAs/hr and remain within 2 FAs/hour. The Train Sentry and the Tandem mean values converge early after about 11 iterations. It was therefore deemed sufficient to iterate the simulation for about 100 times to ensure that the parameters' average value displayed adequate convergence behavior. Furthermore, to quantitatively evaluate the situation, the FA rates were also collected in Table 12.

Table 12 – Mean running false alarm rate variation in terms of the number of iterations in the Monte Carlo simulation. Each entry represents mean FA rate +/- 1 SD (events/hour).

Number of Iterations	Train Sentry Standalone	DD-850 Standalone	Tandem Configuration
10	42 +/- 2	3079 +/- 25	37 +/- 1
20	42 +/- 2	3096 +/- 42	36 +/- 1
50	42 +/- 2	3097 +/- 41	36 +/- 2
100	42 +/- 2	3097 +/- 45	36 +/- 2
500	42 +/- 2	3099 +/- 45	36 +/- 2

Also, given the number of assumptions built into the simulation and the purpose of understanding general trends, the order of magnitude of the mean values of the FA rates was of primary interest. There was little advantage at running the simulation 500 times. Therefore, in the following case studies, a Monte Carlo simulation with N = 100 iterations was deemed sufficient to calculate all final mean SDT parameters.

Before applying the simulation tool developed in Chapter 5 to study alerting performance in three illustrative railroad case studies, the appropriateness of the manufacturer's DD-850 alert threshold setting recommendation was first investigated:.

6.2 DD-850 Threshold Adjustments

The manufacturer, Attention Technologies, Inc., suggested the DD-850 first-level alert threshold to be between 8-10%, and the full alert threshold to be around 12% (Attention Technology DD-850 instructions). The simulation (Chapter 5), which effectively brings together the eye visibility data from the cab environment study (Chapter 3) with SDT data on front end sensor software performance (Chapter 4), predicts that the manufacturer's recommended values are low enough that a very large number of false alarms will result, probably because the running values of PERCLOS are expected to be higher than the manufacturer assumed, due to look-aways and the relatively high front end false alarm rate. As shown in Table 13, the predicted DD-850 PERCLOS averager FA rates in operational conditions are predicted to be 3097 1st level FAs/hr and 2358 2nd level FAs/hr in Table 13. Therefore, one may use this simulation tool to find threshold values that are optimal or satisficing in terms of PERCLOS averager FA rates acceptability.

Table 13 – Summary of FA performance of the three possible detection configurations with DD-850 thresholds at 10 and 12%

Configuration	1st-Level FA rate (events/hour)	2nd-Level FA rate (events/hour)
Train Sentry Standalone	42	0
DD-850 Standalone	3097	2358
Tandem Configuration	36	0

The performance of the DD-850 depends on front end detection reliability and eye visibility fluctuations due to engineer head movements. Therefore, three possible options are available to alter the DD-850 engineering model to improve its inadequate performance: a) one may either increase the alert threshold, b) use the 3 or 5-minute PERCLOS to issue the alerts, or c) make both modifications. The dynamics of the 3 and 5 minute PERCLOS running averages are relatively slow, so they may take too much time to detect a true drowsiness event. On the other hand, as shown in Figure 41, the 1-minute PERCLOS running average hovers mostly between 10 and 20%, with a mean value of 15.6% and a standard deviation of 6.67%. Therefore, by simply increasing the threshold to 35% (mean + 3 SD), it should be possible to eliminate a large number of false alarms. In fact, after several simulation trials, a threshold of 50% seems a better alternative. However, increasing the threshold too much would require more time for the averager to detect a drowsiness lapse. In comparison to the Train Sentry class monitor, which issues an alert after at least 36 seconds of inactivity, the simulation predicts that the DD-850 with a 50% threshold would take at most 30 seconds to detect the absence of open eyes (starting from a 0% PERCLOS level). Therefore, from the arbiter's perspective, having the two instrument alert delays of the same order of magnitude is beneficial to reduce the overall alert latencies, since the AND arbiter's time to alert is dependent on the "slowest" instrument.

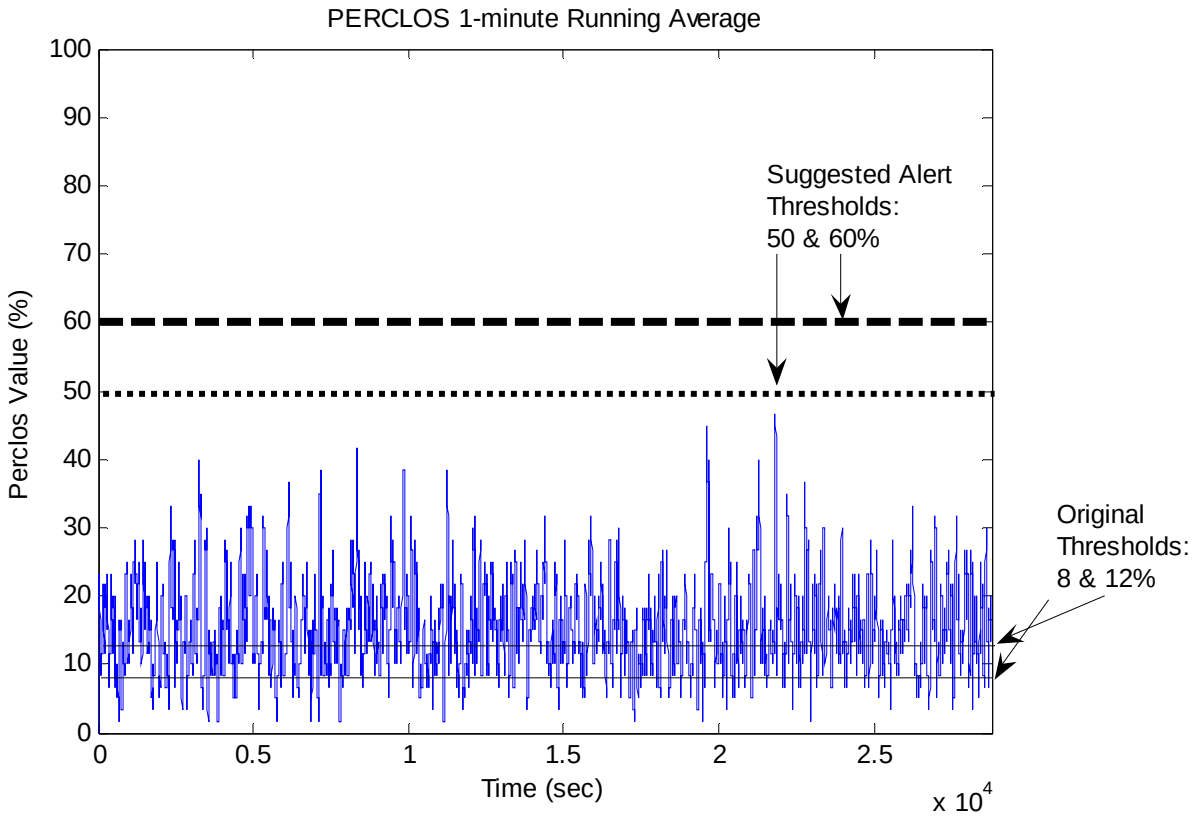

Figure 41 – One–minute PERCLOS running average on a sample 8-hour trip.

Since the alerter has two levels of alerts – a full alert at 12%, and a first-level alert adjustable between 8-10% ($^2/_3$ - $^5/_6$ of 12%) – a potential recommendation from the simulation is to set the first-level alert between 35-50% and the full alert at 60%. After implementing the latter suggestion (with the first-level threshold at 50% rather than at 35% to keep the time between the alerts the shortest, the simulation was run again, and results summarized in Table 14 were obtained.

Table 14 – Summary of FA rate performance of the three possible detection configurations, and improvement due to theDD-850 threshold adjustment of 50% and 60%

Configuration	1st-Level FA rate	2nd-Level FA rate
Train Sentry Standalone	42 (no improvement)	0 (no improvement)
DD-850 Standalone	12 (257% improvement)	< 1 (2357% improvement)
Tandem Configuration	< 1 (350% improvement)	0 (0% improvement)

The performance of the DD-850 improved substantially as a direct result of the threshold adjustment: ~250% decrease in first-level FA rates and ~2300% fold decrease in second-level false alerts. The tandem configuration first-level FA rate decreased by ~350%, and was again lower than the FA rate of both individual alerters. Notice that the simulation predicts that the performance of the tandem configuration is improved at the first alarm level. With regards to the second level alarm, since the Train Sentry auditory alarm is not expected to awaken the operator (i.e. a sleepy operator will remain asleep until automatic train brake application), then there are no second level false alarms as a consequence of the way the simulation was implemented. However, this threshold adjustment was solely conducted at the DD-850 instrument level and not at the system level. To determine the system-wide optimal DD-850 alert threshold one must also examine the SDT conditional probabilities and the time to true alert (TTA), because at the expense of decreasing the FA rate, one may be increasing P(CD|S) or TTA. Thus, the necessary tradeoffs will be made in all three case studies to select optimal DD-850 thresholds to obtain the best overall alerting performance.

6.3 Illustrative Application 1: Regular Nighttime Freight Operations

6.3.1 Simulation Assumptions:

Although in scheduled operations engineers may know their uniform work timetables months ahead of time, they may be operating at hours where they may be fighting their biological clock's sleep drive, for various reasons. They are thus expected to experience some severe drowsiness episodes, and some percentage of these will result in a sleep episode. As previously discussed, there are no solid data revealing the probability of a drowsiness episode, but Raslear and Coplen's (2004) estimate of 0.16 (i.e. once out of 6 trips) based on Pollard's (1996) study was used.

For the Regular Nighttime Freight scenario, a typical schedule of a freight engineer operating on the line between Boston and Selkirk was used. The engineer typically goes to sleep at midnight and wakes up the following day between 6:30am and 8am, depending on how busy he is on that day. He takes no naps in the afternoon. He reports to work at 6:15pm, leaves Boston between

7pm and 8pm, and arrives in Selkirk around 4am. Thus, for the simulation, the CT model was used with midnight sleep time and 8am wakeup time; the alertness level input to the simulation is the segment of the CT time history between 8pm and 4am as shown in Figure 42. The engineer is assumed to operate the train continuously, with no breaks.

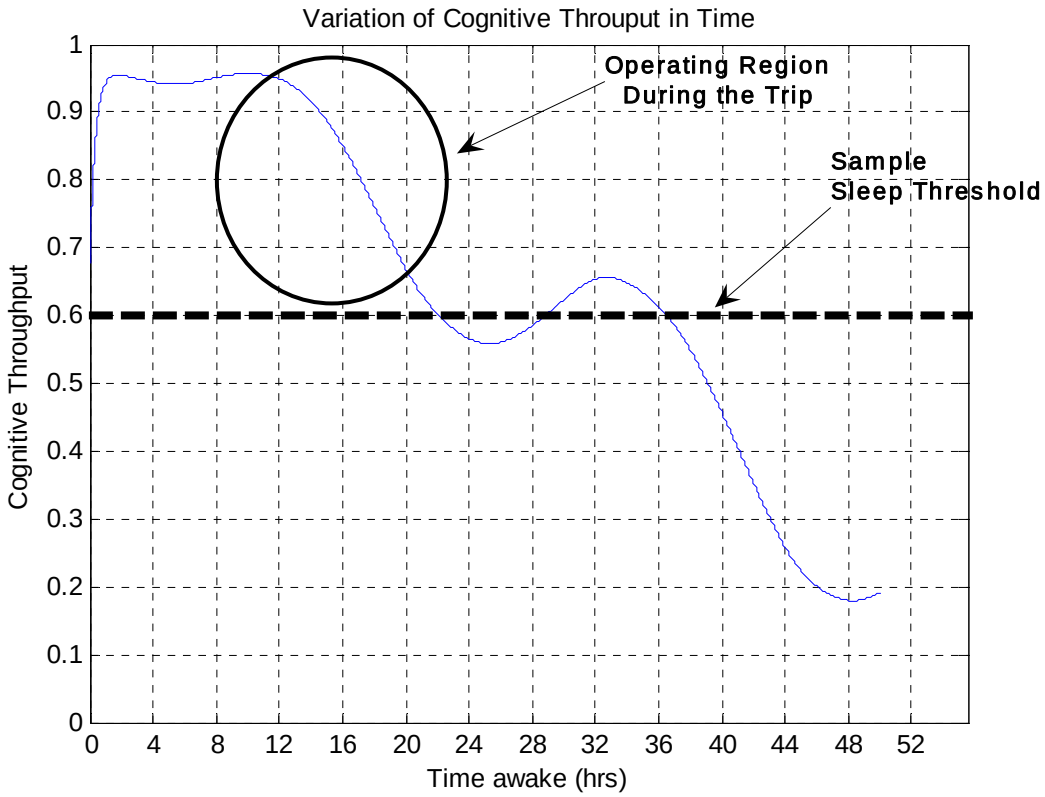

Figure 42 – Circled region comprises the engineer's alertness level during a scheduled nighttime freight trip between 8pm and 4am. A sample threshold is shown: since the alertness level never drops below the threshold, the operator does not experience any drowsiness episodes on this particular trip.

The fatigue effects of the driving task itself were not considered. It was also assumed that the trip is conducted in darkness, < 10 lux as measured in the Cab Characterization Study in Chapter 3. It was also assumed that the DD-850 motion sensitivity problem described in Chapter 4 was resolved, and that the front end displays the same performance as measured under ideal darkness conditions in Chapter 4, Table 9 i.e. with a P(CR) of 89% and a P(CD) of 83%.

6.3.2 Results:

After running the simulation for N = 600 times, to obtain about 100 trips in which the simulated operator experienced a drowsiness event (1:6 ratio), the detection performance was assessed and is tabulated in Table 15 below. As previously mentioned a sleepy operator will remain asleep until automatic train brake application, which then ends the simulation. A direct result of this convention was that the 2nd level FA rate of the Train Sentry is zero; i.e. no false alarms occurred. In general, the 2nd-level false alerts were not very instructive due to the way the simulation was implemented (see section 6.2), with the exception of TTA; therefore, in the analysis, only the 1st-level alerts will be addressed; the significance of the 2nd level alerts however will only be discussed further below.

Table 15 – Performance summary of the three different design options

Configuration	1st –Level Alerts				2nd -Level Alerts			
	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$
Train Sentry Standalone	42	25	0.012	0.64	0	37	0	0.43
DD-850 Standalone	3092	< 1	0.86	0.99	2355	< 1	0.66	0.98
Tandem	36	25	0.010	0.64	0	37	0	0.43

At the 1st-level, the Train Sentry FA rate is of the order of 42 FAs/hour. Of course, this value depends on many variables (type of operation, speed, territory), so an informal study was conducted to verify whether such figure roughly agrees with what is experienced in the field. Two engineers in passenger operations driving between Boston, MA and Worcester, MA recorded the frequency of the 1st-level visual alerts experienced. The type of activity monitoring system and the alerting interstimulus interval were not made available to us. The reported results are listed in Table 1Table 16. Each row represents a certain leg of the trip.

Table 16 – Summary of the 1st-level visual alerts from two engineers in passenger operations between Boston and Worcester, MA.

FA count	Time Period (min)	FA Rate (FAs/hr)
11	37	18
20	20	60
2	9	22
8	9	53
Total: 41	75	32

The reported rate ranged between 18 and 60 FAs/hr, with a mean of 32 FAs/hr. Therefore, the ballpark figure of 42 FAs/hour generated by the simulation is within the same order of magnitude of what is experienced in the field. However, this value should not be seen as an emergent property of the model. The FA rate is dependent on the terrain model and the interstimulus alerting interval, where the former was tuned to yield qualitatively plausible terrain type and operator activity. Therefore, this agreement was a fortunate occurrence.

Due to the lack of data sets describing the frequency of the operator's controls adjustment, the simulation might have shown a discrepancy arising from the way tasks are generated through the Roadbed model. Thus to obtain more accurate results it is appropriate to collect video data and quantify the operator's frequency of controls adjustment.

Since a reduction of 6 FAs/hour was not seen substantial, another way was sought to improve the system's performance. To obtain the best DD-850 alert threshold adjustment, the two-level alert thresholds were varied and the performance of each configuration is listed in Table 17 and Table 18. Table 17 predicts the false alarm rate (i.e. performance as experienced by the operator) and Table 18 describes the probability of occurrence of an FA or CD at a given time from the SDT analysis.

Table 17 - Summary of the 3 configuration FA rate and TTA performance variation as a function of choice of DD-850 threshold

	Sentry			**DD-850**			**Tandem**	
DD-850 Threshold Variation	**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**
10/12*	42	25		3,092	< 1		36	25
30/36	42	29		143	12		2	31
35/42	42	30		65	15		< 1	31
50/60	42	27		14	31		< 1	36

* n/m = 1st level alert threshold set at 'n', and 2nd level alert threshold set at 'm'

Table 18 – Summary of the 3 configuration SDT probabilities performance variation as a function of the DD-850 threshold variation

	Sentry			DD-850			Tandem	
DD-850 Threshold Variation	**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**
10/12*	0.012	0.64		0.86	0.99		0.010	0.64
30/36	0.012	0.61		0.040	0.80		0.00044	0.58
35/42	0.012	0.60		0.018	0.76		0.00021	0.55
50/60	0.012	0.61		0.0014	0.48		0.000028	0.43

In such setting, varying the DD-850 threshold was not expected to affect the Train Sentry operating point. However, if the DD-850 alert was linked to the Train Sentry so that a first level DD-850 alert required a Train Sentry switch acknowledgment, then indeed, varying the DD-850 threshold would affect the Train Sentry's performance. The observed variation may be due to the fact that, in the calculation of TTA and $P(CD|H_0)$, there were less data points available (~100) than for the other variables (~200,000) in the Monte Carlo simulation, such as the FA rate. However, since the focus of this work was on orders of magnitude, the 5 sec TTA variation and the 4% $P(CD|H_0)$ fluctuation were not believed to have a major impact on the overall results and conclusions. Again, combining the DD-850's faster response and the Train Sentry's lower FA rate, the tandem system consistently displayed at least equivalent or superior performance with respect to the status quo (i.e. Train Sentry alone). Of course, since the arbiter must wait for both instruments to concurrently issue an alert, the resulting arbiter will be limited by the TTA of the slowest alerter.

For both the DD-850 and the tandem configuration, an increase in the DD-850 alert threshold decreases the FA rate, increases TTA, and decreases both $P(FA|H_1)$ and $P(CD|H_0)$. Thus, one can reduce the occurrence of FAs at the expense of a longer delay in true alert response. By raising the threshold to 30/36, ~ 3 SDs above the PERLCOS mean value, a considerable change is noticed at the tandem system level with respect to the 10/12 threshold setting; the FA rate decreases by one order of magnitude to 2 FAs per hour while costing about 6 seconds in TTA. Thus, in terms of FA rate and TTA – what the operator notices - the 30/36 and 35/42 settings produce similar Tandem System performance, and are hence believed to bring the Tandem system in a desirable operating point.

The relative impact of varying the threshold on the tandem system can also be appreciated by constructing and examining the ROC[17] curve, as shown in Figure 43. A desirable ROC curve is convex and passes through the upper left corner. It is certainly undesirable for an alerter to display high FA and low CD and hence move its operating point towards the lower right corner. Ideally, one seeks to operate in the top left region, with high CD and low FA. Again, the *optimal* location is dependent on the payoff matrix values. Therefore, depending on the relative costs of FAs, CDs, MDs, and CRs, an optimal solution may not require the operating point to be exactly in the top left corner.

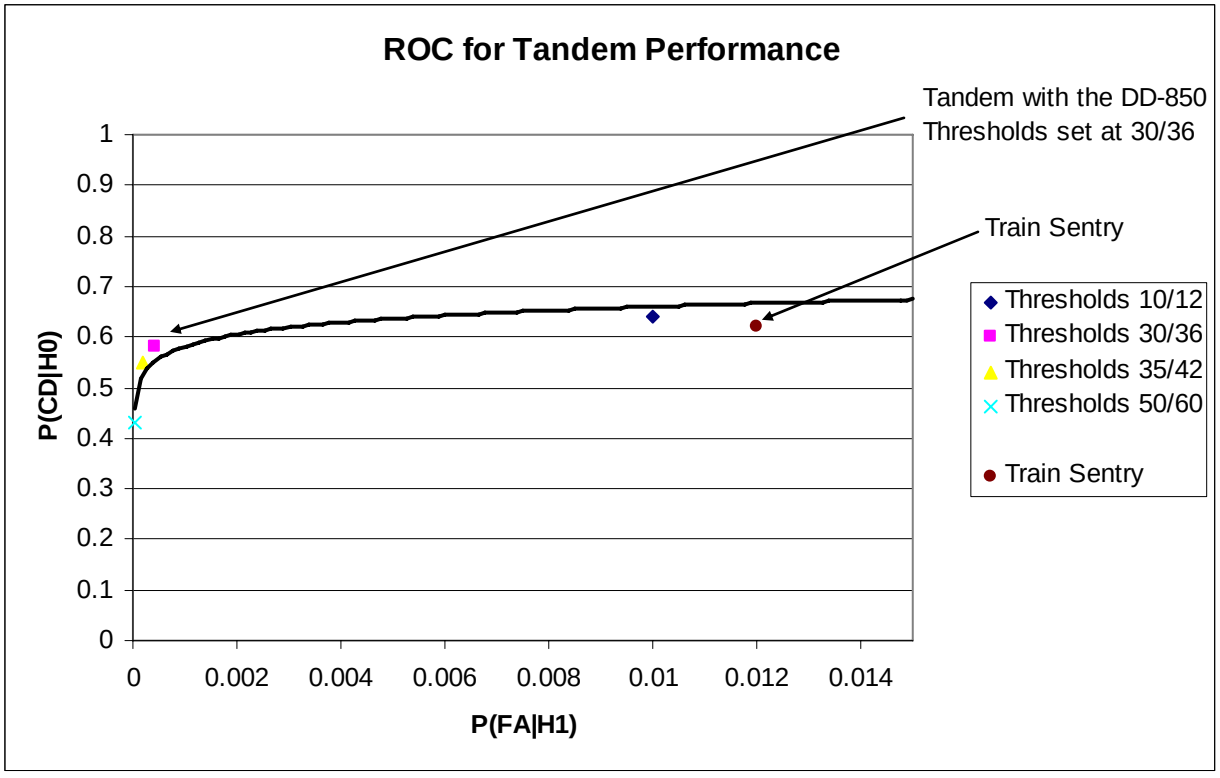

Figure 43 – ROC curve of the Tandem system constructed by varying the DD-850 threshold

Because the values of $P(FA|H_1)$ are relatively small (since $P(FA|H_1)$ is not calculated on a per event basis, but as the percentage of time during the trip an alert is issued while the operator is awake), a zoomed version of the abscissa was shown in Figure 43. Although the differences in the operating points may seem negligible – due to the length of the trip ~28,800 seconds (8

[17] ROC concepts were reviewed in Chapter 2

hours) – it is more instructive to look at the order of magnitude between the different $P(FA|H_1)$ operating points. By examining the ROC curve generated by a logarithmic fit of the 4 tandem operating points, one can note that the 30/36 threshold setting places the system's performance in the desirable top left region in ROC space. However, the 35/42 operating point has characteristics of the same order of magnitude as the 30/36 setting. Therefore, by using the 10/12 threshold Train Sentry performance as a baseline for comparison, the tandem performance has a slightly lower $P(CD|H_0)$ (~ 6-9%) but a $P(FA|H_1)$ that is 2 orders of magnitude lower (despite the fact that the noisy DD-850 has a $P(FA|H_1)$ that is 1.5-3 times greater than the Train Sentry's). Therefore, one may conclude that it is possible to bring the overall system to an operational "sweet spot", a region of satisfactory detection performance, by simply employing an 'AND strategy' tandem system consisting of the described Train Sentry class activity monitor and DD-850 infrared monitor with alerting thresholds set between 30/36 and 35/42. Formal analysis of optimality was not conducted here, but will be addressed in the final discussion.

6.3.3 Discussion:

In this illustrative example, the performance of each solution configuration was assessed and compared with respect to each other given the probability of occurrence of a sleep event. As a standalone monitor, the DD-850 has a relatively high FA rate, but a rapid response. The Train Sentry class activity monitor on the other hand has a lower FA rate and slower response. Therefore, by combining those instruments together with an AND arbiter, the resulting system eliminates the drawbacks of both instruments while capitalizing on their strengths. An adjustment to the DD-850's alerting parameters was warranted to improve the overall operation in light of the general system.

It is appropriate to discuss the rationale behind the DD-850 threshold setting in more detail. The PERCLOS value at a given instant of time is the average of the past 60 seconds or 180 samples. Therefore, a static performance analysis may not be adequate to set the alerting threshold of this instrument. The manufacturer's wish to implement a 10/12 alert may perhaps be acceptable for a person looking straight ahead (not away from the instrument's field of view). However, the difference between the manufacturer recommendation and the conclusion based on simulation

stems from the fact that the simulation analysis included the expected head and eye visibility statistics of engineers operating in the field. Successive look-aways and go-aways are expected to increase the PERCLOS value and make it hover around 15% in practice. Thus the simulation suggests that a more conservative setting is appropriate, increasing the threshold values to 30/36 to account for the engineer's behavioral characteristics, which were not yet available to the manufacturer.

With this simulation tool one may also explore related questions regarding the value and role of the alerters in the locomotive environment. For instance, could a standalone DD-850 with adjusted thresholds replace a Train Sentry monitor? As shown in Table 18 it is possible to make DD-850 performance comparable to the Train Sentry's by setting the threshold between 35/42 and 50/60. Of course, in this simulation, the manufacturer's proportion of 5:6 between the first-level and full alert was maintained to limit the degrees of freedom. However, one could independently vary the first and second level alert thresholds to obtain the required performance level. The simulation suggests that after trial-and-error a threshold that could make the DD-850's signal detection performance appear equal or even superior to the Train Sentry's. However, it should not be forgotten that the DD-850 operates only in darkness, and is less reliable when operators wear glasses. It probably makes more sense to use the DD-850 in a tandem architecture with Train Sentry. The simulation predicts the tandem system will have overall detection performance superior each of the individual standalone monitors.

A critical question is: what benefits in dollars will accrue to railroad companies who install improved alerting systems? As reviewed in Section 2.3.3, the usual SDT based systems design approach is to determine the probability of an event, and to estimate the costs and benefits of CD, MD, CR and FA. However, in the present application, there are several complications. First, as noted earlier, there is little data on the actual incidence of fatigue events in various railroad operations, and none on the probability of a major railroad accident, given that a fatigue event occurs. The required studies have not yet been made. For purposes of this thesis, we assumed that the incidence of sleep events and fatigue events were equivalent, and did not attempt to model the relationship between a sleep event and a severe accident. Second, the alerting systems modeled in this thesis have multiple alerting levels. Should the analysis be confined only to the

costs and benefits of the highest level of alerting, or are there costs of the lower levels that should also be incorporated? Third, when it comes to assigning dollar values to terms in the payoff matrix, the average costs of severe accidents depends on many variables including the type of operation, the cost of the transported freight, the extent of the environmental damage (both natural and human-made), damage to infrastructure, and legal fees. Furthermore, estimates also involve ethical issues requiring setting a dollar price on human life. The total cost of an accident may not be equivalent to the cost incurred by the company if the insurance party covers a certain fraction of the cost. Raslear and Coplen (2004) offered an SDT based analysis of the costs and benefits of a locomotive engineer fatigue prediction model for use in scheduling. By their analysis, the estimated benefit of a correct detection was set at $220,000, the cost of a missed detection was $740,000, the benefit of a correct rejection was $10,000 and the cost of a false alarm was $100,000. Raslear and Coplen are probably correct in emphasizing that the cost of a missed detection ($740,000) greatly exceeds the benefit of a correct detection ($220,000) because a missed detection greatly harms the company's reputation, while a correct detection – though it avoids lawsuits - does not significantly add value to the company. However, their estimates for correct detection and false alarms included anticipated legal costs associated with declaring an employee "unsafe" that apply to prospective fatigue prediction, but arguably not to in cab fatigue detection. The cost of an individual false alarm for an in cab fatigue detection system is more likely on the order of a few or tens of dollars. Referring back to equation 2.3 for the optimal slope of the ROC section 2.3.3, this means that the denominator terms (the dollar cost difference between the benefit of correct detection and the cost of a missed detection) dominate the numerator terms (the dollar cost difference between the benefit of correct rejection and the cost of a false alarm). In effect this means that the costs of missed detections probably should be given great weight in the design of an in-cab fatigue alerting system, and the correct detection performance should be optimized – i.e. one should operate on the right side of the system ROC curve. Though engineers may not like false alarms, they are acceptable because their dollar cost is low compared to the damage to the company's reputation caused by an accident. But the question remains: is there some upper limit on acceptable false alarm rate that is not captured by the traditional SDT cost function analysis? We argue that there is, and that it occurs when the rate is so high that engineers decide to disable the system, or their response to the alarm becomes so automatic that they can respond even while asleep. Evidence indicates that

the Train Sentry, operating on a standalone basis, has a false alarm rate that is already at an unacceptable level. The DD-850 standalone false alarm rate is anticipated to be even higher. We believe the simulation results shown in Table 17 Table 18 and Figure 44 support the view that a Tandem Alerting system will significantly lower the false alarm rate with only a few percent decrement in overall system correct detection probability. The strategy should be to reduce false alarms to a level where engineers do not consider disabling the system, or exhibit preprogrammed responses. Additional research is probably needed to establish what this false alarm level is.

The simulation described in this section assumed night operations with 1:6 probability of occurrence of a sleep event. What happens if the frequency of a sleep event was significantly smaller or higher? This question will be explored in the next two case studies.

6.4 Illustrative Application 2: Extra Board Operations

6.4.1 Simulation Assumptions

In freight extra board operations, the engineer does not and cannot know his/her schedule in advance, and may be working without sufficient sleep and out of synch with the biological clock. Though no scientific data was available, the incidence of drowsiness events in extra board operations was expected to be higher. Based on anecdotal data and Pollard (personal communication), the likelihood of a severe drowsiness event for an extra board operator operating on either side of dawn was assumed to be 1 out of 3 trips.

In the simulation, the extra board engineer was assumed to sleep at midnight, wake up at 8am and remain awake at home, but received no calls during the day. After 18 hours of wakefulness, she decided to retire and then receives a call to work. After 1 hour of commute and 1 hour in the yard, she departs at 4am for an 8-hour trip. The CT model predictions between 4am and 12pm, in Figure 44, were computed and used as the alertness level input to the simulation. The engineer was assumed to operate the train continuously and the fatigue effects of the driving task itself were not considered.

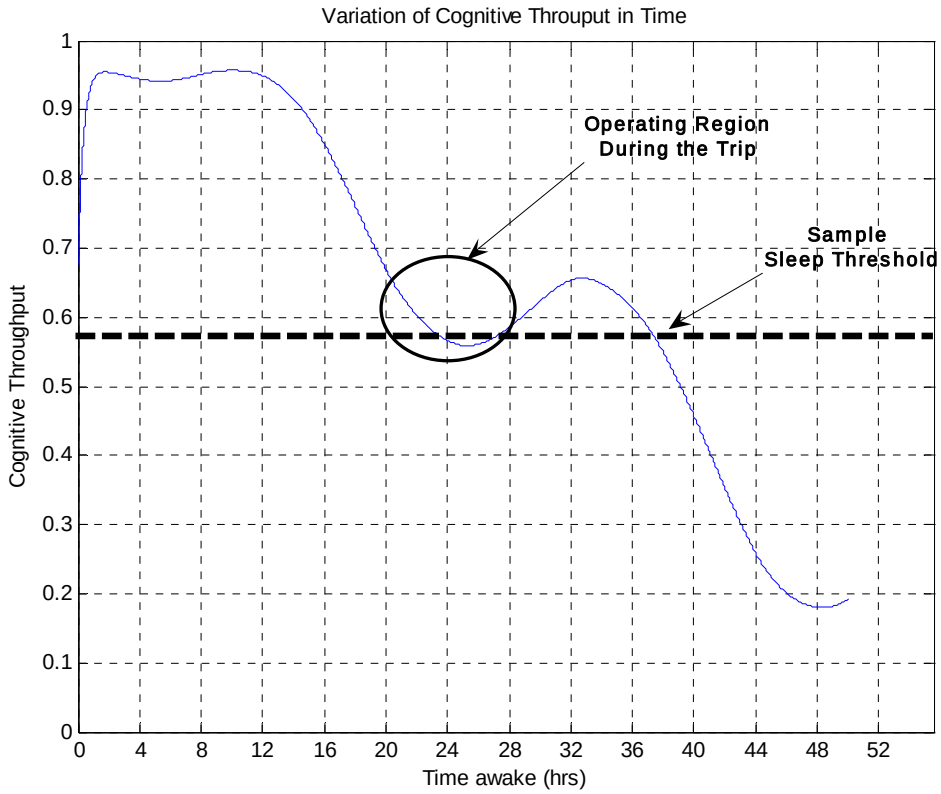

Figure 44 – Circled region comprises the extra board operator's alertness level as indicated by the Cognitive Throughput index during the trip between 4am and 12pm. A sample threshold is shown below which the operator falls asleep.

This example – and the one in the next section - also illustrates one of the limitations of the DD-850 class eye closure monitors. Half of this hypothetical trip occurs in daytime, where the DD-850 cannot operate. For these daytime applications, we assumed that the image based PERCLOS monitor was an optical system capable of operating in daytime. Since such a monitor does not yet exist, for simplicity we assumed the hypothetical daytime PERCLOS monitor had the same SDT performance characteristics as those of the DD-850 in darkness (namely, a CR rate of 89% and a CD rate of 83%).

6.4.2 Results

For this example the simulation was iterated N = 300 times to obtain about 100 trips in which the operator experiences a DE (1:3 ratio). The results were collected in Table 19.

Table 19 – Performance summary of the three different design options

Configuration	1st-Level Alerts				2nd-Level Alerts			
	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$
Train Sentry Standalone	42	29	0.012	0.62	0	42	0	0.41
PERCLOS Standalone	3054	< 1	0.86	0.99	2,329	< 1	0.65	0.99
Tandem	36	29	0.010	0.62	0	42	0	0.41

A comparison of Table 19 and Table 15 indicates that the performance of the different alerting systems does not seem to vary due to the increase (doubling) in the probability of the signal's (i.e. DE) likelihood. There was only a 4 and 5 sec discrepancy observed in the TTA at the 1st and 2nd alert levels respectively. Otherwise, the corresponding entries in both tables are of a matched order and are nearly equivalent to one another. Again the same conclusions obtained in the previous example apply to this type of operation. The tandem AND system decreases the overall FA rate to a level that is below both standalone systems despite a poor PERCLOS monitor FA performance. However, one may improve the performance of the standalone PERCLOS system, and hence of the overall system, by judiciously setting its 1st and 2nd level alert thresholds. The result of the threshold variation experiment was collected in Table 20 and Table 21.

Table 20 - Summary of the 3 configuration FA rate and TTA performance variation as a function of the PERCLOS system threshold variation

	Sentry			PERCLOS			Tandem	
PERCLOS Threshold Variation	**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**
10/12*	42	29		3,054	< 1		36	29
30/36	42	31		166	11		2	32
35/42	42	30		53	16		< 1	32
50/60	42	28		4	31		< 1	37

* n/m = 1st level alert threshold set at 'n', and 2nd level alert threshold set at 'm'

Table 21 – Summary of the 3 configuration SDT probabilities performance variation as a function of the PERCLOS threshold variation

	Sentry			PERCLOS			Tandem	
DD-850 Threshold Variation	**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**
10/12*	0.012	0.62		0.86	0.99		0.010	0.62
30/36	0.012	0.60		0.049	0.82		0.00055	0.57
35/42	0.012	0.60		0.015	0.74		0.00018	0.55
50/60	0.012	0.61		0.0012	0.48		0.000015	0.43

A careful observation of those results and a comparison to Table 17 and Table 18 reveals the same conclusions obtained in the regular freight night operations. The fusion of two imperfect alerters – a PERCLOS class monitor with a generally higher FA rate and a Train Sentry class monitor with a generally slower response (except for the 50/60 threshold setting) – results in a tandem system yielding an overall performance that is superior to that of each alerter on its own. Please refer back to section 6.3.2 for the description of the general trends in the collected data.

Basic ROC analysis principles were again applied to determine the most suitable operating point in Figure 45.

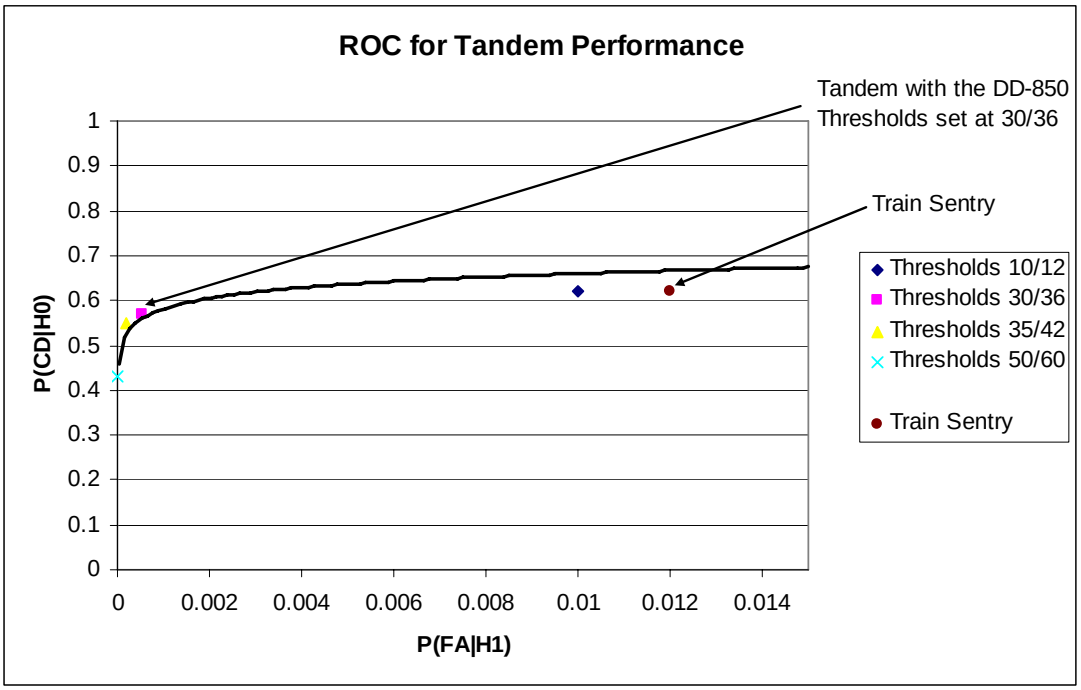

Figure 45 – ROC curve of the tandem system constructed by varying the PERCLOS monitor threshold

Indeed all SDT characteristics and trends are similar to the ones previous example (which was not unexpected given the tabulated probabilities correspondence). The 30/36 and 35/42 threshold setting place the system's performance in the desirable top left region in ROC space and are relatively equivalent within the granularity of the model.

These results strengthen the conclusion that varying the probability of DE incidence does not affect the performance of the different detector configurations. The reason why will now be discussed.

6.4.3 Discussion:

The presence of the signal (DE) does not affect the performance of the detectors. Indeed, this outcome is a direct mathematical result from SDT theory and from the method the variables are computed as explained in section 5.3.10. All calculated conditional probabilities, $P(FA|H_1)$ and $P(CD|H_0)$, are normalized and independent of the probability of the signal's presence and of the trip's length. The FA rate and the TTA, which depend respectively on the values of the two conditional probabilities (the former directly, and the latter inversely) are also unaffected by the DE's likelihood variation. However, the probability of the signal's presence enters directly into the calculation of the cost function and of $\beta_{optimal}$, as listed in equations 2.2 and 2.3. Therefore, given the relative importance accorded to the decision outcomes – i.e. the costs C_{01}, C_{10} and benefits B_{11}, B_{00} – the probability of the signal will scale the value of each decision outcome and will determine its effective contribution to the total cost function.

When data becomes available for the SDT performance of an actual "daylight capable" PERCLOS monitor, and for the probability of a serious accident given a DE, given assumed payoff matrix coefficients, it will become possible to estimate the benefits of adding a "daylight capable" PERCLOS monitor to the locomotive cab, as compared to a Train Sentry Class device, or a Train Sentry and DD-850 IR monitor working in tandem..

For completeness, even though the results were expected to be the same, the third case study examined AS performance in regular daytime operations. In this case the likelihood of an operator experiencing a DE was the smallest of all cases.

6.5 Illustrative Application 3: Regular Daytime Passenger Operations

This example also assumed a "daylight capable" PERCLOS monitor with the same mean SDT performance as the DD-850. This example shows detection performance under conditions where drowsiness events are relatively rare.

6.5.1 Simulation Assumptions

In scheduled daytime passenger operations, engineers know their regular timetables weeks and even months in advance. It thus expected that they remain fully alert throughout their work shift, with a probability of a drowsiness event being close to 0. The likelihood of 1 drowsiness event per 60 trips was arbitrarily chosen as one order of magnitude below the incidence rate of scheduled nighttime freight operations.

The engineer was assumed to go to sleep at midnight, wake up at 8am, and report to work at 9am for an eight hour shift. The CT model was run for 24 hours, and the segment between 9am and 5pm, in Figure 46, was extracted and used as the alertness level input to the simulation. No lunch or other breaks were assumed taken, and the engineer was assumed to operate the train continuously.

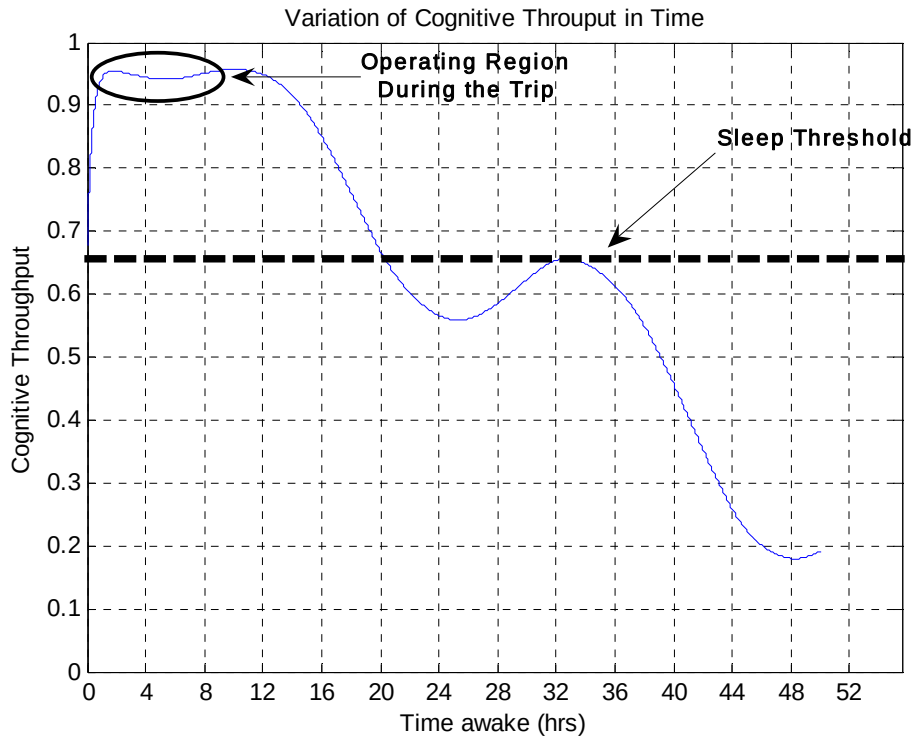

Figure 46 – Circled region indicates the operator's alertness level between 9am and 5pm. This segment is used as the simulation's alertness level input. Alertness level remains above the sample sleep threshold and the operator does not experience any drowsiness episodes in this particular trip.

For this simulation, one could have run N = 6,000 iterations (to obtain 100 trips in which a drowsiness event occurs, to attain an overall ratio of 1:60 trips). However, it would take more than 21 hours to complete one simulation. To reduce this computational time period and still obtain useful answers, the following approach was implemented. First, of the 6,000 iterations, about 100 of them had drowsiness events while the remaining 5,900 did not. Second, earlier results in Table 12 and Figure 40 suggested that running the simulation for either N = 100 or N = 500 full runs would yield the same results because of parameter value convergence. Thus, one could expect that running the simulation without any DE for N = 100 or N = 5,900 yields average values that are nearly similar, or at least of the same order of magnitude. Assuming linearity, the overall mean values, $< X >$, were computed as the following weighted average, where X may be the FA rate, TTA, $P(CD|H_0)$ or $P(FA|H_1)$

$$Overall\left\langle X_{N=6,000}\right\rangle \approx \frac{1*\left\langle X^{DE}_{N=100}\right\rangle + 59*\left\langle X^{No\ DE}_{N=100}\right\rangle}{(1+59)} \quad \ldots(6.1)$$

And where

DE = indicates the trips in which a DE occur

No DE = indicates the trips in which a DE does not occur

The simulation was run from 9am to 5pm. However, the alertness curve in Figure 46 was still rising at the beginning of the trip as the sleep inertia was still dissipating (it rises by about 0.013 in the first 50 minutes of the trip). Thus, along with a 0.0131 dip throughout the middle of the run, this initial rise in alertness caused many instances where the operator fell asleep right at the beginning of the trip, due to the manner the sleep threshold was set (section 5.3.2). This observation was not realistic. To resolve the problem for the simulation, an additional condition was added that prevented the operator from falling asleep in the first 1.5 hours of the trip.

6.5.2 Results

The results of this illustrative study were collected in Table 22.

Table 22 – Performance summary of the three different design options averaged

Configuration	1st-Level Alerts				2nd-Level Alerts			
	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$	FA rate (FAs/hr)	TTA (sec)	$P(FA\|H_1)$	$P(CD\|H_0)$
Train Sentry Standalone	42	31	0.012	0.58	0	44	0	0.38
PERCLOS Standalone	3,086	< 1	0.86	0.98	2,341	< 1	0.65	0.98
Tandem	36	31	0.010	0.58	0	44	0	0.38

Indeed, as expected, by comparing this table to Table 15 and Table 19, one obtains entries of similar order and values. Finally, the overall system performance was examined as a function of the DD-850 threshold variation and was tabulated below.

Table 23 - Summary of the 3 configuration FA rate and TTA performance variation as a function of the PERCLOS threshold variation

	Sentry			PERCLOS			Tandem	
PERCLOS Threshold Variation	**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**
10/12*	42	31		3,086	< 1		36	31
30/36	42	33		126	12		2	34
35/42	42	36		64	15		< 1	37
50/60	42	30		12	29		< 1	37

* n/m = 1st level alert threshold set at 'n', and 2nd level alert threshold set at 'm'

Table 24 – Summary of the 3 configuration SDT probabilities performance variation as a function of the DD-850 threshold variation

	Sentry			PERCLOS			Tandem	
PERCLOS Threshold Variation	**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**
10/12*	0.012	0.58		0.86	0.98		0.010	0.58
30/36	0.012	0.61		0.034	0.81		0.00046	0.57
35/42	0.012	0.60		0.018	0.74		0.00020	0.56
50/60	0.012	0.59		0.0034	0.50		0.000029	0.44

Since the same results and same conclusions apply here, a discussion was not deemed necessary.

6.6 Further Design Investigations

So far, using the simulation modeling tool, the impact of fusing alert decisions of the PERCLOS class alerter and a Train Sentry class activity monitor has been assessed. The effect on performance of varying the PERCLOS alerting thresholds were also determined both at the instrument and system level. The design tool may also be used to further improve the solution design. To do so, one may need to identify the main factors that affect the operation of the tandem configuration. For instance, the AND logic architecture considered so far requires both alerters to be simultaneously consonant and hence its time dynamics become limited by the "slowest" standalone alerter. Therefore, one may have a scenario where, at low speeds - say around 3 mph - the PERCLOS monitor is detecting a drowsy operator. At such low speeds the time window to auditory alert of the generic Train Sentry model used in this simulation is approximately 13 minutes. Thus, the tandem system will not issue the alert after tens of seconds,

but after 13 minutes. This scenario may arise when a train is slowly entering a siding to yield the way to a meeting or passing train. Such a typically boring task (especially in the dark, without any sensory stimulation) may lead to drowsiness episodes that could go undetected by the Train Sentry monitor due to its very long response time at low speeds. As a result, the train may slowly roll through the signal at the far end into the main railway and face the train it was supposed to avoid.

Thus in the low speed operating region, the generic Train Sentry model we used displays some weaknesses but the PERCLOS monitor remains unaffected. To address this concern, the tandem system could potentially be adjusted so that at low speeds it issues an alert whenever *any one* of the two alerters issues an alert. Therefore, at low speeds an OR mechanism could be implemented between the two standalone alerters whereas the same AND logic could be employed at higher speeds. Such a design approach could overcome the Train Sentry's low speed-related weaknesses and benefit from the advantages of the PERCLOS monitor.

6.6.1 Speed-Dependent Arbiter Logic

This speed-dependent arbiter architecture was implemented in the SIMULINK model as shown in Figure 47.

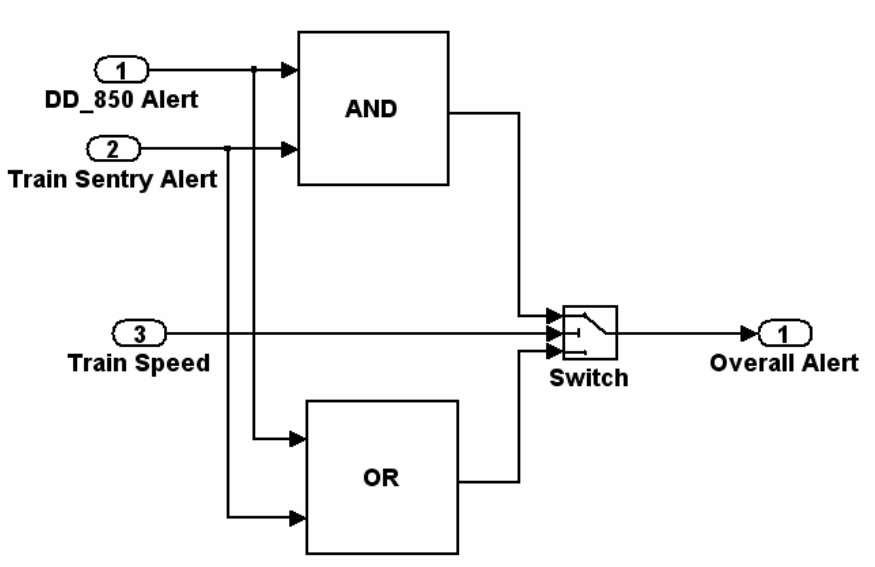

Figure 47 – Speed-dependent arbiter design.

The 'Train Speed' input acts as the variable that controls whether an AND or OR strategy is used. The arbiter 'Switch' has an internal threshold value which may be set by the user to the desired speed below which the OR logic is used. For example, if the 'Switch' threshold is set to 5 mph, then whenever 'Train Speed' is greater than 5 mph, the 'Overall Alert' is computed using the AND logic; otherwise, an OR logic is used to determine the 'Overall Alert'. Using the best tandem system so far proposed by the simulation, i.e. with the DD-850 thresholds set at 30/36, the 'Switch' threshold was varied and tandem performance was reported in Table 25 and Table 26, starting from a threshold value of 0 mph (i.e. a pure AND) to 50 mph (pure OR). Since similar results where obtained when the likelihood of a DE was 1:3 or 1:6, only the 1:3 (Night Extra Board) case was simulated. As before, a "daylight capable" DD-850-like PERCLOS monitor was assumed used. Since in this simulation the train does not operate on a continuous set of speeds, the switch threshold points were chosen according to the train's discrete speed settings listed in section 5.3.5 to cover the system's behavior in all of the train's operating conditions.

Table 25 - Summary of the 3 configuration first level FA rate and TTA performance variation as a function of the 'Switch' threshold variation

	Sentry			PERCLOS			Tandem	
Switch Threshold Variation	**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**		**FA rate (FAs/hr)**	**TTA (sec)**
0 (pureAND)	42	30		166	12		2	31
3	42	27		126	12		6	25
25	42	28		131	12		7	26
35	42	29		131	12		12	25
40	42	28		131	11		25	23
45	42	31		131	12		37	22
50 (pureOR)	42	29		125	11		166	11

* n/m = 1st level alert threshold set at 'n', and 2nd level alert threshold set at 'm'

Table 26 – Summary of the 3 configuration SDT probabilities performance variation as a function of the 'Switch' threshold variation

	Sentry			PERCLOS			Tandem	
Switch Threshold Variation	**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**		**$P(FA\|H_1)$**	**$P(CD\|H_0)$**
0 (pureAND)	0.012	0.60		0.049	0.82		0.00041	0.57
3	0.012	0.61		0.035	0.80		0.0016	0.61
25	0.012	0.61		0.036	0.80		0.0019	0.60
35	0.012	0.61		0.037	0.81		0.0032	0.62
40	0.012	0.60		0.037	0.83		0.0071	0.66
45	0.012	0.60		0.037	0.79		0.010	0.66
50 (pureOR)	0.012	0.60		0.035	0.80		0.046	0.82

Table 25 and Table 26 show detection performance across the train speed spectrum by using an arbiter that switches from an AND to an OR strategy below a certain speed threshold. As expected, varying the switch threshold does not affect the $P(CD|H_0)$, TTA, $P(FA|H_1)$ and FA rates of the standalone alerters, which remain within the same order of magnitude. However, for the tandem system, as the threshold decreases, $P(FA|H_1)$ and FA rate decrease while $P(CD|H_0)$ and TTA increase. Indeed, as the threshold increases towards 50 mph, either alert (originating from the PERCLOS or Train Sentry) will be issued by the tandem system thus reducing the speed response but increasing the incidence of false alarms. Of course, an OR strategy is viewed as the worst case configuration based on SDT principles since it alerts whenever any of the monitors issue an alert. An ROC analysis in Figure 48 may be used to qualitatively find the threshold that most reduces occurrence of FAs while displaying the highest CD performance (lowest TTA). Without the payoff matrix, one may not find the optimal threshold location.

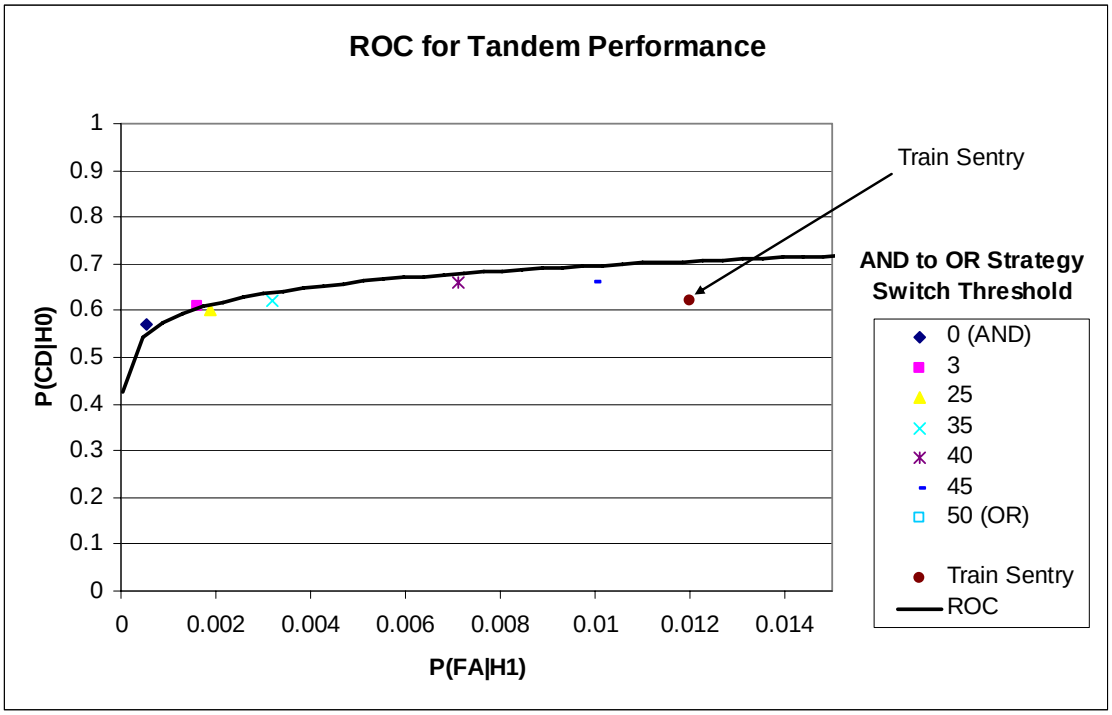

Figure 48 – ROC curve of the tandem system constructed by varying the arbiter's threshold above which an AND strategy is used while below which an OR strategy is used.

The two most effective strategies that move the tandem system to the top left operating region are the ones with the threshold set at the lowest speeds modeled: 3 mph and 25 mph. Switching between AND and OR logic at 3 mph raises the FA rate by a factor of three, but reduces TTA by 6 seconds, and raises the $P(CD|H_0)$ by four percent as compared to the AND based logic alone The Train Sentry class activity monitor takes more than one minute to issue a visual alert when train speed is below 25 mph. Therefore, this AND/OR switching solution[18] is suggested to enhance the response, at the expense of a modest increase in FAs.

[18] Disabling the DD-850 or the arbiter with a 0 mph train speed input, may be an additional modification to obtain a better alerting response so that when the train is at rest only (i.e., not at slow speeds) the operator may move around, or take a nap.

6.7 Model Limitations

The model-based calculations and results are of course not expected to be exact. They are dependent on the stated assumptions (see Chapter 5), but are expected to provide useful estimates for design purposes. As was shown, the main conclusions do not depend strongly on the probability of a sleep event, which was the major area of uncertainty in structuring the simulation. If parts of the model are elaborated in more details – for example the representation of the roadbed or the operator's activities – and are incorporated into the model, the output of the simulation is expected to become more reliable. No matter how sophisticated, the results of any simulation must ultimately be validated by field studies with humans involved. The current simulation by itself may not solve the alerting system design problem; however, it points the way to potential solutions.

6.8 Conclusion

This chapter has assessed and compared the performance of several Alerting System architectures, including a standalone DD-850 monitor (and in some cases, a "daylight capable" equivalent) and standalone Train Sentry monitors and a tandem alerter configuration, with several different arbiter architectures. The individual and system component parameters (DD-850 threshold and AND/OR strategy switch threshold) were varied to determine the best qualitative overall system SDT performance based on ROC analysis. The best solution based on these simulations appears to be the use of a Train Sentry Activity Monitor in combination with a daylight capable PERCLOS monitor arranged in a tandem configuration with the data arbiter acting as an AND gate at high speeds and as an OR gate at low speeds. It is argued that due to the nature of the in-cab alerting problem, in optimizing the system, a high cost should be placed on missed detections, due to the potential damage to the railroad company's reputation, and a relatively small penalty assigned to false alarms. Nonetheless, it is argued that the false alarm rate associated with the current Train Sentry class systems is unacceptably high, since engineers disable it and occasionally attempt to shut it off even when sleeping. The simulations suggest that by adopting a tandem alerting system design, the false alarm rate can be reduced by an order of magnitude, while maintaining good missed detection performance.

Finally, this simulation tool may be readily utilized to rapidly analyze and optimize different design options both at the individual and system level. The tool may be used as a platform to test new architectures, identify their benefits and drawbacks, and obtain as much information as possible before committing time and money to their hardware implementation. The cost of using this simulation tool is insignificant with respect to the investment involved in building deploying an actual system.

The final chapter will summarize the main contributions and findings of this thesis. It will also provide a set of recommendations for future research and development.

Chapter 7

Conclusion

7.1 Summary and Conclusions

The overall objective of this thesis was to develop an improved method to design fatigue management technologies for locomotive environments. The main contribution of this work represents the first effort of its kind where a Signal Detection Theory (SDT) engineering perspective was applied in conjunction with a stochastic framework to optimize the design of alerters in simulated operational settings. A simulation design tool was developed following steps that lie at the intersection of systems engineering, human factors, SDT, and mathematical modeling (both deterministic and stochastic). Each step is described in a chapter of its own.

In Chapter 1, the drowsiness detection problem was introduced and its importance highlighted. Millions of dollars and many lives can be saved if an operator's drowsiness level is properly monitored and alerts are judiciously issued. The main challenge of drowsiness detection stems from the absence of a single reliable physical indicator of drowsiness that can be measured without interfering with the operator's task. Many signs and symptoms collectively but not uniquely characterize drowsiness. Current alerters are inadequate partly because they typically monitor one modality and fail to consider the inherent characteristics of the locomotive operator tasks and physical environment. Furthermore, no quantitative techniques are currently being employed to assess alerting system performance. Existing devices are being typically designed and developed in an impromptu fashion without being subjected to performance criteria that yield effective and reliable alerters.

Chapter 2 reviewed the elements of a new approach. The key idea is the use of an estimation theory framework to combine information across relatively noisy detectors in tandem – each

responding to a different modality, namely, slow eyelid closure and engineer physical activity – to isolate the common drowsiness component and obtain an improved estimate of the operator's state. From a systems engineering standpoint, the problem of an alerting system deciding, in presence of uncertainty, whether an operator is awake or asleep may be approached as an SDT problem. Within such mathematical framework, an alerter's performance may be conveniently and quantitatively characterized. A cost function is defined that relates the probabilities of false alarms and correct detections, their respective costs and benefits, as well as the likelihood of a drowsiness event. Using this cost function, system components may then be optimized in the context of overall system value, rather by attempting to optimize performance at the individual component level, as is often done in practice. The problem of alerter system design addressed here is too complex to be *solely* addressed with classical SDT concepts. The detectors are nonlinear, the noise characteristics are non-Gaussian, and the monitors have time-dependent dynamics and multiple alerting levels. Therefore, Monte Carlo stochastic simulation methods were used to estimate the SDT parameters of a generic Train Sentry class activity monitor and a generic infrared eye closure monitor (based on the performance of Attention Technology's DD-850 prototype). Three types of simulations were run, separately for each standalone device, and when operating in tandem.

However, as detailed in Chapter 3, it was deemed important to first understand the interaction between the alerting system and its environment (which includes the human operator) so that the subsystems and the final system may be developed accordingly. Since no such data was available, the next logical step was to conduct a passenger and freight cab characterization study. The main objective was to quantify the important physical aspects of the locomotive cab and engineer behavior that are primarily pertinent to the performance of image-based eye closure (e.g. PERCLOS) monitors. The particular information sought related to head and eye visibility, specifically the incidence and duration of look-aways, go-aways, and intervals of visibility of one or both eyes. Physical measurements were made on 8 passenger trips in F-40, AEM7, and Acela cabs and 5 freight trips in SD-70 and AC4400 cabs. Instrumentation included an infrared head tracker, video camera, photometer, radiometer, accelerometer, and sound level meter. Engineer head position and eye/head visibility behavior were found to be quantitatively similar in both freight and passenger cabs, though cab geometries varied. While seated at the console,

the engineer's head remained within a 9.4" wide x 4.6" high x 8.4" deep box approximately 90% of the time, well within view of a 40º field-of-view console-mounted eye-closure monitor camera. However, about 6% of the time, the head is turned so far the eyes cannot be seen, and 2% of the time the engineer works elsewhere in the cab. Most of the look-aways were less than 5 sec., but go-aways were longer. Whether these generate false alarms depends on the detector system logic. Environmental characteristics were: Illumination at operator's eye level: 342 mean, peak 1250 Lux, Irradiance 0.1–30 mW/cm^2. Linear accelerations: 0.1g on all axes (95% range). Sound level range: 66dB to 95dB. It was concluded that image-based alertness monitors must be robust to wide ranges of ambient illumination, and the significant physical movement that locomotive engineers are required to make.

Chapter 4 summarized the results of a bench test that sought to evaluate the DD-850's front end image analysis SDT performance under ideal laboratory conditions with the subject looking near the camera in total darkness. Findings included a large dependence of the detector's ROC performance on head movement and eye color. The SDT parameters of the front end data stream were characterized for subsequent use in a MATLAB simulation of the entire instrument, detailed in Chapter 5. As human eye color cannot be standardized, eliminating or substantially reducing the dependence of the instrument's software detection performance on subject head movement is strongly recommended.

To assess the proposed multi-modal solution entailing the use of the PERCLOS and engineer activity detectors on both a standalone and tandem basis, and to more broadly support the design, analysis, and optimization of future alertness monitoring devices (both as individual functional components and as a part of a multi-detector system), a stochastic simulation tool was developed. Its implementation in MATLAB and SIMULINK was described in Chapter 5. Based on Chapter 3's cab study and Chapter 4's bench tests, engineering models of a generic Train Sentry class activity monitor and an infrared PERCLOS monitor with performance based on tests of the Attention Technologies Model DD-850 were developed. The SIMULINK model has several user specified inputs that drive it. An adapted version of Jewett's Cognitive Throughput Model (CTM) was used to predict the time course of the operator's alertness level. A heuristic Markov process describing variations in the roadbed configuration was used to model the operator's task

stream, while another Markov process, constructed based on eye and head visibility data obtained in the cab study, modeled variations in operator's eye visibility over time. In general, the simulation proceeds as follows. The PERCLOS model describes the presence or absence of eye closure, while the Train Sentry captures the activity input. However, the eye closure and the activity signals are both gated through an awake/asleep signal resulting from the modified CTM. Thus, when the simulated operator is asleep, eyes are closed and the operator makes no adjustments to the train controls. Finally, an AND logic based arbiter fuses the alerts from both standalone monitors to issue its own alert. The development of the overall simulation tool required making several tradeoffs between modeling reality and answering the study's proposed key questions. This simulation may be used as a proof-of-concept or pathfinder tool to rapidly assess different designs and their associated issues before actually implementing them. The model is modular and flexible, permits removal or update of current functionalities, and allows ease of inclusion of further alerters and arbiter architectures.

Finally, a simulation study using the SIMULINK model is described in Chapter 6. The model was used to examine different case scenarios and assess the performance of the alerters individually and in tandem. Three types of railroad operations were defined which were estimated to cover a 30 fold range in the probability of the operator experiencing a sleep event. These operations were regular nighttime freight, extra board freight, and regular daytime passenger. SDT probabilities of Correct Detection, False Alarms, False Alarm Rates and Total Time to Alert metrics were calculated and used to contrast the different designs. This simulation study led to the following major findings and predictions:

- **Current Performance Status:** The simulated Train Sentry class activity monitor performance remained almost invariant across the explored design space with mean false alarm (FA) rate of 42 FAs/hour and a time to alert (TTA) delay of ~ 30 seconds. This performance represents the "status quo" and the main basis for comparison.
- **Standalone Vs Tandem with AND logic:** A key result of the simulation was that a tandem system with an AND logic arbiter consistently reduces the FA rate an order of magnitude below that of both standalone detectors. This was achieved with only a loss in correct detection performance of a few percent. The overall response behavior displayed a dependency on the slowest instrument. Therefore, one should adjust both alerters to

exhibit similar time dynamics, provided the resulting system FA rate is not negatively impacted.

- **DD-850 PERCLOS Detector Alert Threshold Setting:** The DD-850 averager output alerting threshold variation at the instrument level affects the tandem operation at the system level. Increasing the threshold decreases the FA rate of the tandem system, but increases the TTA.
    - Attention Technology Inc. recommends the use of 1st and 2nd level threshold settings of 8-10% and 12% (expressed as '8-10'/12) PERCLOS. Although such settings may be appropriate if the person is looking straight at the instrument, the simulation indicates the thresholds may be unsuitable for a train operator who frequently looks away from the camera, creating an apparent artifactual eye closure.
    - A DD-850 threshold setting of 30/36 produced the best simulated tandem detection performance. Tandem performance is predicted to be 2 FAs/hour and 32 seconds in TTA. The tandem design represents an order of magnitude FA reduction with respect to status quo (i.e. Train Sentry alone) at the expense of 1 or 2 seconds of TTA.
- **Type of Operation:** Although the cost function associated with the performance of the different designs depends on the probability of the drowsiness signal's presence, the SDT rates and probabilities of the alerters were not greatly affected.
- **System Optimization Considerations:** The SIMULINK simulation allowed prediction and comparison of the ROC curves for each candidate alerting system architecture. The best solution based on simulations is the use of a Train Sentry Activity Monitor in combination with a daylight capable PERCLOS monitor arranged in a tandem configuration with the data arbiter acting as an AND gate at high speeds and as an OR gate at low speeds. The following "concept" figure illustrates the main ideas without relaying the quantitative values obtained from the simulation.

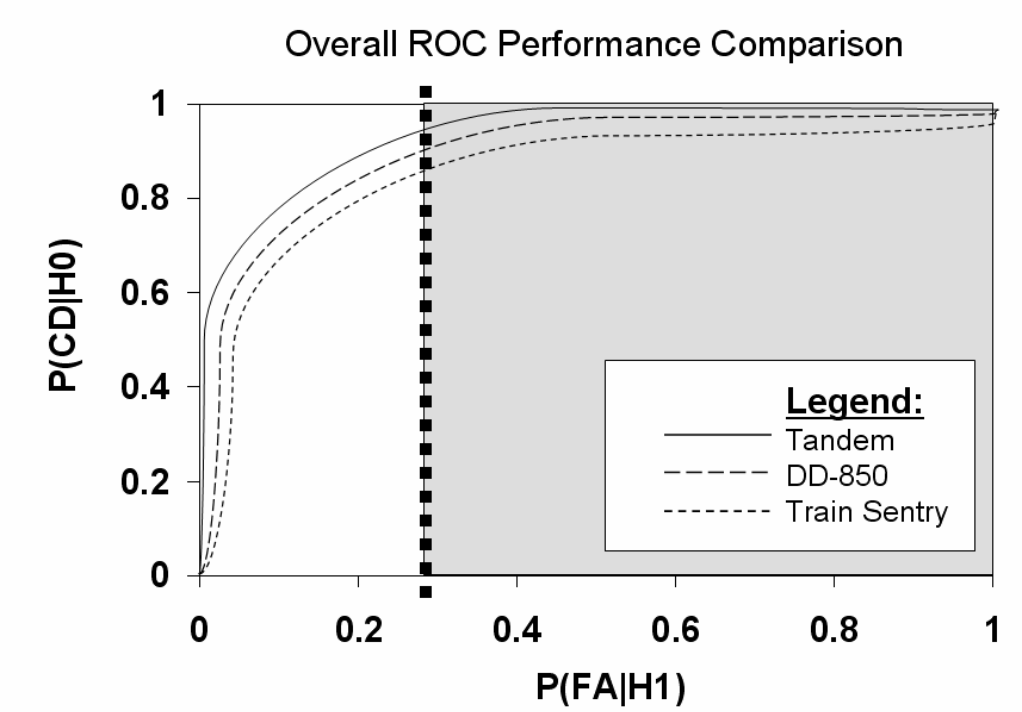

Figure 49 – Concept Figure: Compares "theoretical" ROC curves of the three different architectures.

First, where on each ROC curve the system should operate depends on the relative costs associated with missed detections and false alarms, the benefits of correct detections and correct rejections, the probability of a sleep episode for the given type of railroad operation, and the probability that a sleep episode will cause a serious accident. Statistical data on the latter two factors is not yet available, but need to be obtained or better estimated. However, it was argued that for the in-cab alerting application, the costs of a missed detection are so high (due to impact on the company's business reputation) as compared to the costs of false alarms that they dominate the design problem. The system should be designed to have the smallest missed alarm probability and the highest acceptable false alarm rate as possible. It was also argued that the current Train Sentry standalone system first level false alarm rates – on the order of 1/minute – are unacceptably high, since operators may sometimes disable the system, and others are believed to respond automatically with button pushes, even while asleep. This argument is illustrated in Figure 49 where it is suggested that there exists a certain $P(FA|H_1)$ (vertical dotted line) above which the SDT model does not apply (the grayed region to the left of the vertical dotted line). In the gray region, the payoff function becomes nonlinear, and so the linear payoff matrix equations are no longer valid.

Nonetheless, the simulations indicated that adopting a tandem alerter system architecture (the full non-dotted line) reduced the false alarm rate by an order of magnitude and improve the total time to alert, at the expense of only a few percent in missed alarm probability.

- **Monitoring PERCLOS in Daylight:** The DD-850 model PERCLOS monitor only works in darkness. However, the model was used to predict the performance of a "daylight capable" monitor with similar performance. Given data on the probability of severe accidents and payoff matrix coefficients, the model could be used to compute the economic advantages of developing a "daylight capable" PERCLOS monitor.
- **Further Optimization:** The generic train Sentry logic assumed here requires one or more engineer activities per mile of travel. However, at low speeds typical of yard operations, the total time to alert can be very long. To counter the weakness of the Train Sentry's slow response at low speeds, one may implement an arbiter architecture with an AND logic at speeds above 25 mph and an OR logic below. It was shown that the mean FA rate is 6 FAs/hour with a TTA of 25 seconds.

In conclusion, this simulation represents a useful design tool for rational design and development of alerting systems for locomotive cabs. This version of the simulation incorporates relatively detailed models for the PERCLOS and activity monitors, and the head movements made by engineers, but a relatively ad-hoc and simplified model for the roadbed and engineer activity that drives it. The simulation development was intended as a proof of concept exercise, and the simulations results focus only on orders of magnitude. Results are dependent on the stated assumptions. Improving and demonstrating the validity of the simulation is an important goal for future development. As a first effort, the results are encouraging and suggest that the model developed in this thesis can be expanded into a useful tool for the design and analysis of various alerting system architectures. It may be utilized as a platform to test and tackle new ideas, identify their benefits and drawbacks, and obtain as much relevant information as possible before committing time and money to their hardware implementation. This system-level investigation provided answers to some questions but also raised others, including the identification of areas where important experimental data is lacking. Resolving these deficiencies warrants further data

collection and research efforts to improve the validity of the simulation to ultimately use it as a formal design tool.

7.2 Recommendations for Future Development

The logical evolution of the model defines the subsequent research agenda. Recommendations are made at several levels which are ultimately linked.

The first level recommendations involve experimental data collection:

1) **Characterize the frequency of the train operator's control adjustments.** Collect video data of different engineers adjusting the train controls, in the field or in a simulator setting. Video images may be played back for a human observer to quantify and classify the control adjustment frequency. The stochastic features of the resulting control adjustments time series can be used to develop a Markov or Hidden Markov process that statistically models the operator's tasks in a Monte Carlo simulation. Such data is not currently available and is required to obtain a more reliable simulation output and Train Sentry performance evaluation. Moreover, in this simulation, the potential correlation between the engineer's head movements and tasks was not considered. Future simulations might consider inclusion of such a relation.

2) **Initiate more aggressive efforts at characterizing the statistics of the drowsiness signal, despite operational constraints.** In this work, it was assumed that every drowsiness event led to a sleep event. Thus, the simulations could have overestimated the probability of sleep (1:3, 1:6, and 1:60). Since the ROC detector performance was not a function of sleep incidence, the stated conclusions were not affected. However, recall that performance is driven by rates and not probabilities. Therefore, it is crucial to determine the frequency and duration of actual sleep episodes (in addition to drowsiness episodes) in the field or in realistic simulator studies, as well as better estimates of the probability that an actual sleep episode will cause a serious accident. One must understand in more depth the properties of the signal being simulated and detected, at least for extra board operations where the likelihood of drowsiness is more prevalent.

3) **Focus on determining precise estimates of the costs and benefits of the associated decision outcomes, i.e. the payoff matrix entries.** Cost and benefit estimate accuracy is a crucial part of SDT cost analysis and system optimization. It is the common quantitative ground used to compare different design options. Of course, one may provide recommendations based on detection and safety performance. However, safety is not the only yardstick managers use to make decisions. Financial considerations have a high impact on the choice of the final solution design.

4) **Determine what the maximum acceptable false alarm rate is for an in cab alerter.** Is an order of magnitude reduction in first level false alarm rate from 1/min to several/hour, probably achievable using a tandem alerter, sufficient to prevent operators from disabling the system or responding even while asleep?

The results of this data collection endeavor can be implemented in MATLAB/SIMULINK and added as model components.

The second level of recommendations involves potential modifications and improvements to the proposed alerting system architecture itself.

1) Develop a more detailed Train Sentry model. The existing model is generic. If a more specific study of certain operating situations requires more accurate estimates, one can use more detailed and solid technical specifications and performance data of activity based alerter models.

2) Examine the potential inclusion of a feature recognition-based (non-infrared) video PERCLOS system for daytime monitoring. Although the performance of such systems may not be perfect, an improved overall performance may be obtained with a tandem system. It may complement or replace the DD-850 in daytime hours. The detection performance of this system can be measured, modeled, and implemented in the simulation.

3) **Examine additional types of alerting strategy methods.** With an increasing number of alerters, the use of a state-space approach may be appropriate for the arbiter design due to its compact notation, straightforward implementation and possible verification of certain system properties. The state of the system may be encoded as a vector with the coordinates of its tip having representing an alerter reading. As the readings change, the vector moves in hyperspace. Alert thresholds or safe operating regions may be characterized by a multidimensional region of space, so that when the vector goes outside of this region, an alert is issued. The main disadvantage of such an approach includes the possibility of having a very large number of reachable states, thus making it difficult to fully analyze a very complex real-life system with too many tandem alerters. On the other hand, one possible approach would be to use formal state-estimation methods, such as an Extended Kalman-Filtering approach.

In conclusion, we believe that this simulation study provides enough ground to build a prototype system and test the proposed solution in a simulator (the US DOT Volpe Center simulator for example). For the future, it is suggested to keep developing this tool in parallel with iterative verification and experimental validation to use it as a formal engineering design tool.

Appendix

Please refer to the attached CD-ROM for the simulation tool code in MATLAB and SIMULINK.

References

1. Boivin D.B., Best Practices Compendium of Fatigue Countermeasures in Transport Operations. TP 13620E. 2000. Pro Tempo Inc., for Transportation Development Centre (TDC), Montreal, Que., Canada.
2. Borbély A.A., A two process model of sleep regulation. Human Neurobiol. 1982, 1: 195-204.
3. Brown I.D., (1994). Driver fatigue. Human Factors, 36, 298–314.
4. Brown E.N., Choe Y., Luithardt H., and Czeisler C.A., A statistical model of the human core-temperature circadian rhythm. Am J Physiol Endocrinol Metab, September 1, 2000; 279(3): E669 - 683.
5. Buck L., and Lamonde F. (1992). Critical Incidents and Fatigue in Canadian Railway Operation Reported by Locomotive Engineers. TP 11185E (1992) 1-G-6. Humansystems Inc. for Transportation Development Centre, Montreal, Que., Canada.
6. Chung C.A., "Simulation Modeling Handbook: A Practical Approach", CRC Press, 2004
7. Crawford A. (1961). Tiredness and visual reaction time among young male nighttime drivers: A roadside survey, Accident Analysis and Prevention, 26, 617-624.
8. Dinges D.F., Mallis M.M., Maislin G., Powell J.W., Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management, DOT HS 808 762, April, 1998.
9. Federal Railroad Administration, (2002). Five Year Strategic Plan for Railroad Research, Development, and Demonstration: [Online]. Available: http://www.fra.dot.gov/downloads/Research/rdv0202.pdf [accessed 12/16/04]. U.S. Department of Transportation, Washington D.C.
10. Flach, P.A., The many faces of ROC analysis in machine learning. Tutorial notes. University of Bristol, UK, 2004. http://www.cs.bris.ac.uk/~flach/ICML04tutorial/ROCtutorialPartI.pdf
11. Folkard S., and Åkerstedt T., Towards a model for the prediction of alertness and/or fatigue on different sleep/wake schedules. In: A. Oginski, J. Pokorski and J. Rutenfranz (eds) Contemporary Advances in Shiftwork Research 1987, pp. 231-240, Krakow, Medical Academy.
12. Gamst F. C., US Railroad history relating to fatigue, safety culture, and technology. Introductory commentary on the issues of the September 2002 TRB subcommittee meeting (in press).
13. Gamst F.C., Fatigue Counter-Measures Long Overdue, 2005. United Transportation Union. http://www.utu.org/worksite/detail_news.cfm?ArticleID=18888
14. Grace R., Byrne V.E., Legrand J.M., Gricourt D.J., Davis R.K., Staszewski J.J., Carnahan B., A Machine Vision Based Drowsy Driver Detection System for Heavy Vehicles, Proceedings of The Ocular Measures of Driver Alertness Conference, pp 75-86, April 26-27, 1999, FHWA-MC-99-136.
15. Grant, J. S. (1971). Concepts of Fatigue and Vigilance in Relation to Railway Operations. Ergonomics, 14(1): 111-118.

16. Griffin M.D., and French J.R., Space vehicle design. Reston, VA: American Institute of Aeronautics and Astronautics, Inc., 2004.
17. Hancock P.A., & Desmond P.A. (Eds.) (2001). Stress, workload and fatigue. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
18. Haworth N.L., Regan M. & Larsson T., (2000) The Effectiveness of Driver Vigilance Control Systems on Locomotives. Proceedings of the Fourth International Conference on Managing Fatigue in Transportation. Fremantle.
19. Heeger D. "Signal Detection Theory" Department of Psychology, New York University, 2003. http://www.cns.nyu.edu/~david/sdt/sdt.html
20. Howard R.A., 1971. Dynamic Probabilistic Systems, volume II. New York, NY: John Wiley & Sons.
21. Horne J.A., Reyner L.A., Sleep related vehicle accidents. Br Med J 1995; 310:565–7.
22. International Maritime Organization (IMO), "Guidance on Fatigue Mitigation and Management", Maritime Safety Committee Circ.1014, 12 June 2001. http://www.imo.org/includes/blastDataOnly.asp/data_id%3D2476/1014.pdf
23. Jewett M.E. and Kronauer R.E., Interactive Mathematical Models of Subjective Alertness and Cognitive Throughput in Humans. J Biol Rhythms 1999 14: 588-597.
24. Karnali L.A., Performance and Alertness of Railroad Engineers on Long Duration Trips. Master's Thesis, Massachusetts Institute of Technology, 2004.
25. Keewell E., and Narayanan A., Intelligent Bioinformatics.West Sussex, England: John Wiley & Sons Ltd, 2005.
26. Kithil P.W., Jones R.D., McCuish J., Driver alertness detection research using capacitive sensor array. Advanced Safety Concepts, Santa Fe, N.M./ Complexica, Inc., Santa Fe, N.M./ Mesa Analytics and Computing, Santa Fe, N.M. 2001. 3 p. Report No.: SAE 2001-01-3057.
27. Klein D.C., Moore R.Y., and Reppert S.M., Suprachiasmatic Nucleus: The Mind's Clock. New York: Oxford Univ. Press, 1991
28. Kogi K., Ohta T., Incidence of near accidental drowsing in locomotive driving during a period of rotation. J Hum Ergol 1975;4:65–76.
29. Kronauer R.E., Czeisler C.A., Pilato S.F., Moore-Ede M.C., and Weitzman E.D. (1982) Mathematical model of the human circadian system with two interacting oscillators, Amer. J. Physiol., 242: R3-R17.
30. Kuchar J., A Unified Methodology for the Evaluation of Hazard Alerting Systems, Ph.D. dissertation, Cambridge: Massachusetts Institute of Technology, 1995.
31. Mallis M.M., Deroshia C.W., Circadian rhythms, sleep, and performance in space. Aviation, Space and Environmental Medicine 2005; 76(6, Suppl.), B94-107
32. Markowitz H.M., Portfolio Selection, second edition, Blackwell (1991).
33. Moir, I., Seabridge A., Design and development of aircraft systems: an introduction. London: Professional Engineering Pub., c 2004.
34. NASA System Engineering Handbook NASA SP-610S, June 1995 http://ldcm.gsfc.nasa.gov/library/Systems_Engineering_Handbook.pdf, US Government Printing Office, Washington DC, 1995
35. National Aeronautics and Space Administration, Ames Research Centre (1995) Fatigue Resource Directory. (on line) at: http://olias.arc.nasa.gov/zteam/fredi/home-page.html
36. Oudonesom V., Evaluation of Techniques for Vigilance Measurements, Master's Thesis, Massachusetts Institute of Technology, 2001.

37. Oxford English Dictionary (Online, 2005). Oxford, England: Oxford University Press, [2000-].
38. Papoulis A., Pillai U., Probability, random variables, and stochastic processes. Boston: McGraw-Hill, c2002.
39. Philip P. Sleepiness of occupational drivers. Ind Health 2005; 43(1):30-3
40. Pollard J. K. (1991). Issues in Locomotive Crew Management and Scheduling. USDOT Report # DOT/FRA/RRP-91-06.
41. Pollard J.K. Locomotive engineer's activity diary. Washington, DC, U.S. Department of Transportation, 1996.
42. Pollard J., Popkin S., Oudonesom V., Fatigue Monitoring Devices: The Quest for the Grail Continues. Proceedings of the XVIth International Symposium on Night and Shiftwork Equity and Working Time, November 17-21, 2003, Santos, Brazil.
43. Pulse Electronics Inc., Advanced Locomotive Cab Alerter Final Report. 1993.
44. Rabiner L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proc. of IEEE, 77, 257--286.
45. Raslear T.G., Coplen M., Fatigue models as practical tools: diagnostic accuracy and decision thresholds. Aviat Space Environ Med. 2004 Mar; 75(3 Suppl):A168-72.
46. Schibler U., Ripperger J., Brown S.A., 2003. Peripheral circadian oscillators in mammals: time and food. J. Biol. Rhythms 18, 250–260.
47. Shanmugan K.S., Breipohl A.M., Random Signals: Detection Estimation and Data Analysis. New York. Wiley, 1988
48. Sherry P. (1998). Current status of Fatigue Countermeasures in the Railroad Industry. Denver, Colo.: [Online]. Available: http://www.du.edu/~psherry/fatigue/index.html [accessed 12/16/04]. University of Denver, Intermodal Transportation Institute.
49. Sussman D. & Coplen M. (2000). Fatigue and Alertness in the United States Railroad Industry Part I: The Nature of the Problem. Transportation Research Part F: Traffic Psychology and Behaviour, 3(4), 211-220.
50. Van Trees, H.L. Detection, estimation, and modulation theory. New York: Wiley, 1969.
51. Van Trees, H.L., Probability and Random Processes, 1971, video lectures, Cambridge, Mass.: Center for Advanced Engineering Study, Massachusetts Institute of Technology.
52. Vossloh Kiepe, Deadman's Switch [Online]. Available: http://www.kiepe-elektrik.com/english/bahnkomponenten/totmannschalter.htm [accessed 10/21/05] Vossloh Kiepe Electrical Systems
53. Wastney M.E., Patterson B.H., Linares O.A., Greif P.C., and Boston R.C., Investigating Biological Systems Using Modeling: Strategies and Software. San Diego: Academic Press, c1999.
54. Wickens C. D. & Hollands, J.G., Engineering psychology and human performance, Upper Saddle River, NJ: Prentice Hall, 2000.
55. Wierwille W.W., Ellsworth L.A., Wreggit S.S., Fairbanks R.J., & Kirn C.L. (1994). Research on Vehicle-Based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness. Final Report. (DOT HS 808 247). Washington, D.C.: National Highway Traffic Safety Administration.
56. Wright H.R., Lack L.C., and Kennaway D.J. (2004) Differential effects of light wavelength in phase advancing the melatonin rhythm. Journal of Pineal Research 36, 140–144.

57. Wylie C.D., Shultz T., Miller J.C., et al. Commercial motor vehicle driver fatigue and alertness study: project report, U.S. Department of Transportation. Washington, DC: DOT; 1996; Report No.: FHWA-MC-97–002.