U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

DOT HS 808 (TBD) **March 1999**

A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipients's Catalog No.

4. Title and Subtitle
A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road

5. Report Date

6. Performing Organization Code

7. Author(s)
Louis Tijerina, Mark Gleckler, Duane Stoltzfus, Scott Johnston, TRC Inc.; Michael J. Goodman, NHTSA; Walter W. Wierwille, Center for Transportation Research, VA Polytechnic Institute and State University

8. Performing Organization Report No.

9. Performing Organization Name and Address
National Highway Traffic Safety Administration
Vehicle Research and Test Center
P.O. Box 37
East Liberty, OH 43319

10. Work Unit No. (TRAIS)n code

11. Contract of Grant No.

12. Sponsoring Agency Name and Address

National Highway Traffic Safety Administration
400 Seventh Street, S.W.
Washington, DC 20590

13. Type of Report and Period Covered

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This study involved the collection of real-world driving data from a small sample of drivers, thought to be at heightened risk, to identify periods of drowsiness and inattention. Data included a variety of engineering measures including video of the driver and the road scene. One objective of the study was the identification of periods of drowsiness and inattention, documented on video, that would be made available for public education and outreach programs. A second objective was to validate, in a naturalistic driving setting, the drowsy driver detection algorithms developed by Wierwille, et al. in a simulator environment. Participants' personal vehicles were instrumented with the MicroDAS instrumentation system and all driving during the data collection was fully discretionary and independent of study objectives. The study thus offered the opportunity to implement highly unobtrusive data collection in subjects own vehicles with the absence of an experimenter in an effort to gather naturalistic data with a minimum of experimental artifacts. Results highlight the importance of lanekeeping variation as a key predictor variable for detecting drowsiness while driving, although the drowsy detection algorithm did not perform as well as in the simulator studies. An attempt to relate algorithm results to the prediction of driver inattention was inconclusive. The results are discussed in terms of theoretical, and procedural issues associated with inattention, drowsiness and driver responses to false positive epochs. It is suggested that the use of a multiplicity of approaches for addressing drowsy and inattentive driving would be most effective, and recommendations are made for future research on both technological and behavioral interventions.

17. Key Words

18. Distribution Statement
Document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No of Pages	22. Price
Unclassified	Unclassified		

Form DOT F1700.7 (8-72) **Reproduction of completed page authorized**

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF TABLES

1.0 INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) estimates that each year in the United States, approximately 1,550 people are killed and 40,000 people are injured in crashes related to, if not primarily caused by, drowsy driving (Knipling and Wang, 1994). Furthermore, it is believed that many more crashes are caused by drowsy driving than are indicated by crash reports (Wang, Knipling, and Goodman, 1996). For example, driver inattention has been implicated as a major causal factor in many different crash types (Tijerina, 1996), yet the possibility exists that drowsiness or fatigue contributes to inattention while driving.

Drowsy driving can be caused by a combination of sleep loss, driving when circadian rhythms are low (early morning hours and mid- afternoon), or driving for long periods of time. A 1994 survey conducted by the Institute for Traffic Safety and Fact Finders, Inc. found that about half of the New York State drivers polled reported driving while drowsy within the previous year and 25 percent of the drivers reported having fallen asleep at the wheel (Institute for Traffic Safety, 1998). These results have been corroborated by other surveys as well (e.g., Garder and Alexander, 1995). This troubling trend suggests a need to address the problem of drowsy driving through a variety of means. For these reasons, the Department of Transportation (DOT) is conducting basic and applied research into the problem of drowsy driving and driver inattention.

Educational and outreach programs are one means to address the problem of drowsy driving. The National Sleep Foundation's "Drive Alert...Drive Alive" program, for example, provides instructional videos, cassettes and brochures to educate the driving public on the nature of the problem and steps to counteract it. Sleep disorders that contribute to drowsy driving are also addressed in some of these types of materials. These types of programs provide useful information to the driving public about the problem of drowsy driving and are the target of public information and education programs under development at the National Highway Traffic Safety Administration (Butler, 1998). An interesting research effort might involve assessing the impact such educational and outreach programs have in reducing the incidence of drowsy driving and related crashes over time.

Road design changes are another means to mitigate or reduce the incidence of drowsy driving or inattention-related crashes. One such road design change is the installation of rumble strips along the shoulder of selected roadways to alert the driver who is drifting off the road due to sleepiness and/or inattention (Wood, 1994; Garder and Alexander, 1995). In principle, drivers take corrective action before they leave the roadway completely. The National Sleep Foundation (1997) empaneled a group of experts to review existing research on the safety impacts of continuous shoulder rumble strips. The expert panel determined that rumble strips are beneficial in reducing roadway departure crashes on rural interstates and similar roadways. However, there was insufficient data to demonstrate a benefit on other types of roads. Furthermore, there were data that suggested that shoulder rumble strips might cause crash migration, i.e., a shift in the crash location to another location without rumble strips. This was interpreted as a temporary postponement of a drowsy driver-related crash. The panel's report recommended that rumble strip deployment be accompanied with educational and outreach programs to inform drivers that hitting the rumble strips may indicate severe impairment and therefore the potentially impaired driver should pull off the road as soon as possible to rest.

Research indicates that people are not necessarily good at gauging how drowsy they are or when they are likely to fall asleep. Itoi, Cliveti, Voth, Danth, Hyde, Gupta, and Dement (1993), for example, conducted a laboratory study on sleep deprived test participants to address this question. The participants engaged in a one-hour computer exercise in which they predicted the likelihood of sleep over the next 2-minute interval by using a 0% to 100% likelihood scale. They also reported any perceived physical signs of drowsiness by selecting from icons on the computer screen that represented involuntary eye closure, yawning, involuntary head nodding, and the like. The results indicated widely varying abilities of individuals to predict sleep onset, with people predicting, on average, a likelihood of only 55% that they would fall asleep prior to their first sleep onset. Self reports of physical indicators of drowsiness were, on average, higher prior to sleep onset than prior to epochs where sleep did not occur. It was reported that test participants whose likelihood estimates seemed to ignore the frequency of physical indicators tended to be poor predictors. Also, those test participants whose physical indicators of drowsiness did not necessarily provide a strong indication of sleep onset also tended to be poor predictors. Together, these results were interpreted to mean that people's inability to judge sleep onset may be related either to a lack of meaningful physical warning signs in some persons or a failure to appreciate the significance of these indicators in other persons.

Technology may be able to provide earlier or more robust warnings of degraded driving status so as to preclude a drowsy driving-related crash. Drowsy driver detection algorithms and approaches have been a topic of considerable research in recent years. A key ingredient in the development of such algorithms is selection of an appropriate "criterion" measure for drowsiness. Such a measure of drowsiness should ideally be valid (i.e., relate to observed performance decrements in meaningful ways) and reliable (i.e., consistently vary with levels of sleepiness, circadian troughs, or time on task). Numerous physiological measures of driver arousal or wakefulness exist (Wylie, Shultz, Miller, Mitler, and Mackie, 1996). These include electroencephalograms (EEGs), heart rate and heart rate variability, core body temperature, and various measures of the eyes.

Eyelid closure has recently proven to be among the most robust and meaningful operational indicators of drowsiness while driving. Wierwille, Wreggitt, Kirn, Ellsworth, and Fairbanks, 1994) demonstrated that PERCLOS, defined as the proportion of a time interval that the eyes were 80% to 100% closed (exclusive of blinks), was highly correlated with other physiological indicators of drowsiness and was a useful criterion for drowsiness prediction algorithms that employed driver-vehicle performance measures as predictors. PERCLOS has an operational relevance that is hard to dispute. Wierwille et al. (1994) found that it was indicative of sleep onset and was connected to poor performance in visual tasks. In driving, the authors point out, "...it seems obvious that if a driver's eyelids are closed, the ability to operate a vehicle would be greatly hampered." (Wierwille et al., 1994, pg. 13).

Numerous attempts have been made to correlate various measures of driving to drowsiness. Variations in lanekeeping, steering inputs, and speed maintenance have been reported as a function of level of fatigue (Wierwille et al., 1994). Of the various measures, those related to steering inputs and lanekeeping appear to be the most promising indicators of drowsiness or impaired driving performance. When the eyes are closed due to drowsiness, visual inputs to the driver are temporarily halted. The driver may hold the steering wheel at a nearly fixed angle during eye closure and the vehicle continues in a straight line. However, given road crown, road surface variations, wind gusts, and variations in road geometry, the vehicle may drift out of lane. The driver who successfully completes a trip will, when subjective uncertainty about the driving situation builds up, open the eyelids to assess the driving situation. This, in turn, will be associated with a corrective steering input and more or less sudden shift in lane position back to lane center. This process is consistent with models of the driver as a intermittent controller (Senders, Kristofferson, Levison, Dietrich, and Ward, 1967). Speed and accelerator pedal variations have been reported by some researchers (e.g., Artaud, Planque, Lavergne, Cara, de Lepine, Tarriere, and Geuguen, 1994; Siegmund, King, and Mumford, 1996) but these are more variable in their relation to drowsiness levels than the steering and lanekeeping effects just described.

The National Highway Traffic Safety Administration (NHTSA) has sponsored research into the drowsiness detection problem. Of particular interest in this report is the drowsy driver research program recently completed by Wierwille and his associates (Wierwille, Lewin, and Fairbanks, 1996a, 1996b, 1996c). This research focused on the development of a vehicle-based driver drowsiness detection system. This is a system of continuous, unobtrusive measurements of driving performance (e.g., steering wheel inputs, lanekeeping performance) and an algorithm to classify a driver as "drowsy" or "not drowsy". Such a detection system would eventually be integrated into a driver-system interface to present warning signals to the driver and possibly countermeasures to drowsiness as well (e.g., cool air, mint scent, seat shaker).

During the course of research into drowsy driver detection algorithms, Wierwille and his associates conducted extensive studies in the driving simulator located at the Vehicle Analysis and Simulation Laboratory at the Virginia Polytechnic Institute and State University. The general research paradigm was as follows. Test participants were asked to rise by 7:00am on the testing day and carry on with normal activities but avoid naps. At 6:00pm they were picked up by an associate experimenter, taken to dinner, and asked to avoid stimulants during and after dinner. The test participant was brought to the laboratory after dinner and engaged in various activities but avoided naps. At approximately midnight, the test participant entered the simulator and began a simulator driving session that lasted roughly 2.75 hours. During this time, various driver-vehicle performance measures were captured, and video of the driver's face was recorded. Drowsiness was operationally defined by PERCLOS, the percentage of a sample driving period (generally 3 to 6 minutes) where the eyes were 80 to 100 percent closed. (Other measures of drowsiness were developed, but they will not be discussed here). Various statistical analyses were conducted, the most robust of which proved to be multiple linear regression. This method of detection algorithm development allowed Wierwille and his associates to set thresholds after the model was fitted to the data.

Algorithms to estimate PERCLOS by means of the regression models had multiple R values ranging from approximately 0.79 to 0.87, depending on what predictors were used in the algorithm. Late in the program, however, these early algorithms were found to be ineffective in detecting drowsiness (Wierwille et al., 1996b). It was surmised that earlier simulator studies had not sufficiently emphasized good lanekeeping as would be expected in normal driving. During the subsequent test and evaluation studies, however, the simulator study paradigm was modified to encourage better lanekeeping by means of monetary penalties for lane departures, a "laneminder" device that sounded an auditory alarm when lane boundaries were exceeded, and a seat vibration "rumble" cue that was presented when lane excursions were excessive.

New studies on the driving simulator were conducted with the new protocol that emphasized good lanekeeping. The best algorithm developed to estimate PERCLOS, referred to as algorithm F4e-3, had a multiple-R value of $R = 0.48$, with a threshold of 0.012 (i.e. a classification criterion of an estimated total of (0.012)x(180 seconds) or approximately 2 seconds or more in a 3-minute driving period or epoch with eyelids 80 percent or more closed, disregarding eye blinks) . The classification accuracy for drowsy vs. non-drowsy states was a false alarm rate of 0.038 or 3.8% and a miss rate of 0.51 or 51.0% (Wierwille et al., 1996a, pg. 57). To improve classification accuracy, Wierwille and his associates then logically OR'd the estimated PERCLOS (ePERCLOS) algorithm with LANEX. LANEX is defined as the proportion of a driving interval the vehicle exceeded a lane line (threshold

0.06667 or 12 seconds for a 3-minute driving interval or epoch). This approach lead to a reconsideration of the phenomenon of interest from drowsy driving detection alone to drowsy driving or impaired lanekeeping performance. By this redefinition, the false alarm rate was reduced to 0.002 or 0.2 % and the miss rate was reduced to 0.211 or 21.1%. Wierwille and his colleagues (Wierwille et al., 1996c) concluded that constructing conditional OR classification schemes by using both estimated PERCLOS algorithm output and observed LANEX was necessary to significantly improve classification accuracy levels. Lane exceedences are taken as prima facie safety relevant to road departure, lane change, and other types of crashes. Theoretically, intended lane changes or other lane departure maneuvers could be distinguished by the driver's use of directional signals.

LNMNSQ values are highly correlated with lane departures so are considered safety relevant as well. LNMNSQ is defined as the lane position mean square error about lane center within a 3-minute driving epoch. A threshold and criterion of 3.00 feet^2 was applied. Estimated PERCLOS was logically OR'd with LNMNSQ to detect degraded driving performance or drowsiness. This lead to a false alarm rate of 0.005 or 0.5 % and a miss rate of 0.1973 or 19.73 %. The research program concluded at this point with the recommended PERCLOS algorithm F4e-3 and a recommended driver interface. It was beyond the scope of that effort to validate the algorithms on the road or test the efficacy of driver warnings or countermeasures in a real world setting.

There were two objectives in the present study. One objective was to collect real-world driving data from a small sample of drivers, including face video, video of the road scene, and various engineering measures. These data would then be processed to determine periods of drowsy and inattentive driving and selected segments would then be available for use as exhibits in NHTSA-sponsored educational and outreach programs. A second objective of this effort was to use real world naturalistic driving data to examine the Wierwille et al. (1996) drowsy driver detection algorithm's performance in a real world driving environment as compared to the simulator environment in which it was originally developed. The test participants were chosen as representatives of drivers thought to be at heightened risk for drowsy driving. They included shift workers, military personnel on leave, and students traveling during breaks in classes. The test participants were not informed of specific interests of the study in an effort to avoid any reactive behavior that might bias the results away from naturalistic driving. However, they were aware of the instrumentation in their vehicles and that a variety of measures were being collected.

2.0 METHOD

2.1 Test Participants

College students were recruited by placing display advertisements in the school newspaper and by posting flyers in public places on campus. Shift workers were recruited by pre-shift announcements by management, in plant TV announcements, and posting flyers in public places at the plant. Military personnel were recruited by placing display advertisements in a base newspaper and posting flyers in public places on base. All recruitment took place in central Ohio.

The advertisements specified the test participant must be from the population of interest and be planning a drive of at least 5 hours duration each way (or a commute of at least 45 minutes for the shift workers). The ads also specified that the volunteer be healthy, have a good driving record, and drive a passenger car. The ads explained that the purpose of the study was to collect data on long distance drives or commutes, how the data would be collected, how much they would be compensated for participation, and the time frame in which test participants were needed.

When prospective test participants called NHTSA's Vehicle Research and Test Center (VRTC), the program was explained in more detail. The program was only described as a study of long distance driving; drowsy driving was not discussed. If the prospective test participants were still interested, eligibility and trip information would be taken. Then this information was reviewed and participants were selected.

The participants were selected from the available pool based on many criteria. Foremost, they had to meet the below listed requirements:

a. have a valid drivers license without restrictions. If eyeglasses or contacts are required by the test participant for driving, they must be worn for testing.

b. have a driving record with no DUI convictions, no more than one crash in the last five years, no more than two moving violations in the last three years

c. have at least the automobile insurance coverage required by state law

d. have a minimum of two years of driving experience;

e. be at least 21 years of age and not more than 55 years of age.

f. not have any diagnosed sleep-related disorder or physical handicaps;

g. not be under the influence of any emotional stress that may impair driving ability.

h drive a late model car in good mechanical condition.

I. be willing to sign a waiver to allow video and engineering data taken of the driver while driving to be used without constraint by NHTSA for educational outreach and research purposes.

From the reduced pool of those who met the above requirements, test participants who were most likely to experience drowsy driving because of their planned itinerary were selected. Once selected, a driving abstract was obtained from the Ohio Bureau of Motor Vehicles to verify the information taken over the telephone, and participants would be called back and asked to participate. If they agreed to participate, their vehicles would be scheduled for pick-up. Table 1 provides descriptive information on the sample of test participants analyzed in this study. They were paid between $310.00 and $390.00, depending on the length of time their vehicle was “leased” to VRTC for instrumentation work and when the data collection was carried out.

TABLE 1. Test Participant Description, Vehicle Driven, and Trip Description

Participant Code	Description	Vehicle Driven	Trip Descriptions
1	Male, 51 years of age, shift worker	1989 Ford Taurus	Worked from 2:45 p.m. to 11:15 p.m. His drive to work varied from 50 to 60 minutes, predominantly on a rural divided highway
2	Male, 30 years of age, college student	1988 VW Sirocco	Drove to Washington D.C. from Columbus, Ohio and back. This trip normally takes about 7 hours, one way.
3	Male, 32 years of age, college student	1992 Honda Accord	Drove to New Jersey from Columbus, Ohio and back. This trip normally takes about 8 hours. Individual stated he made this trip almost every weekend.
5	Male, 53 years of age, shift worker	1987 Honda Accord	Worked from 10:30 p.m. to 7:00 a.m. His drive to work varied from 45 to 50 minutes and took place on rural highways and a rural interstate highway.
6	Male, 34 years of age, college student	1990 Honda Civic	Drove to Coventry, Conn. from Columbus, Ohio and back. This trip normally takes about 11 hours.
7	Male, 28 years of age, college student	1994 Honda Civic	Drove to Richmond, Virginia from Columbus, Ohio and back. This trip normally takes about 7 hours.
8	Male, 35 years of age, military personnel	1995 Dodge Intrepid	Drove to East Saint Louis, Illinois from Dayton, Ohio and back. This trip normally takes about 7 hours.

Note: Test Participant 4 was not included because of data loss. Two additional test participants (Test Participants 9 and 10) were also not included because of data loss.

2.2 Apparatus

Each test participant's personal vehicle was equipped with MicroDAS instrumentation. The system of sensors and processing captured steering wheel position, travel speed, lane position, lane exceedences, lateral and longitudinal accelerations, and yaw rate. In addition, digitized video of the road scene and driver eye glance behavior were also recorded and synchronized to the engineering data . Infrared lighting was used so as to allow high quality video of the driver's face to be taken at night. All data were captured at a sampling rate of 30 Hz. Further details on MicroDAS are provided in Barickman (1998). The MicroDAS was programmed to collect data only when the vehicle was traveling approximately 49 mph or faster. The MicroDAS continued recording until the vehicle travel speed fell below 43 mph. Such minimum travel speed criteria were suggested by Wierwille et al. (1996a) as indicative of a condition relatively more likely to induce drowsy driving and for which a usable algorithm for drowsiness detection might be obtained. Eye glance video was later manually reduced to obtain eye closure data needed to determine the observed measure of drowsiness, PERCLOS. It should be noted that the MicroDAS lane trackers are essentially cameras whose images are processed to discern the line pavement marking from the surrounding pavement. Worn lane lines, specular reflections, and other phenomena contributed to loss of lane tracker data.

2.3 Procedure

At the time a test participant's vehicle was picked up for MicroDAS installation, the individual was asked to read and sign an informed consent and waiver of confidentiality form. This form described the program in general terms, the risks, benefits, confidentiality, compensation, and requirements. The individual was also asked to sign a vehicle lease agreement, which gave the owner's permission for VRTC to take the vehicle and install MicroDAS. The return of the vehicle was scheduled and delivered upon completion of MicroDAS installation. Depending on the individual test participant's driving plans, the pickup of the vehicle for MicroDAS removal was also scheduled. The test participant then completed the planned driving, during which data was collected. When the vehicle was returned at the end of the test, after removal of the MicroDAS, the test participants were paid and informally interviewed. They were asked about any problems or observations about MicroDAS or any problems or unusual events that occurred during their driving. Additionally, the shift workers were interviewed about their sleep habits.

Care was taken to ensure that test participants were not alerted to a special interest in drowsy driving. The informed consent form, for example, described the study purpose only in general terms, i.e., to gather data on how people normally drive in a variety of driving conditions and evaluate the MicroDAS instrumentation (both true). No attempts were made to collect sleep logs, subjective sleep scale ratings, or any other data that might prompt a reaction on the part of the test participant. Also, the driving undertaken by the test participants was fully discretionary and had been planned regardless of involvement in this study. These factors, together with the absence of an experimenter on the drives, the unobtrusive nature of MicroDAS, and its installation into test participant's own vehicles, represent a comprehensive effort to gather naturalistic driving data with a minimum of experimental artifacts.

2.4 Measured Variables

A list of the measured variables is provided in Table 2. These variables include those used to assess the Wierwille et al. (1996a) drowsy driver detection algorithm. Additional variables in the table were also collected so that new algorithms might be developed from the collected data.

TABLE 2. Response Variables Collected/Calculated for Drowsy/Inattentive Driving Study
(Source: Wierwille, Lewin, and Fairbanks, 1996a; Wierwille, personal communication, 1998)

Measured Variable	Definition	Collection/Calculation
PERCLOS	The proportion of time that a test participant's eyes were 80% to 100% closed.	Manual data reduction eye closure data from face camera video. Magnitude and duration of eye closure used to calculate PERCLOS
STVELV	The variance of steering wheel velocity, where steering velocity is measured in degrees/sec. Thus, STVELV is in units of (degrees/sec)^2.	Steering wheel position was recorded at approximately 0.1 degree resolution. This signal was post-processed (filtered)and then differentiated to derive STVELV, applying standard sample estimates of variance. The steering velocity noise was within 2 degrees / sec.

TABLE 2. (Continued)

LGREV	The number of times that the steering wheel movement exceeded 15 degrees after steering velocity passed through zero.	LGREV and MDREV were post-processed from measured steering wheel position and calculated steering wheel velocity with a 2 deg/sec deadband (i.e., operational definition of zero velocity).
MDREV	The number of times that the steering wheel movement exceeded 5 degrees, but did not exceed 15 degrees, after steering velocity passed through zero.	See LGREV comments.
LNMNSQ	The mean square of lane position with respect to lane center, measured in $feet^2$.	Calculated from two rear-mounted, downward-facing cameras that captured approximately a 7-foot field of view each. Video output was processed through a VRTC-built video processing board to determine lane line position with respect to vehicle reference points. Resolution of the system averaged approximately 1inch.
LANVAR	The variance of lateral position relative to lane center, in $feet^2$.	See previous comments. Note that LANVAR is equal to LNMNSQ only when the mean lane position is zero (i.e., lane center).
LANEX	The proportion of time any part of the vehicle is outside the lane boundary .	Derived from lane position sensor information.
INTAC DEV INTAC DEV*	The standard deviation of the lateral velocity of the vehicle, measured in volts where one volt equals 73.34 feet/sec. INTACDEV* is INTACDEV in feet/second. It was assessed with stepwise regression (see text).	Lateral acceleration calculated from yaw rate (in degrees/second) and travel speed (in feet/second). This is then integrated and filtered via a "leaking integrator function" applied to derive lateral velocity, which is then converted to volts.
HPYAW RTDEV and HPYAW VAR	The standard deviation of yaw rate after high pass filtering, in units of degrees/sec. HPYAWRTVAR is the square of HPYAWRTDEV.	Yaw rate sensor output is high pass filtered in post-processing with a corner frequency of 0.08Hz. The purpose of high pass filtering is to remove the effects of curvature on the yaw rate signal. The result is a measure that reflects changes in control of vehicle alignment

3.0 RESULTS

The analysis of the reduced video and engineering data was conducted in several stages. First, quality assurance criteria were established such that a) only full 3-minute epochs were analyzed, b) eye closure data had to be available for 2.5 minutes or more of each analyzed 3-minute epoch, and c) engineering data (specifically lane position data) had to be available for 2.5 minutes or more for each analyzed 3-minute epoch. Video data loss might have arisen due to equipment malfunction en route (e.g., a system crash or face camera misalignment) or insufficient data in the face video (e.g., due to shadows or variations in driver head position) to allow the manual data reductionist to reliably measure eye closure. Engineering data such as the lane position channel might be unavailable due to such factors as worn lane line pavement markings, shadows on the lane markings, or specular reflections off the pavement from bright sunlight. These quality assurance criteria were based on the fact that 3-minute epochs were used in the final recommended estimated PERCLOS (ePERCLOS) algorithm, algorithm F4e-3 in Wierwille et al. (1996a). Furthermore, when lane information is not available for a substantial period of time, no reasonable degree of detection accuracy can be achieved with steering inputs alone (Wierwille et al., 199c, pg. 40). In all, 283 3-minute epochs were retained for further analysis.

A total of 20 epochs of drowsiness occurred among the 283 epochs analyzed. These were operationally defined by means of an observed PERCLOS value greater than or equal to 0.012. Table 3 presents these 20 epochs. By looking down the SUBJECT column, one may see that test participants 1, 3, 5, 6, 7,and 8 all had at least one drowsy driving epoch each. However, test participants 3 and 8 account for the majority of the drowsy driving epochs observed. Within a file (FILENAME), the sequence of drowsy driving epochs (EPOKSEQ) were sometimes consecutive, indicating drowsy driving that extended beyond a single 3-minute period. For example, test participant 3 had consecutive epoch sequence numbers 20 and 21 for file 3070 and test participant 8 had consecutive epoch sequence numbers 14 and 15 for file 8007. The longest single bout of drowsy driving that contributed to these PERCLOS values was approximately 13 seconds for test participant 3, epoch sequence 26 in file 3070. There were no drowsiness-related crashes or serious mishaps recorded during the study. One close call was recorded; a heavy truck ahead of one test participant blew a tire and its tread flew toward the test participant's windshield. Luckily, the test participant made a successful evasive maneuver (lane change) and avoided the tread.

TABLE 3. True Drowsiness Epochs

True Drowsiness Epochs

SUBJECT	FILENAME	EPOKSEQ	PERCLOS	LNMNSQ	LANEX
1	1122	13	0.016	0.94155	0.00000
3	3070	19	0.032	3.07173	0.06972
3	3070	20	0.017	4.44423	0.15518
3	3070	21	0.024	3.58716	0.07776
3	3070	26	0.041	4.36884	0.09091
3	3072	8	0.018	3.03197	0.03113
3	3072	11	0.044	5.00633	0.10175
3	3072	18	0.019	1.78454	0.00000
3	3109	2	0.031	4.45339	0.06549
3	3109	3	0.057	2.61978	0.01973
3	3109	5	0.057	1.27367	0.01745
3	3110	7	0.031	1.49403	0.00563
5	5007	1	0.050	1.60610	0.00000
6	6011	14	0.014	0.69562	0.00000
7	7005	21	0.012	1.57067	0.04486
8	8007	14	0.016	1.30125	0.00683
8	8007	15	0.016	0.72378	0.00631
8	8007	28	0.013	0.52367	0.00000
8	8023	38	0.014	1.50358	0.00000
8	8024	15	0.014	2.11609	0.00000

Note: Drowsiness operationally defined by an observed PERCLOS value greater than or equal to 0.012.

Inasmuch as detection of degraded driving was also of interest in this project, degraded driving as defined in terms of lanekeeping measures were also examined. Table 4 presents the instances of degraded lanekeeping, given no drowsy driving occurred. The top portion of the table shows epochs with degraded lanekeeping, operationally defined with a lane mean square lane position value for the epoch of 3.0 ft^2 or greater. The lower portion shows epochs with observed lane exceedence (LANEX) values of 0.06667 or greater for a 3-minute epoch. Interestingly, all of these degraded lanekeeping events occurred with test participant 3. These were the data to which driver status monitoring assessments were directed.

TABLE 4. True Degraded Driving Epochs, Exclusive of Drowsy Driving

True Degraded Driving Epochs,LNMNSQ Criterion, Excluding Drowsy Driving

OBS	SUBJECT	FILENAME	EPOKSEQ	PC80PCT	LNMNSQ
1	2	2018	14	.000	3.79630
2	2	2021	1	.000	3.08035
3	3	3070	3	.000	5.18908
4	3	3070	10	.000	3.41268
5	3	3070	17	.000	3.57719
6	3	3072	4	.000	5.21237
7	3	3072	5	.000	5.18546
8	3	3072	6	.007	4.29919
9	3	3072	12	.000	3.66169
10	3	3072	15	.000	3.81875
11	3	3072	16	.006	4.25305
12	3	3086	5	.000	3.05906
13	3	3101	5	.000	3.12766
14	3	3102	3	.000	7.86231
15	3	3102	5	.000	3.64000

True Degraded Driving Epochs, Lanex Defined, excluding Drowsy Driving

OBS	SUBJECT	FILENAME	EPOKSEQ	PC80PCT	LANEX
1	3	3070	3	0	0.11290
2	3	3070	10	0	0.07202
3	3	3072	4	0	0.11723
4	3	3072	5	0	0.09600
5	3	3072	15	0	0.08989
6	3	3102	3	0	0.21120

Note: The LNMNSQ criterion was an observed lane position mean square of 3.0 ft^2 or more. The LANEX criterion was an observed lane exceedence of 0.06667 or more.

3.1 Drowsiness Detection Results

Derived by means of backward stepwise multiple linear regression, Wierwille et al. (1996a) recommended Algorithm F4e-3 to generate estimates of PERCLOS (ePERCLOS):

$$\text{ePERCLOS} = -0.00304 + 0.000055\,(\text{STVELV}) - 0.00153\,(\text{LGREV}) - 0.00038(\text{MDREV}) + 0.003326(\text{LNMNSQ}) + 0.00524(\text{LANVAR}) - 0.00796(\text{INTACDEV})$$

where all terms are defined in Table 2. This model had an multiple correlation coefficient of R = 0.482 in the report by Wierwille et al. (1996a) and all regression coefficients were statistically significantly different from zero at least a 0.05 alpha level. The algorithm was intended for use with predictor values calculated over 3-minute epochs. It should be noted that multiple regression models, in maximizing multiple correlation, take advantage of any correlated errors or specific variations in the original derivation study (Grimm and Yarnold, 1995). Therefore, the measures of association and accuracy of prediction (or classification) are expected to be lower or shrink somewhat when the regression equation is applied to a new sample of data. How much shrinkage arises depends on the degree of similarity between the original model building data set and the new sample.

Algorithm F4e-3 was applied to the naturalistic driving data set in order to assess its classification performance. Figure 1 depicts a scatterplot of observed PERCLOS, versus ePERCLOS as calculated from algorithm F4e-3. There is a small but significant positive correlation between PERCLOS and ePERCLOS, $r = 0.274$, $p < 0.0001$, $N = 283$. Table 5 presents the drowsiness classification accuracy of algorithm F4e-3's ePERCLOS with the selected naturalistic driving data set. The classification assumed an ePERCLOS threshold and a PERCLOS criterion for drowsy driving of 0.012. This amounts to cumulative eye closure (80% or more eyelid closure) of approximately 2 seconds or more over a 180 second epoch, disregarding blinks. As indicated in Table 5, when drowsy driving was not present (as operationally defined by observed PERCLOS less than 0.012), the classification accuracy was approximately 94%, close to the 96% reported by Wierwille et al. (1996a, 1996c) with the same threshold and criterion values. When drowsiness was present, the correct detection or true positive rate was 35% as compared with 48.5% in Wierwille et al. (1996a, 1996c). This suggests the ePERCLOS algorithm extrapolated relatively well in terms of its intended purpose from the simulator to real world driving.

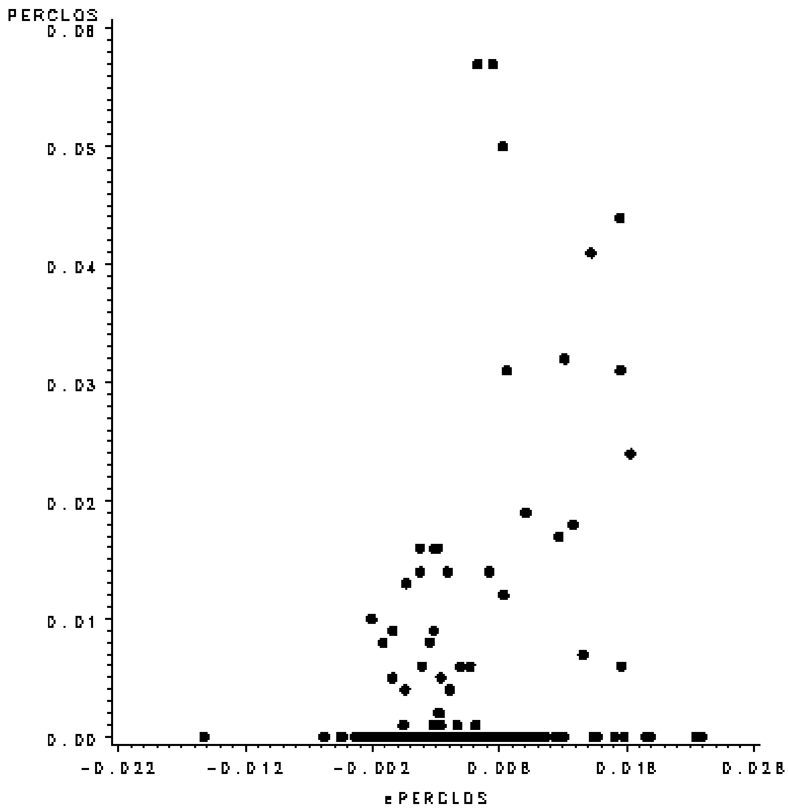

Figure 1. Scatter Plot of observed PERCLOS versus estimated PERCLOS (ePERCLOS) calculated from algorithm F4e-3, N =283 3-minute epochs from 7 test participants.

TABLE 5. Drowsiness Classification Accuracy of Algorithm F4e-3 calculated ePERCLOS

		Observed PERCLOS		
		Low	High	Total
ePERCLOS, Algorithm F4e-3 Output	High (i.e. Drowsy)	16	7	23
	Low (i.e. Not Drowsy)	247	13	260
	Total	263	20	283
	% Correct	93.9%	35.0%	Overall: 89.7%

Note: ePERCLOS threshold = 0.012; PERCLOS criterion = 0.012.

An effort was made to refit the same set of predictors to the naturalistic data by means of multiple linear regression using the SAS PROC GLM routine (SAS Institute, 1990). The intention was to assume that the set of predictors in algorithm F4e-3 were a “best” subset of predictors and simply adjust the regression weights to improve the fit to the current data set. This effort led to substantially different results in the multiple regression, as indicated in Table 6. Unlike the results reported in Wierwille et al. (1996a) no predictor except for LNMNSQ had a regression weight significantly different from zero. Furthermore, this model led to no substantially larger multiple R than the correlation between PERCLOS and the original algorithm. For this reason, algorithm F4e-3 was retained for further assessment.

TABLE 6. Multiple regression results to refit Algorithm F4e-3 to naturalistic driving data collected in this study.

```
Dependent Variable:PERCLOS
                                      Analysis of Variance

                                         Sum of          Mean
                Source          DF      Squares        Square      F Value       Prob>F

                Model            6      0.00151       0.00025        4.208       0.0005
                Error          276      0.01651       0.00006
                C Total        282      0.01802

                    Root MSE       0.00773     R-square       0.0838 Multiple R = 0.289
                    Dep Mean       0.00227     Adj R-sq       0.0639
                    C.V.         340.90941

                                      Parameter Estimates

                              Parameter       Standard    T for H0:
              Variable  DF     Estimate          Error   Parameter=0    Prob > |T|

              INTERCEPT  1     -0.000928     0.00157240        -0.590        0.5555
              STVELV     1  -0.000024132     0.00026202        -0.092        0.9267
              LGREV      1  -0.000031506     0.00148892        -0.021        0.9831
              MDREV      1   0.000004708     0.00022534         0.021        0.9833
              LNMNSQ     1      0.001780     0.00058252         3.055        0.0025
              LANVAR     1      0.001718     0.00141590         1.213        0.2261
              INTACDEV   1     -0.010286     0.02075054        -0.496        0.6205
```

Wierwille et al. (1996a) also recommended that algorithm F4e-3 be logically OR’d with observed LANEX to enhance classification performance. The criterion applied to LANEX was a measured value of 0.06667 or greater for a 3-minute epoch. This translates into traveling with any part of the vehicle exceeding a lane boundary for approximately 0.06667 x (180 seconds) or approximately 12 seconds or more, cumulatively, within a 3-minute period. Table 7 presents the classification results of this approach applied to drowsiness detection. As can be seen in comparison to Table 5, there is no change in the classification accuracy over algorithm F4e-3 alone. Examination of the false positive cases revealed they were the same in both instances.

TABLE 7. Drowsiness Classification Accuracy of Algorithm F4e-3 OR'd with observed LANEX

		Observed PERCLOS		
		Low	High	Total
ePERCLOS, Algorithm F4e-3 OR Observed LANEX Output	High (i.e. Drowsy)	16	7	23
	Low (i.e. Not Drowsy)	247	13	260
	Total	263	20	283
	% Correct	93.9%	35.0%	Overall: 89.7%

Note: ePERCLOS threshold = 0.012; LANEX threshold = 0.06667; PERCLOS criterion = 0.012.

Wierwille et al. (1996a) also applied algorithm F4e-3 logically OR'd with LNMNSQ. Table 8 shows that, relative to algorithm F4e-3 alone, this results in a slight increase in false positives with no increase in true positive detections. Such results indicate this approach is not recommended if the goal is to maximize the number of correct classifications.

TABLE 8. Drowsiness Classification Accuracy of Algorithm F4e-3 OR'd with observed LNMNSQ

		Observed PERCLOS		
		Low	High	Total
ePERCLOS, Algorithm F4e-3 OR LNMNSQ Output	High (i.e. Drowsy)	19	7	26
	Low (i.e. Not Drowsy)	244	13	257
	Total	263	20	283
	% Correct	92.8%	35.0%	Overall: 88.7%

Note: ePERCLOS threshold = 0.012; LNMNSQ threshold = 3.00; PERCLOS criterion = 0.012.

In an attempt to be as parsimonious as possible, a classification was made using observed LANEX alone for classifying drowsiness (as defined by observed PERCLOS). That is, an observed LANEX of 0.06667 or greater for a 3-minute epoch to predict drowsiness. Table 9 shows that this simple rule reduces false alarms to approximately 2.3 % but it does so at the expense of greater numbers of missed detections of drowsiness (approximately 75%).

TABLE 9. Drowsiness Classification Accuracy with observed LANEX alone

		Observed PERCLOS		
		Low	High	Total
Observed LANEX Alone Output	High (i.e. Drowsy)	6	5	11
	Low (i.e. Not Drowsy)	257	15	272
	Total	263	20	283
	% Correct	97.7%	25.0%	Overall: 92.6%

Note: LANEX threshold = 0.06667; PERCLOS criterion = 0.012.

A simple classification scheme based on observed LNMNSQ alone was also attempted. For this scheme, a drowsy classification was estimated if the observed LNMNSQ value for a 3-minute epoch was 3.00 ft^2 or greater. Table 10 indicates that this rule is as good as Algorithm F4e-3 in terms of hit rate (35%) and even better in terms of false alarm rate (5.7% as opposed to 6.1% with Algorithm F4e-3).

TABLE 10. Drowsiness Classification Accuracy of observed LNMNSQ alone

		Observed PERCLOS		
		Low	High	Total
Observed LNMNSQ alone Output	High (i.e. Drowsy)	15	7	22
	Low (i.e. Not Drowsy)	248	13	261
	Total	263	20	283
	% Correct	94.3%	35.0%	Overall: 90.1%

Note: LNMNSQ threshold = 3.00; PERCLOS criterion = 0.012.

A final analysis was carried out in an attempt to arrive at a better classification algorithm based on selection of predictor variables from all of the measures in Table 2. Stepwise, forward selection, and backward elimination methods were applied to the data set using the SAS® PROC GLM procedure (SAS Institute, 1992). The results of all three selection procedures were the same. The solution (Table 11) for the "best" estimated PERCLOS value (i.e, BESTePERC) is:

$$\text{BESTePERC} = -0.00011822 + 0.00225082 * \text{LANEX} + 0.07517621 * \text{LANVAR}$$

TABLE 11. Best-Fit Model obtained with Stepwise, Forward, and Backward Regression

Multiple R = 0.3063268 R-square = 0.09383618 adj. R-square = 0.0873894

	DF	Sum of Squares	Mean Square	F	Prob>F
Regression	2	0.00169070	0.00084535	14.50	0.0001
Error	280	0.01632689	0.00005831		
Total	282	0.01801759			

Variable	Parameter Estimate	Standard Error	Type II Sum of Squares	F	Prob>F
INTERCEP	-0.00011822	0.00092225	0.00000096	0.02	0.8981
LANVAR	0.00225082	0.00114123	0.00022682	3.89	0.0496
LANEX	0.07517621	0.02081958	0.00076026	13.04	0.0004

Figure 2 presents a scatter plot of observed PERCLOS versus BESTePERC . In comparison to Figure 1, again one sees a small positive correlation between PERCLOS and its estimate (multiple $R = 0.306$, $p < 0.0001$, $N = 283$).. Like Figure 1, it is also plain that the epochs indicative of drowsiness (i.e., those with y-axis values of 0.012 or greater) are spread over a fairly wide range of BESTePERC values. This suggests that the distributions of drowsy and non-drowsy epochs overlap in terms of BESTePERC, much like they did with algorithm F4e-3's ePERCLOS.

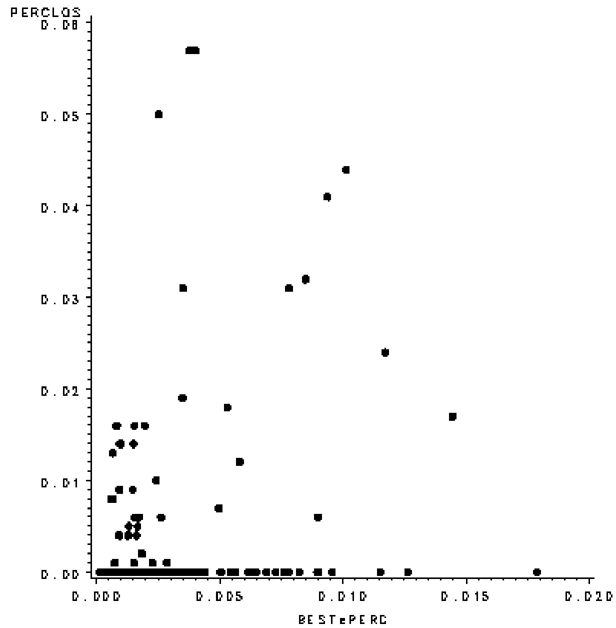

Figure 2. Scatter Plot of PERCLOS versus BESTePERC.

Table 12 shows the classification accuracy of BESTePERC with a threshold of 0.012. With this threshold, the number of false alarms drops to 2 epochs but the number of true positives also drops. The threshold was varied in an attempt to improve the classification accuracy but no threshold value could be found that surpassed the classification performance of LNMNSQ alone.

TABLE 12. Drowsiness Classification Accuracy with "Best" Derived Algorithm

Observed PERCLOS

		Low	High	Total
BESTePERC	High (i.e. Drowsy)	2	1	3
"best fit" model Output	Low (i.e. Not Drowsy)	261	19	280
	Total	263	20	283
	% Correct	99.2 %	5.0 %	Overall:92.6 %

Note: BESTePERC threshold = 0.012; PERCLOS criterion = 0.012.

Furthermore, since the entire data set was used to obtain the BESTePERC results, no validation sample was set aside for validation. It is expected that using only half of the sample would result in poorer fit of the BESTePERC model. Finally, since the simple approach of using observed LNMNSQ alone did not require any estimation at all, (i.e., observed LNMNSQ was used to determine if it was above or below 3.00 ft^2), no further analysis of BESTePERC was carried out.

Table 13 shows the false alarm cases generated with each of the detection approaches. Inspection of the contents of this table reveal several interesting and important results. First, 17 out of 19 epochs belong to the same driver, Test Participant 3. This suggests that Test Participant 3 was in some sense idiosyncratic relative to the other drivers. Second, all of the false alarm epochs share in common high lanekeeping variability or lengthy lane exceedences (or both). These are often, but not always, associated with higher ePERCLOS values. This suggests that degraded lanekeeping is the key, not only to drowsiness detection, but also to false alarms. A further consideration of degraded lanekeeping is offered in the next section.

TABLE 13. False Alarm Cases for Drowsiness Detection

OBS	SUBJECT	FILENAME	EPOCH	PERCLOS	LNMNSQ	ePERCLOS (Algo.F4e-3)	LANEX
1	2	2018	14	.000	3.79630	0.019518	0.02724
2	2	2021	1	.000	3.08035	0.004807	0.00000
3	3	3070	3	.000	5.18908	0.023937	0.11290
4	3	3070	4	.000	2.86242	0.015490	0.04997
5	3	3070	10	.000	3.41268	0.015437	0.07202
6	3	3070	17	.000	3.57719	0.010005	0.04631
7	3	3070	18	.000	2.64038	0.013098	0.02558
8	3	3072	4	.000	5.21237	0.019496	0.11723
9	3	3072	5	.000	5.18546	0.017852	0.09600
10	3	3072	6	.007	4.29919	0.014622	0.04798
11	3	3072	12	.000	3.66169	0.012716	0.04473
12	3	3072	15	.000	3.81875	0.012443	0.08989
13	3	3072	16	.006	4.25305	0.017679	0.06457
14	3	3086	4	.000	2.27819	0.012799	0.00515
15	3	3086	5	.000	3.05906	0.019946	0.02740
16	3	3101	5	.000	3.12766	0.010446	0.00000
17	3	3102	3	.000	7.86231	0.023498	0.21120
18	3	3102	5	.000	3.64000	0.015692	0.04713
19	3	3110	6	.000	2.52050	0.017141	0.00915

To recapitulate, the following results were obtained with the naturalistic driving data. Algorithm F4e-3 provided a substantial degree of classification accuracy (relative to the upper limit represented by the simulator work of Wierwille et al., 1996a; 1996c). However, observed LNMNSQ with a criterion of 3.00 ft^2 provided the best classification performance of any other approach tried. It appears that lanekeeping variation is a key predictor variable for detecting drowsiness while driving.

3.2 Drowsiness or Degraded Performance Detection Results

It was mentioned earlier that Wierwille et al. (1996c) attempted to improve classification accuracy by methods that logically OR'd algorithm F4e-3 with LANEX or with LNMNSQ. No comparisons with their classification accuracy were provided in the previous section for a subtle reason. In the previous section, the phenomenon of interest was drowsiness detection only. Wierwille et al. (1996a), on the other hand, assessed drowsiness or degraded driving performance (as evidenced by degraded lanekeeping). Their line of reasoning was that lane exceedences are prima facie evidence of driving performance that can result in crashes such as roadway departures. Thus, they characterized them as indicative of degraded driving performance with probability 1.0. The present authors attempted a similar analysis.

Algorithm F4e-3 was logically OR'd with LANEX to arrive at a "drowsy driver or degraded performance" detection approach. The phenomenon to be detected was redefined and the 283 selected epochs were reclassified as low versus high drowsiness or degraded performance driving periods. Both the criterion and threshold applied for LANEX was a measured value of 0.06667 or greater for a 3-minute epoch. This translates into traveling with any part of the vehicle exceeding a lane boundary for approximately 0.06667 x (180 seconds) or approximately 12 seconds or more, cumulatively, within a 3-minute period. By reclassifying any such lane exceedences as, de facto, degraded performance, this has the effect of moving data points from the drowsiness algorithm classification matrices from classification as missed detections, false alarms, or correct rejections to classification as correct detections if the LANEX data for a given epoch was above threshold (Wierwille et al., 1996c, pg. 35).

Table 14 presents the classification results of algorithm F4e-3 OR'd with LANEX. As indicated, the reclassification to identify drowsiness or degraded lanekeeping performance lead to an improvement in classification: 96.1% correct classifications in the absence of degraded driver status, 50% correct detections (true positives) when degraded driver status was present. Inspection of the true positive and false positive epochs revealed that the incorporation of LANEX and operational redefinition of degraded driver status lead to a migration of false positive cases to the true positive category. No other migration was observed. By way of comparison, Wierwille et al. (1996a) reported classification performance of 0.25% false positives and 78.87% correct detections.

TABLE 14. Classification Accuracy of Algorithm F4e-3 calculated ePERCLOS OR observed LANEX for Drowsiness or Degraded Performance Detection (LANEX-defined)

		Observed PERCLOS or Degraded LANEX-Defined Lanekeeping		
		Low	High	Total
ePERCLOS OR LANEX Algorithm Output	High (i.e., degraded status)	10	13	23
	Low(i.e., no degraded status)	247	13	260
	Total	257	26	283
	% Correct	96.1%	50.0%	Overall: 91.9%

Note: ePERCLOS threshold = 0.012; LANEX threshold and criterion = 0.06667; PERCLOS criterion = 0.012.

A similar analysis was conducted by use of algorithm F4e-3 OR'd with LNMNSQ. The criterion and threshold for degraded lanekeeping performance was set to LNMNSQ = 3.00 ft^2. Again, by reclassification of the phenomenon to be detected, application of this revised algorithm improved classification accuracy. As indicated in Table 15, the false positive rate fell to 1.6 percent of 3-minute epochs while the true positive rate increased to 62.8%. This is lower, though similar to the 0.5% false positive and 80.28% true positive rates reported in Wierwille et al. (1996a; 1996c) from the simulator research.

TABLE 15. Classification Accuracy of Algorithm F4e-3 calculated ePERCLOS OR observed LNMNSQ for Drowsiness or Degraded Performance Detection

		Observed PERCLOS or Degraded LNMNSQ-Defined Lanekeeping		
		Low	High	Total
ePERCLOS OR LNMNSQ Algorithm Output	High (i.e., degraded status)	4	22	26
	Low(i.e., no degraded status)	244	13	257
	Total	248	35	283
	% Correct	98.4%	62.8%	Overall: 94.0%

Note: ePERCLOS threshold = 0.012; LNMNSQ threshold and criterion = 3.00 ft^2; PERCLOS criterion = 0.012.

Next, an attempt was made to find a simpler prediction scheme by use of observed LANEX alone. A criterion and threshold value of 0.06667 for LANEX was applied to recategorize the epochs and reclassify them. The results are presented in Table 16. Note that there are no false positives but the true positive rate is only 42.3%, less than that reported for the previous two classification approaches.

TABLE 16. Classification Accuracy of LANEX Alone for Drowsiness or Degraded Performance Detection (LANEX-defined)

		Observed PERCLOS or Degraded LANEX-Defined Lanekeeping		
		Low	High	Total
Measured LANEX Criterion Alone as Predictor	High (i.e., degraded status)	0	11	11
	Low(i.e., no degraded status)	257	15	272
	Total	257	26	283
	% Correct	100%	42.3%	Overall: 94.7%

Note: LANEX threshold and criterion = 0.06667; PERCLOS criterion = 0.012.

A similar approach was carried out using observed LNMNSQ to redefine driver status as drowsiness or degraded performance (as indicated by LNMNSQ greater than or equal to 3.00 ft ^{2} within a 3-minute epoch). Table 17 presents the results of this classification. In this instance, the false alarm rate again fell to 0.0% and the hit rate grew to 62.8%. The suppression of false alarms surpasses even the best results reported by Wierwille et al. (1996c) for algorithm F4e-3 OR'd with LNMNSQ. However, the hit rate is below the 80.28% correct detections reported by Wierwille et al. (1996a).

TABLE 17. Classification Accuracy of Observed LNMNSQ alone for Drowsiness or degraded Performance (LNMNSQ-defined)

	Observed PERCLOS or Degraded LNMNSQ-Defined Lanekeeping			
		Low	High	Total
Measured LNMNSQ Criterion Alone as Predictor	High (i.e., drowsy)	0	22	22
	Low(i.e., not drowsy)	248	13	261
	Total	248	35	283
	% Correct	100%	62.8%	Overall: 95.4%

Note: LNMNSQ threshold and criterion = 3.00 ft^{2} ; PERCLOS criterion = 0.012.

3.3 Further Analysis of False Positive Epochs in Terms of Curve Cutting

One possibility to explain the higher lanekeeping variability present in the False Positive epochs for drowsiness detection alone is that it relates to a voluntary strategy on the part of the driver. For example, drivers sometimes negotiate curves by crossing the lane line in the curve, a strategy referred to as "curve cutting" (Koenig, Pape, and Pomerleau, 1995). This might be done to reduce the perceived centrifugal forces acting upon the driver and vehicle while going through the curve. If this was the driver's strategy during false positive epochs, a warning by the driver status monitoring system might be perceived as a nuisance rather than as an indicator of degraded performance.

To evaluate this hypothesis, the yaw rate channel was low-passed filtered at a nominal 0.08 Hz cutoff frequency. The intent was to provide a rough surrogate measure of road "curviness" during an epoch by filtering out the higher-frequency yaw rate variability that might be attributable to the driver's inputs rather than the demand of the road geometry. This was done for 17 false positive epochs from Test Participant 3 and 20 true negative epochs from Test Participant 3. The results are graphically depicted as an empirical quantile plot in Figure 3 in terms of low-pass filtered yaw rate variance (LPYAWRTVAR). A quantile is the proportion of the observations that fall at or below the scale value of the data (Chambers, Cleveland, Kleiner, and Tukey, 1983). Thus, Q(0.85) defines that value of LPYAWRTVAR at or below which a fraction 0.85 of the observations fall. The difference between a percentile and a quantile is that the percentile refers to a percentage of the data and the quantile refers to a fraction of the data. Note that in the Figure the False Positive observations are plotted as "F" and the True Negative observations are plotted as "T". If there was a substantially higher low-pass yaw rate variance associated with the False Positives (indicative of increased road curviness), the curve of "F" points would be shifted to the right and separated from the curve of "T" points. There is some evidence of this in the plot but only beyond about 0.5 $(deg/s)^2$. Further, the separation between the False Positive and True Negative points does not appear to be substantial. Thus, the low-pass filtered yaw rate variance data are suggestive of increased road curviness during False Positive epochs, but they are not definitive.

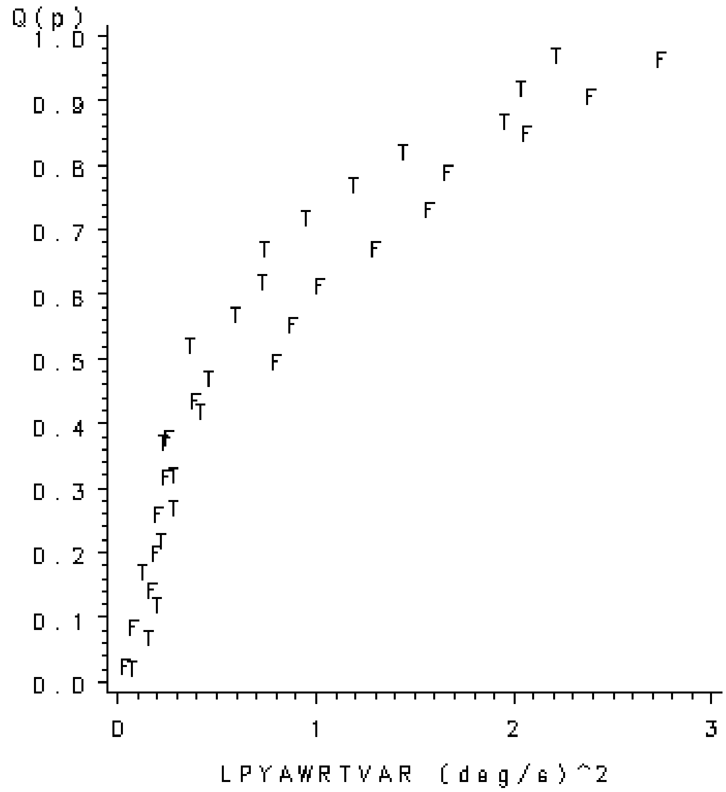

Figure 3. Empirical Quantile Plot of Low-Pass Filtered Yaw Rate Variance (LPYAWRTVAR), by False Positive (F) epochs and True Negative (T) epochs, for Test Participant 3

3.4 Further Analysis of False Alarm Epochs in Terms of Driver Eye Glance Behavior

Vision is the primary information channel a driver uses to drive safely. It is for this reason that measures of driver eye glance behavior have been widely used in highway safety research (Wierwille, 1993). An intriguing hypothesis about the false alarms epochs is that at least some portion of them might be prompted by substantial periods of eyes-off-road ahead time. This might arise as a result of a variety of in-vehicle or roadside distraction sources. Some of the in-vehicle distractions from the road scene ahead observed in the data set include looking at a road atlas, changing a compact disc, drinking beverages, and eating food or snacks. If during a distraction, lanekeeping was to be degraded, a driver status monitoring system might nonetheless provide value to the driver in alerting him or her to that fact.

To examine this hypothesis, selected epochs associated with Test Participant 3 were reanalyzed, this time in terms of driver eye glance behavior. The video of the driver's face was manually reviewed and glance locations and durations were manually reduced from the video for the false positive and true negative epochs. Then a new variable was created as an analogue to PERCLOS. This variable was the Eyes Off Road-ahead Time Proportion (EORTP). This was defined as the proportion of a 3-minute epoch that the driver's eyes were diverted from the road ahead. Figure 4 depicts the empirical quantile plot for the EORTP measure, by False Positive ("F") and True Negative ("T") epochs, for Test Participant 3. Contrary to expectation, the True Negative epochs actually often had longer EORT proportions and the range was wider. Thus, this data set does not support the hypothesis that the driver status monitoring schemes evaluated here are sensitive to or indicative of significant eye glance behavior away from the road scene ahead.

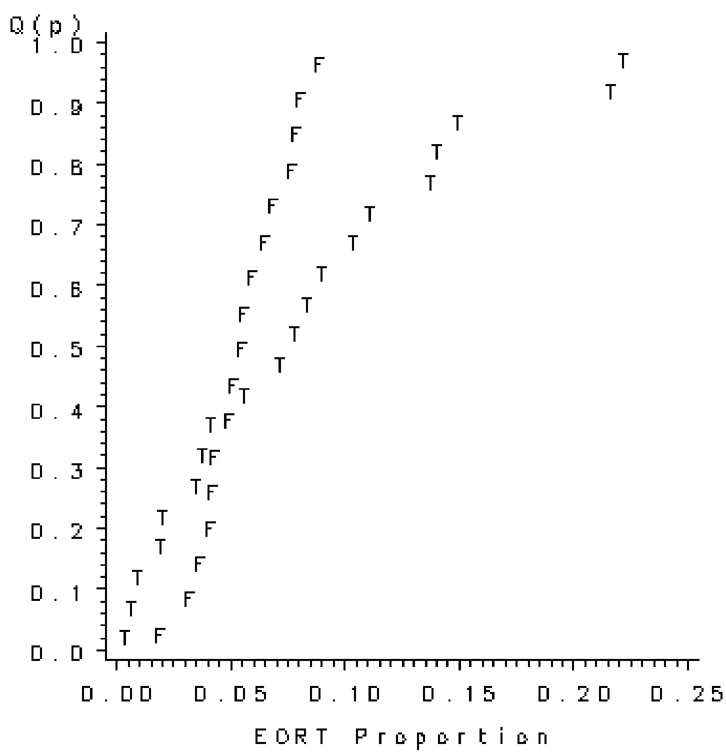

Figure 4. Empirical Quantile plot of Eyes-Off-Road-Time proportion, by False Positive (F) epochs and True Negative (T) epochs, for Test Participant 3

3.5 A Qualitative Analysis of False Alarm Epochs as Preliminary to Drowsiness Epochs

Table 18 presents the false positive and true positive epochs for Test Participant 3, sequenced in time. In this table, the filename data are rank-ordered (i.e., filename 3070 was created before filename 3072) and epoch numbers within a filename are sequentially ordered by time (e.g., epoch 17 was the 3-minute period prior to epoch 18). A close review of the sequence of those false positive epochs along with the drowsiness epochs (both true positives and false negatives) generated by this test participant reveals that they were often interleaved. (Missing epochs in a sequence were removed from analysis because they did not meet the quality assurance filters described earlier). This suggests that the erratic driving that triggered a false positive indication may have been an attempt by the driver to fight off the effects of drowsiness. If so, these false positives were in themselves drowsiness-related. An alternative interpretation is that this interleaving represents some form of cyclical effect of drowsiness wherein driving performance is degraded while eye closure is unaffected, followed by periods when both performance is degraded and eye closure increases.

TABLE 18. False Positive (FP) and Positive (P) Epochs of Drowsiness, arranged in temporal sequence, Test Participant 3

Code[a]	SUBJECT	FILENAME	EPOKSEQ
FP	3	3070	3
FP	3	3070	4
FP	3	3070	10
FP	3	3070	17
FP	3	3070	18
P	3	3070	19
P	3	3070	20
P	3	3070	21
P	3	3070	26
FP	3	3072	4
FP	3	3072	5
FP	3	3072	6
FP	3	3072	12
FP	3	3072	15
FP	3	3072	16
P	3	3072	8
P	3	3072	11
P	3	3072	18
FP	3	3086	4
FP	3	3086	5
FP	3	3101	5
FP	3	3102	3
FP	3	3102	5
P	3	3109	2
P	3	3109	3
P	3	3109	5
FP	3	3110	6
P	3	3110	7

Note: a. P indicates a Positive epoch (either True Positive or False Negative). FP indicates a "False Positive" epoch

4.0 DISCUSSION

This research effort represents a first attempt to validate the drowsy driver detection algorithm of Wierwille et al. (1996a) using real world driving data. This study is unique for several reasons. Test participants were passenger car drivers rather than commercial vehicle drivers who are more often the subject of drowsy driving research (e.g., Wylie et al., 1996). Drivers' personal vehicles were instrumented for unobtrusive data collection in a naturalistic setting rather than being provided with an unfamiliar test vehicle. Drivers were not informed of the specific interest in drowsy driving nor were any procedures used that might prompt a sensitivity to this issue (e.g., sleep questionnaire, sleep logs, etc.). Drivers made trips of their own choosing rather than those prescribed by the researchers. And drivers drove on real roads, rather than in a simulator or a test track.

The underlying logic of using deterioration in driving performance to indicate drowsiness is sound on several counts. Ideally, one would monitor PERCLOS directly. Work is ongoing to monitor eye closure directly (American Trucking Associations Foundation, 1996; Richard Grace, personal communication, 1998). Presently, available prototype systems are not sufficiently robust to be used in automobiles. Furthermore, the technical challenges associated with unobtrusive eye closure measurement over a wide range of driving conditions (e.g., variations in lighting, use of sunglasses) are daunting. This suggests that indirect as well as direct measurement of drowsiness (via PERCLOS) are worth further development effort.

The mechanisms behind performance deterioration with increased fatigue or drowsiness are only partially understood. However, three sources are of particular relevance to this issue. First, Brown (1994) describes fatigue effects in terms of general and selective withdrawal of attention. General withdrawal of attention involves degraded vehicle control as well as degraded object and event detection. These effects are thought to be due primarily to eye closure. Selective withdrawal of attention is more insidious in that it leaves lanekeeping performance intact yet degrades object and event detection. This type of effect is due to central processing degradation that might manifest itself in "lost in thought" crash contributors. Clearly, the algorithmic approach assessed here is compatible with detection of general withdrawal of attention effects. It may be less applicable to selective withdrawal of attention effects.

A second source relevant to relating drowsiness to degraded driver performance is Hockey (1986). In his review of the fatigue literature, Hockey points out that a central feature of the fatigue state is an aversion to effort. That is, the hypothesis is that prolonged work effects manifest themselves in terms of less active control over behavior and the selection of easy but more risky alternatives. Research by Holding (1974) is cited wherein it was experimentally demonstrated that with prolonged work, test participants chose low effort/low probability of success options. This way of thinking about fatigue in the driving context lead to predictions of less effort applied to "crisp" vehicle control with nominal increased (though subjectively acceptable) crash risk.

A third source that suggests how drowsiness might affect degraded driver performance comes from optimal control models of driving performance. In such models (Levison, 1989), control activities are pursued in an attempt to minimize a cost function expressed in terms of control precision (e.g., lanekeeping error) and control effort (e.g., steering inputs). Depending on the driver's strategy, these parameters may be traded off to some extent, including that suggested by Holding (1974) under conditions of drowsiness.

The results of the drowsy detection algorithm F4e-3 found that it performed less well than in the simulator studies. However, classification was actually relatively good. An attempt to improve the fit of algorithm F4e-3 to the real world data set was largely unsuccessful. In particular, except for LNMNSQ, no other predictor variable from algorithm F4e-3 had a regression weight significantly different from zero. There are many possible reasons for this. However, one of the most plausible reasons why variables such as INTACDEV or STVELV had no predictive power in the real world data set is because of range restrictions in the real world data. It is possible that, in the simulator, drivers exhibited a greater range of variability in lateral acceleration or steering inputs. The simulated environment did not, for example, include roadside appurtenances or ditches that are common in the real world. This would suggest that drivers on the road simply could not and did not let their driving deteriorate to the same degree as the test participants in the Virginia Tech research.

Further evidence lending support to this explanation comes from the Virginia Tech research program itself. Early on, simulator studies were conducted in such a fashion that fatigued drivers were not penalized for poor lanekeeping. In later studies, this protocol was changed to emphasize better lanekeeping and new models were needed as a result. It is plausible to carry this progression further from the simulator to the real world and find even greater penalties for poor lanekeeping (e.g., bodily injury and property damage!) that affect the form and performance of detection algorithms.

A very simple rule, observed LNMNSQ of 3.00 ft^2 or greater, was found to produce the best drowsiness classification accuracy. Allen, Parseghian, and Stein (1996) reported that as the standard deviation of lane position increased beyond about 0.8 ft (relative to lane center), the probability of a lane departure goes up dramatically. The square root of the LNMNSQ criterion of 3.00 ft^2 is 1.73 ft, well beyond the cutoff of Allen et al. (1996). The threshold value of 3.00 ft^2 therefore represents an extreme performance profile, one that merits attention.

The possibility exists that increased lanekeeping variability might sometimes result from deliberate driver action. In particular, curve cutting was examined indirectly by plotting the variance of low-pass filtered yaw rate both for False Positive and True Negative epochs. The plots provide some indication that False Positive epochs sometimes involved curvier road segments than the true negative epochs but the effect appears to be small. The implications of deliberate driver action on a driver status monitoring system are discussed later in this section.

It was thought possible that false positives were related to driver visual inattention that manifested itself in degraded lanekeeping. One indication of this would be longer times the driver looked away from the road scene ahead for the false positive epochs as opposed to the true negative epochs. To examine this hypothesis, empirical quantile plots of the eyes-off-road time proportion (EORTP) were prepared. Unexpectedly, the EORT proportions were often longer for the true negative epochs, i.e., the periods where no degraded lanekeeping performance was manifested. Thus, no evidence in support of the hypothesis was found.

A third look at the nature of the false positive epochs was attempted by examining their relationship to true positive epochs or drowsiness. It is interesting to note that, in the present data set, Test Participant 3 generated 17 out of 19 false positive epochs. A close review of the sequence of those false positive epochs along with the true drowsiness epochs generated by this test participant revealed that they were highly interleaved. This suggests that the erratic driving that triggered a false positive indication may have been an attempt by the driver to fight off the effects of drowsiness. If so, these false positives were in themselves drowsiness-related and so might not be perceived as nuisance alarms by such a driver. An alternative interpretation is that this interleaving represents some form of cyclical effect of drowsiness wherein driving performance is degraded while eye closure is unaffected, followed by periods when both performance is degraded and eye closure increases. Further research is needed to determine how drivers might process warnings under such situations.

Wierwille et al. (1996a) expressed the opinion that, though it is not desirable to have a system that misses detections, it is less desirable to have a system that produces a large number of false alarms since this erodes driver acceptance. As mentioned previously, in one portion of their algorithm development they redefined the phenomena to be detected as drowsiness <u>or</u> degraded driving. The latter was operationally defined in terms of observed rather than estimated measures of lanekeeping. This had a salutary effect in eliminating virtually all false alarms. Elsewhere, Wierwille has pointed out that a driver who has a drowsy driver detection system installed in his or her vehicle would have to receive instruction to maintain good lanekeeping performance. (It should be kept in mind that the test participants in the naturalistic driving study received no such instruction). If the driver wished to depart the lane, (e.g., to make a lane change, to cut a curve, or to deviate to avoid an object in the lane), the driver might be instructed to engage the turn signals, even if only momentarily. The detection system could then be programmed to delete a segment of data from the determination of estimated PERCLOS, thereby reducing false alarm rates substantially. Theoretically, this would leave only unintended lane deviations to be detected. It is an open research issue how a driver would react to a driver performance monitoring system and the instruction to maintain lanekeeping except with purposely deviating. It would be desirable to repeat the study with a larger number of test participants and instructions to use turn signals when intentionally departing the lane.

It is clear that detection algorithms presented here for drowsy or degraded driving are imperfect. There are many missed detections of drowsiness regardless of the approach taken here. In such cases, above-threshold PERCLOS values were not accompanied by any steering or lanekeeping anomalies. In one instance, a test participant had an eye closure bout that lasted approximately 13 seconds. Despite driving at highway speeds, the video and MicroDAS data revealed no lane departures. Presumably this occurred because there were no wind gusts, road surface variations, objects, or other "forcing functions" to perturb the vehicle for that segment of driving. Thus, there will sometimes be periods of drowsy driving in which there are no disturbances to the vehicle that result in drowsiness indicators rising above threshold for detection. It should be recognized, however, that performance degradation and eye closures do not always occur simultaneously. However, they usually do occur in time proximity. Thus, averaging should increase the probability of detection. Wierwille, et al. (1996a) have generally used 3- and 6-minute averages for this very reason.

It is likely to be the case that the best in-vehicle driver status monitoring system will use a multiplicity of approaches which are collectively better than any single approach alone. The present research has found that a single observed measure, with a fixed threshold value, reduced all false alarms (when drowsy driving <u>or</u> degraded lanekeeping were to be detected) and provided an approximately 63% hit rate. These results indicate a promising direction for future research and development to provide driver support to combat the potentially lethal effects of drowsiness and degraded lanekeeping while driving. It should be noted, however, that the lane tracking data were frequently missing or of unreliable quality. As mentioned earlier, the lane trackers used in MicroDAS are essentially cameras whose images are processed to discern the lane line pavement marking from the surrounding pavement. Worn lane lines, specular reflections, and other phenomena contributed to lanekeeping data loss. More robust systems (i.e., ones that process the optical flow field rather than lane lines) are therefore needed to implement even the simple rules that depend on measured lanekeeping.

If a multiplicity of approaches or technologies are needed for driver status monitoring, a multiplicity of approaches are needed to address the overall problem of drowsy and inattentive driving. It was mentioned earlier that drivers whose vehicles are equipped with such systems must be instructed in the system's use. More generally, education and outreach programs could provide a pamphlet or videotape to individuals when they get their driver's license renewed. The materials contained therein would provide information on the dangers of drowsy or inattentive driving, signs or symptoms to note, proper sleep hygiene, scientifically based facts on drowsiness countermeasures, and the like. Perhaps a nominal charge (e.g., 25 cents for a pamphlet or $2 for a videotape) would provide an incentive to the driver to actually review the material.

Under certain circumstances, the benefit of a drowsy driver detection system depends on the extent to which the driver will adhere to a warning to pull off the road and rest. For some people, the infrastructure plays a role in their willingness to stop driving. People are sometimes reluctant to pull off the road unless there is a rest stop available. Thus, there could be more rest stops put in at the side of the road, including interstate, state, county, and possibly township roads. The rest stops might only need to be a small gravel area. If no rest stops are available then drivers would have to improvise as best they could (e.g., pull off onto a side street to rest). It is ironic that while efforts are underway to improve drowsiness detection and driver awareness of the dangers of fatigue while driving, some states are closing rest stops to save money.

Just as there is concern about driver acceptance, so there should be concern about risk compensation or over-reliance on a driver status monitoring system. It is important to research the extent and circumstances under which this driver reaction to the technology might arise as well as means to counteract it. Indeed, if the level of missed detections achieved with the best approach still leaves perhaps 30% to 40% of drowsiness periods undetected, this might have the benefit of counteracting over-reliance. For example, some degree of missed detections (either caused by inherent system limitations or built-in random error) might be found such that, while the driver finds value in the system, he or she does not over rely on the system. Given that all sensor systems occasionally fail, over reliance on such technologies presents a real concern for broad scale implementation. Beyond this, there is concern that a driver may use such a system to keeping driving (e.g., “I can keep going cause it will warn me if necessary”). This type of unintended consequence also merits assessment in future research.

The question of what can be done once drowsiness has been detected should be researched further. More in-depth studies or surveys may find methods of fighting the affects of drowsiness and determine what methods truly work under what conditions. Example countermeasures could be mint odors or mint candy, loud music, cold air, short duration exercise, caffeine, chewing ice cubes or fruit snacks, shaking of the head, rocking fore/aft, taking a 10 to 20 minute nap, or getting involved in a mentally stimulating activity such as listening to a controversial talk show.

How should an ideal warning system work? An intelligent system would, by definition, perform like another attentive person in the vehicle and would, therefore, have situational awareness. The system would know the condition of the driver, the driver's habits, weather conditions, lighting conditions, time of day, and sense when the driver is not driving like he or she normally does. Conceivably, the driver could also "tell" this intelligent monitor and warning system how he or she is doing or feeling, to adjust the system sensitivity up or down, as prevailing conditions dictate. Such an intelligent system is clearly some distance off in the future, though research to develop it should be ongoing now.

In addition to technology, education and outreach, and infrastructure interventions, dealing effectively with the drowsy or inattentive driver problem should also encompass law enforcement. For example, if there have been any cases where the legal system has found a driver guilty of intentional manslaughter due to driving while drowsy, then all drivers should be made aware of this. There are current penalties in place for driving while drunk and outreach programs incorporate them as appropriate. In principle, then, there might be strict penalties if it can be shown without a doubt that a person had an accident due to driving while profoundly drowsy, thereby placing himself or herself in peril, along with fellow road users.

Additional research could be conducted to determine the usefulness of a system that needs three minutes of data collection and processing prior to any kind of drowsiness warning to a driver. It would be useful to determine the effectiveness of a system that provided drowsiness warnings within shorter and longer time intervals. A driver status monitoring system that monitored driver performance (e.g., unintended lane departure) would presumably have to warn the driver very quickly to avoid a roadway departure, opposite direction, or lane change crash, for instance. In the final recommended system design proposed Wierwille et al. (1996a), an instantaneous roadway departure system was integrated into the longer-term status/performance monitoring system in an attempt to "catch" rapidly occurring lapses in driver alertness and attention.

Due to pressures involved with our society, people frequently do drive even though they know that they are drowsy. They attempt to maximize the use of their time by sometimes engaging in distracting activities while driving. These problems of drowsy and inattentive driving will most likely only get worse, as schedules get more hectic. Therefore, a multi disciplinary approach to improving highway safety is not only advisable, it may be imperative.

5.0 REFERENCES

Allen, R. W., Parseghian, Z., and Stein, A. C. (1996). A driving simulator study of the performance effects of low blood alcohol concentration. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 943- 951). Santa Monica, CA: Human Factors and Ergonomics Society.

American Trucking Associations Foundation (Ed.). *Proceedings: Technical conference on enhancing commercial motor vehicle driver vigilance*. Alexandria, VA: American Trucking Associations Foundation.

Artaud, P. Planque, S., Lavergne, C., Cara, H., de Lepine, P., Tarriere, C., and Gueguen, B. (1994). *An on-board system for detecting lapses of alertness in car driving* (Paper No. 94-S2-O-08). Paper presented at the 14th International Technical Conference on Enhanced Safety of Vehicles, Munich, Germany, May 23-26, 1994.

Barickman, F. (1998). *Intelligent data acquisition for intelligent transportation research* (SAE Technical Paper No. 981198). Warrendale, PA: Society of Automotive Engineers

Brown, I. D. (1994). Driver fatigue. *Human Factors,* 36(2), 298-314.

Butler, L. (1998). *A further focus on fatigue.* Workshop presentation given at the Lifesavers 16 National Conference on Highway Safety Priorities, Cleveland, OH, March 29 - April 1, 1998.

Garder, P., and Alexander, J. (1995). *Fatigue related accidents and continuous shoulder rumble strips (CSRS*) (Paper No. 950132). Paper presented at the Transportation Research Board 74th Annual Meeting, January 22-28, 1995, Washington, DC.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical methods for data analysis*. Belmont, CA: Wadsworth Publishing.

Grimm, L. G., and Yarnold, P. R. (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association Press.

Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.) *Handbook of perception and performance, Volume II* (pp. 44-1 to 44-49). New York: John Wiley and Sons.

Holding, D. H. (1974). Risk, effort, and fatigue. In M. G. Wade and R. Martens (Eds.), *Psychology of motor behavior and sport*. Urbana, IL: Human Kinetics. Cited by Hockey (1986).

Institute for Traffic Safety. (1998). *Summary of New York State's comprehensive approach to addressing drowsy driving*. Albany, NY: Institute for Traffic Safety Management and Research.

Itoi, A., Cilveti, R., Voth, M., Dantz, B., Hyde, P., Gupta, A., and Dement, W. C. (1993, February). *Relationship between awareness of sleepiness and ability to predict sleep onset.* Washington, DC: AAA Foundation for Traffic Safety.

Knipling, R. R., and Wang, J.- S. (1994, November). *Crashes and fatalities related to driver drowsiness/fatigue*. NHTSA Research Note . Washington, DC: National Highway Traffic Safety Administration.

Koenig, M., Pape, D. and Pomerleau, D. (1995, September). *Run-off-road collision avoidance using IVHS countermeasures Task 4 Interim Report: Volume II: RORSIM manual*. (Contract No. DTNH22-93-C-07023). Washington, DC: National Highway Traffic Safety Administration.

Levison, W. H. (1989). The optimal control model for manually controlled systems. In G. R. McMillan et al. (Eds.), *Applications of human performance models to system design* (pp. 185 - 198). New York: Plenum Press.

National Sleep Foundation. (1997, June). *Use of continuous shoulder rumble strips: A consensus report of the National Sleep Foundation*. Washington, DC: National Sleep Foundation.

SAS Institute. (1990). *SAS/STAT users guide: Volume 2* (Fourth edition). Cary, NC: SAS Institute Inc.

Senders, J. W., Kristofferson, A. B., Levison, W. B., Dietrich, C. W., and Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, *195*, 15 - 32.

Siegmund, G. P., King, D. J., and Mumford, D. K. (1996). *Correlation of steering behavior with heavy-truck driver fatigue* (SAE Technical Paper Series Paper No. 961683). Warrendale, PA: Society of Automotive Engineers.

Tijerina, L. (1996). A taxonomic analysis of crash contributing factors and prospects for ITS crash countermeasures. *Proceedings of the Third Annual World Congress on Intelligent Transport Systems*. (Compact Disc). Orlando, FL., October 14 - 18, 1996.

Wang, J.-S., Knipling, R. R., and Goodman, M. J. (1996). The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System. In *40th Annual Proceedings: Association for the Advancement of Automotive Medicine* (pp. 377-392). De Plaines, IL: Association for the Advancement of Automotive Medicine.

Wierwille, W. W. (1993). *Visual and manual demands of in-car controls and displays*. In B. Peacock and W. Karwowski (Eds.), Automotive Ergonomics (pp. 299-320). London: Taylor and Francis.

Wierwille, W. W., Lewin, M. G., and Fairbanks, R. J. III. (1996a, September). *Final Report: Research on vehicle-based driver status/performance monitoring, Part I* (Tech. Report No. DOT HS 808 638). Washington, DC: National Highway Traffic Safety Administration.

Wierwille, W. W., Lewin, M. G., and Fairbanks, R. J. III. (1996b, September). *Final Report: Research on vehicle-based driver status/performance monitoring, Part II* (Tech. Report No. DOT HS 808 638). Washington, DC: National Highway Traffic Safety Administration.

Wierwille, W. W., Lewin, M. G., and Fairbanks, R. J. III. (1996c, September). *Final Report: Research on vehicle-based driver status/performance monitoring, Part III* (Tech. Report No. DOT HS 808 638). Washington, DC: National Highway Traffic Safety Administration.

Wierwille, W. W., Wreggit, S. S. Kirn, C. L., Ellsworth, L. A., and Fairbanks, R. J. (1994, December). *Research on vehicle-based driver status/performance monitoring: Development, validation, and refinement of algorithms for detection of driver drowsiness* (Tech Report No. DOT HS 808 247). Washington, DC: National Highway Traffic Safety Administration.

Wood, N. E. (1994). *Shoulder rumble strips: A method to alert "drifting"* drivers (Paper No. 940312). Paper presented at the Transportation Research Board 73rd Annual Meeting, January 9-13, 1994, Washington, DC.

Wylie, C. D., Shultz, T., Miller, J. C., Mitler, M. M., and Mackie, R. R. (1996, October). *Commercial motor vehicle driver fatigue and alertness study : Project Report* (Report No. FHWA-MC-97-002). Washington, DC: Federal Highway Administration Office of Motor Carriers.