Detection of Driver Impairment from Drowsiness

Timothy Brown
University of Iowa
United States
John Lee
University of Wisconsin-Madison
United States
Chris Schwarz
University of Iowa
United States
Dary Fiorentino
DF Consulting
United States
Anthony McDonald
University of Wisconsin-Madison
United States
Eric Traube
National Highway Traffic Safety Administration
United States
Eric Nadler
Volpe National Transportation Systems Center
United States

Paper Number: 13-0346

ABSTRACT

Drowsy driving is a significant contributor to death and injury crashes on our nation's highways accounting for more than 80000 crashes and 850 fatalities per year. Recent research using data from the 100-car study found that drowsy driving contributed to 22% to 24% of crashes and near-crashes observed. This paper describes an approach that detects impairment from drowsiness in real time using inexpensive vehicle-based sensors to detect drowsiness-related changes in drivers' behavior.

Data were collected on the National Advanced Driving Simulator from 72 volunteer drivers. Three age groups (21-34, 38-51, and 55-68 years of age) drove through representative situations on three types of roadways (urban, freeway, and rural) at three times of day (9 am-1 pm, 10 pm-2 am, and 2 am – 4 am) representing different levels of drowsiness.

Driving data indicated that a complex relationship exists wherein driving performance improves with low levels of drowsiness in the early night session before degrading in the late night session. This study demonstrates the feasibility of detecting drowsiness with vehicle-based

sensors. Results show that alcohol and drowsiness impairment do not allow for a single algorithm to detect both types of impairment; however a single algorithm approach with different training data for the different types of impairment may be successful. To detect impairment due to either alcohol or drowsiness, a more complex approach is necessary where separate algorithms are combined to work with each other. These results suggest promise in a vehicle-based approach to detecting and differentiating multiple types of impairment.

INTRODUCTION

Exact counts of the number of crashes caused by drowsiness are hard to obtain due to the use of varying methodologies. The Gallup organization surveyed drivers and estimated that during the 5 years prior to 2002 and found that 1.35 million drivers may have been involved in a drowsy driving related crash [1]. A National Highway Traffic Safety Administration (NHTSA) report, which contained crash report data between 2005 and 2009, attributed 83,000 crashes per year and 886 fatal crashes per year to drowsy, fatigued, or sleeping drivers. Over the five-year period these causes resulted in 5,021 fatalities.

Other research methods and driver populations lead to different estimates for the percentage of drowsy driving crashes. The 100-car naturalistic driving study found that drowsy driving contributed to 22% to 24% of crashes and near-crashes observed [2]. In a report to congress, NHTSA stated that 3.2 % of crashes were related to actual sleep [3]. An estimated 1% of all large-truck crashes, 3–6 % of fatal heavy-truck crashes, and 15–33% of fatal-to-the-truck-occupant-only crashes have been attributed to driver fatigue as a primary factor [4]. Although the methodologies result in different estimates, all point to a significant problem.

According to the National Sleep Foundation's 2009 annual Sleep in America survey, 28 percent of drivers had driven while drowsy at least once per month in the past year. Of those that drove while drowsy, 28% have fallen asleep [5]. A survey conducted in 2003 found that 37% of drivers have fallen asleep for at least a moment (nodded off) while driving at least once in their driving career, while 8% of them had done it in the last six months.

Of those encountering a sleeping episode, 58% of drivers were on a multi-lane interstate highway, and 92% of them were startled awake and of those who were startled awake, 33% wandered into another lane or shoulder, 19% crossed the centerline and 10% ran off road [1]. Drowsy driving is not only common in the United States, it was found that one in five Canadian drivers have admitted to nodding off or falling asleep at least once while driving [6] and that driver fatigue contributes to at least 9% to 10% of crashes in the UK [7].

Clearly, there is cause for concern about the rate of drowsy driving and the resultant crashes, injuries and fatalities. This concern creates a need for research to facilitate the development of technological approaches that will reduce the number of lives lost due to drowsy driving. The present aim is to extend Impairment Monitoring to Promote Avoidance of Crashes using Technology or IMPACT, a program of research into detecting alcohol-impaired driving based primarily upon vehicle-based measures to the domain of drowsy driving [8]. IMPACT has developed alcohol detection algorithms for all drivers (general algorithms) and algorithms that take into account individual driving differences (individualized algorithms). This work explores how well

the previously developed algorithms that detect impairment from alcohol are able to detect drowsiness, and how best to modify those algorithms, if necessary, to detect both. The algorithms that were previously developed to detect alcohol impairment were effective at levels comparable to the Standardized Field Sobriety Test in eight to twenty-five minutes.

Although there were many objectives to this research, this paper will focus on the following:

- Can existing algorithms for detecting impairment be used to detect drowsiness?
- Can real-time algorithms reliably detect drowsiness in advance of a drowsiness-related mishap?

METHOD

Participants

Data were collected from 72 volunteer drivers from three age groups (21-34, 38-51, and 55-68 years of age) driving through representative situations on three types of roadways (urban, freeway, and rural). Participants drove at three times of day (9 am-1 pm, 10 pm-2 am, and 2 am – 4 am) to induce different levels of drowsiness.

To be eligible, participants were required to:

- Possess a valid US driver's license
- Have been licensed driver for two or more years
- Drive at least 10,000 miles per year
- Have no restrictions on driver's license except for vision
- Not require the use of any special equipment to drive.

Procedure

An initial telephone interview was conducted to determine eligibility for the study. Applicants were screened in terms of health history, current health status, medication and drug usage, morning/evening tendencies [9], and sleep apnea history [10]. Pregnancy, disease, or evidence of sleep apnea or being a night person were excluded from the study as were those taking prescription medications that cause drowsiness.

Each participant participated in four sessions over three visits. The two overnight drives occurred on a single night. The daytime and nighttime data collection visits were separated by three days and the order of these visits and scenario event sequence were counterbalanced

On study Visit 1 (screening), each participant informed consent was obtained. They then provided a urine sample for the drug screen and, for females, the pregnancy screen. During a five-minute period following these activities, the participant sat alone in the room where subsequent measurements of blood pressure, heart rate, height, and weight were made.

Cardiovascular measures were taken and compared to acceptable ranges (systolic blood pressure = 120 ± 30 mmHg, diastolic blood pressure = 80 ± 20 mmHg, heart rate = 70 ± 20) to assess eligibility for the study. If participants met study criteria, they completed demographic surveys. These surveys included questions related to crashes, moving violations, driver behavior, sleeping, and driving history. Participants viewed an orientation and training presentation that provided an overview of the simulator cab and the secondary task they were asked to complete while driving.

The task consisted of the participant turning on the CD player and sequentially advancing the CD player to two tracks provided in an auditory cue and then turning off the CD player.

Participants then completed the practice drive and completed surveys after their drive about how they felt and about the realism of the simulator. The practice drive included making a left hand turn, driving on two- and four-lane roads, and changing CDs. If the participant remained eligible, baseline EEG measurements were recorded. Prior to their study visits, participants were provided with activity monitors and activity logs to verify sleep preceding the visits.

During the daytime-alert visit, participants were asked to not ingest any caffeine. Logs were reviewed to verify a normal night's sleep (at least six hours) the preceding night. Their BAC was checked to ensure that they were not under the influence of alcohol. Participants were then fitted with the wireless B-Alert X-10 EEG cap [11], [12] and electrodes to record their EEG and heart rate. The participants then entered the simulator and eye tracking calibrations were completed.

Prior to beginning the drive, the participants also completed a questionnaire about their current sleepiness level, the Stanford Sleepiness Scale [13], and a version of the Psychomotor Vigilance Task or PVT (Cognitive Media Iowa City, IA) based on the Psychomotor Vigilance Task [14], [15]. The participants drove through the simulation scenario.

Following the drive, participants were again administered the Stanford Sleepiness Scale, the wellness survey, PVT, a Retrospective Sleepiness Scale, and a simulator realism survey. The Retrospective Sleepiness Scale required subjective judgments of drowsiness at specified scenario locations. The B-Alert cap was then removed.

During the nighttime-drowsy visit, participants were instructed to restrict beverage consumption to water after 12:00 pm on the day of their overnight visit, to minimize caffeine intake. Participants were picked up at their homes after having eaten dinner, and transported to the simulation facility to arrive around 7pm. Logs were reviewed to verify a normal night's sleep (at least six hours) the preceding night and that they did not take any naps during the day. Caffeine intake was reviewed and if caffeine was consumed after noon on the day of the overnight drive, the participant was either rescheduled or dropped from the study. Participants were then fitted with the B-Alert monitoring device.

A variety of activities were provided to keep participants awake including activities on an iPad, reading, playing computer games, etc. They were monitored to ensure they did not fall asleep or converse with other participants. If participants began to fall asleep, they were engaged by a researcher to keep them awake. The participants completed the Stanford Sleepiness Scale every 30 minutes until they drove. One hour prior to their drive, they were taken to a private room to wait. They completed a PVT at this time, and also at 30 minutes prior to the drive. Participants were escorted to the simulator between 22:00 and 01:00 for their first drives. Once in the simulator, eye tracking calibration procedures were performed, and the B-Alert electrode connection was verified. Before starting the drive, the participants completed a PVT and Stanford Sleepiness Scale. After the drive, participants completed the Stanford Sleepiness Scale, a Wellness Survey, a PVT, and a Retrospective Sleepiness Scale.

Participants were then escorted back to a separate waiting area where TV, movies, reading, computer games, etc. were available. A Stanford Sleepiness scale was administered every 30 minutes until their next drive. One hour prior to their second drive times, participants were again taken to a private room to wait. They completed a PVT one hour prior to the drive and also at 30 minutes prior to the drive. Participants were escorted to the simulator between 02:00 and 05:00 for their second drives. Once in the simulator, eye tracking calibration procedures were performed, and the B-Alert connection was verified. Before starting the drive, the participants completed a PVT and Stanford Sleepiness Scale. After the drive, participants completed Stanford Sleepiness Scale, a Wellness Survey, a PVT, a retrospective sleepiness scale, and a realism survey. The B-Alert system was then removed and they were transported home

Apparatus

The National Advanced Driving Simulator (NADS), shown in Figure 1, made it possible to collect representative driving behavior data from drowsy drivers in a safe and controlled manner. This is the highest fidelity simulator in the United States and allowed for precise characterization of driver response. Drivers' control inputs, vehicle state, driving context, and driver state were captured in representative driving situations (see Figure 2).

Simulator Scenario

Each drive was composed of three nighttime driving segments. The drives started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph (see Figure 3 and Figure 4) with signal-controlled and uncontrolled intersections. An interstate segment followed that consisted of a four-lane divided expressway with a posted speed limit of 70 mph. Following a period in which drivers followed the vehicle ahead, they encountered infrequent lane changes associated with the need to pass several slower-moving trucks (see Figure 5). The drives concluded with a rural segment that was composed of a two-lane undivided road with curves (see Figure 6); followed by a gravel road segment; and then a 10-minute section of straight rural driving.

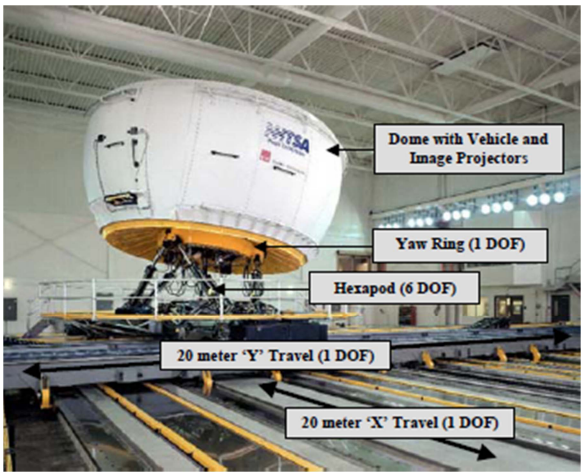

Figure 1. The NADS-1 high-fidelity driving simulator.

Figure 2. An urban driving scene from the NADS-1 simulator.

Figure 3. Approach to curve in urban drive

Figure 4. Straight roadway segment in urban drive

Figure 5. Passing truck on Interstate.

Figure 6. Approach to rural curve

RESULTS

Sensitivity of Scenarios to Drowsiness

An analysis of common driving metrics of variability in speed and lane keeping demonstrates the sensitivity of the NADS-1 to drowsiness. Driving data indicated that a complex relationship exists wherein driving performance improves with low levels of drowsiness in the early night session before degrading in the late night session (see Figure 7). Session time of day did not interact with age, gender, or roadway situation.

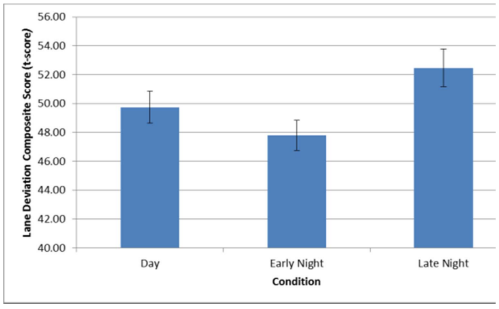

Figure 7. Lane deviation scores by Drowsiness Condition

Detecting Impairment from Drowsiness

The primary objectives for algorithm development and evaluation include:

- Evaluate existing algorithms for detecting impairment for detection of drowsiness
- Determine if real-time algorithms can reliably detect drowsiness in advance of a drowsiness-related mishap

An initial set of algorithms was selected for evaluation based on prior studies in this line of research and a review of the literature. For those selected from the open literature, only those algorithms with enough detail for real-time implantation were considered. The algorithms used in the evaluation are documented in Table 1.

Table 1. Summary of Algorithms

Label	Algorithm	Inputs	Outputs
PC	PERCLOS [16]	Eye closure	Continuous percentage Drowsy binary
PC+	PERCLOS+ [17]	Eye closure, lane departure	Drowsy categorical (low, moderate, severe)
SB	Steering-based [18]	Steering angle, steering rate	Drowsy binary
EEG	EEG [19]	Scalp electrical activity	Continuous probability Drowsy binary
DT	Decision Tree[8]	Multiple measures of driver performance	Intoxicated binary
MDD	Multi-Distraction Detection [20]	Eye gaze location	Continuous PRC Visual binary Cognitive binary
TLC	Time-to-lane-crossing[19]	Lane position, lane heading angle	Drowsy binary
SRF	Steering random forest [19]	Steering wheel angle	Drowsy binary
BN	Bayes Net [19]	Multiple measures of driver performance, eye closure, eye closure rate	Intoxicated categorical (none, moderate, severe)

Table 2 shows algorithm performance in detecting drowsiness, as defined by drivers' ratings of sleepiness using the SSS after they completed the drive. Drowsiness was indicated by post SSS of five or greater and alertness by post SSS of three or less. In this table, the algorithms were assessed according to how well they differentiated between drivers with a rated sleepiness score of three or less and those with a score of five or greater.

Three standard criteria were used to assess algorithm performance in detecting and distinguishing impairments: accuracy, positive predictive performance (PPP), and area under curve (AUC). Accuracy measures the percent of cases that were correctly classified, while PPP measures the degree to which those drivers that were judged to be drowsy were actually drowsy. An algorithm can correctly identify all instances of impairment simply by setting a very low decision criterion, but such an algorithm would misclassify all cases where there was no impairment. The relationship between the true positive detection rate (sensitivity) and false positive detection rate (1-specificity) is represented by the receiver operator characteristic (ROC) curve. AUC represents the area under the receiver operator curve, which provides a robust and simple performance measure. Perfect classification performance is indicated by an AUC of 1.0, and chance performance is indicated by .50. AUC is an unbiased measure of algorithm performance, but accuracy and PPP are more easily interpreted, so all three are used in describing the algorithms.

Surprisingly, all algorithms performed poorly with only the PERCLOS algorithm having a confidence interval that did not include.50. The mean AUC for the PERCLOS algorithm was only.61, meaning

that if the driver was drowsy the algorithm would only have a 61% chance of correctly detecting the drowsiness.

Table 2 Impairment detection algorithm performance based on post-drive sleepiness ratings with 95% confidence intervals

Label	Algorithm	AUC	PPP	Accuracy
MDD	Multi-Distraction Detection	.51 (.45-.61)	.59 (.55-.62)	.55 (.53-.55)
EEG	EEG	.58 (.48-.65)	.54 (.53-.55)	.59 (.56-.61)
PC	Perclos	.63 (.53-.70)	.60 (.59-.60)	.59 (.55-.61)
PC+	Perclos+	.53 (.43-.60)	.59 (.58-.60)	.54 (.53-.59)
SB	Steering-based	.55 (.48-.62)	.59 (.58-.59)	.56 (.54-.59)
BN	Bayes Network	.45 (.38-.57)	.48 (.45-.51)	.49 (.47-.51)

The algorithm developed to detect distraction (MDD) performed very poorly, comparable to a random classifier. Similarly, the Bayes Network trained to detect alcohol impairment also performed very poorly, and algorithms developed to detect drowsiness performed almost as poorly. Overall, these results show that algorithms developed to detect other impairments will not necessarily detect overall drowsiness as determined by SSS rating.

In switching to the second objective, it is not surprising that algorithms detecting impairment defined by the drowsiness condition performed poorly when the variability of drowsiness across conditions, drivers, and the drive are considered. The transient nature of drowsiness suggests that algorithms that detect impairment associated with driving mishaps, such as lane departures, might be substantially more sensitive.

To assess this possibility, real-time algorithms were developed using data from a small time window, with a focus on data surrounding lane departures. The continuous data consists of driver and vehicle data recorded at 60 Hz for the entire drive. Each record of these datasets was coded as alert or drowsy according to three definitions: the drowsiness condition (Day, Early Night, Late Night), a linear combination of PVT, pre-post and retrospective SSS, and the presence or absence of a lane departure.

Ten-fold cross validation was used to assess each algorithm, producing a measure of accuracy, PPP, AUC, timeliness and corresponding confidence interval for each algorithm. Timeliness is defined by the AUC of the ROC curve measured at six seconds before the lane departure. ROC curves summarize the performance graphically.

Data mining algorithms (Bayesian Networks and Random Forest), designed to detect and classify drowsiness in real time, successfully detected drowsiness six seconds before it resulted in a lane departure (see Figure 8and Figure 9. These algorithms were based on time-to-lane-crossing (TLC) and steering behavior using sensor data already available in cars. They performed better than PERcentage of CLOSure of the eyelid (PERCLOS), see Figure 10, which uses eye-tracking cameras that are not currently available in the vehicle fleet. We have demonstrated that inexpensive vehicle-based sensors can be used to successfully detect driver impairment

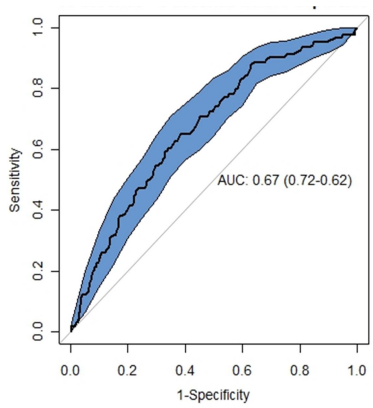

Figure 8. Receiver operator characteristic of steering

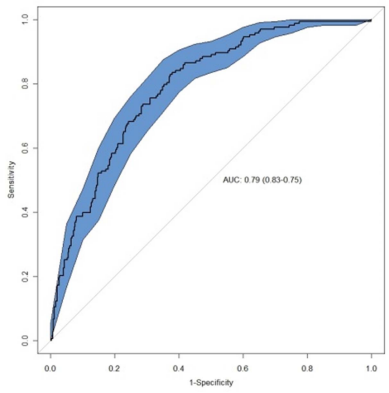

Figure 9. Receiver operator characteristic of TLC algorithm

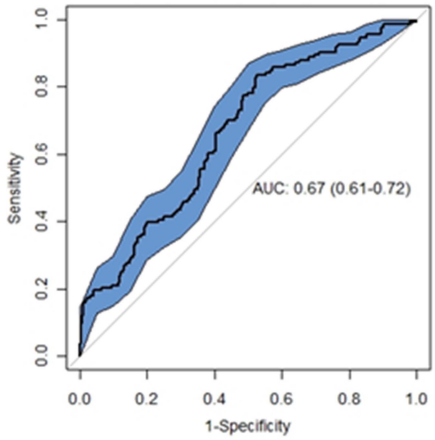

Figure 10. Receiver operator characteristic of PERCLOS algorithm

It was found that such algorithms could be generalized to detect both alcohol impairment and drowsiness with additional training or by combining multiple algorithms. However, algorithms that were trained to detect alcohol impairment did not perform well when simply applied to drowsiness and vice versa.

Drowsiness has a strong transient component, as compared with intoxication which is longer-lasting. In fact a Bayes Net was able to differentiate intoxication from the combination of intoxication and drowsiness, showing that the symptoms of the former do not necessarily mask those of drowsiness.

CONCLUSIONS

This study demonstrates the feasibility of detecting drowsiness with vehicle-based sensors. Results show that the differences in alcohol and drowsiness impairment do not allow for a single algorithm to detect both types of impairment; however similar algorithms trained independently may be successful. To detect impairment due to either alcohol or drowsiness, a more complex approach is necessary where separate algorithms are combined to work with each other. These results suggest promise in a vehicle-based approach to impairment detection including multiple types of impairment.

Future research should focus on examining distraction related impairment to evaluate the extent to which distraction can be detected when drivers are impaired from alcohol or drowsiness, and the extent to which impairment from alcohol, drowsiness and distraction can be distinguished. Then other types of impairments may also be considered, such as drugs and age-related cognitive decline.

Additional research should evaluate the extent to which existing impairment

detection algorithms are capable of detecting impairment from medications or illicit drugs. Many over the counter medications are known to produce drowsiness; however, because these medications produce a more uniform level of drowsiness compared to the transient nature of the natural onset of drowsiness, this type of impairment should be tested to determine if the algorithms developed to detect drowsiness as part of this research would detect driving impaired by medications or illicit drugs.

ACKNOWLEDGEMENTS

We would also like to thank the staff at the National Advanced Driving Simulator who put in long days and nights to collect the data for this project.

REFERENCES

[1] D. Royal, "National Survey of Distracted and Drowsy Driving Attitudes and Behavior:2002 Volume I: Findings," National Highway Traffic Safety Administration, Washington, DC, DOT HS 809 566, Apr. 2003.

[2] S. Klauer, T. Dingus, V. Neale, J. Sudweeks, and D. Ransey, "Impact of Driver Inattention On Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data," National Highway Traffic Safety Administration, Washington, DC, DOT HS 810 594, 2006.

[3] National Highway Traffic Safety Administration, "National Motor Vehicle Crash Causation Survey: Report to Congress," National Highway Traffic Safety Administration, Washington, DC, DOT HS 811 059, 2008.

[4] R. Knipling and T. Shelton, "Problem Size Assessment: Large Truck Crashes Related Primarily to Driver Fatigue," in *Proceedings of the Second International Large Truck Safety Symposium*, Knoxville, TN, 1999.

[5] National Sleep Foundation, "2009 Sleep in America Poll: Summary of Findings," National Sleep Foundation, 2009.

[6] D. Beirness, H. Simpson, and K. Desmond, "The Road Safety Monitor 2004: Drowsy Driving," Traffic Injury Research Foundation, Ottawa, ON, 2005.

[7] G. Maycock, "Sleepiness and driving: the experience of U.K. car drivers," *Accident Analysis & Prevention*, vol. 29, pp. 453–462, 1997.

[8] J. Lee, D. Fiorentino, M. Reyes, T. Brown, O. Ahmad, J. Fell, N. Ward, and R. Dufour, "Assessing the Feasibility of Vehicle-Based Sensors to Detect Alcohol Impairment," National Highway Traffic Safety Administration, Washington, DC, DOT HS 811 358, 2010.

[9] A. Adan and H. Almirall, "Horne and Osterberg Morningness Eveningness Questionnaire - a Reduced Scale," *Personality and Individual Differences*, vol. 12, no. 3, pp. 241–253, 1991.

[10] T. Brown, B. Dow, D. Trask, E. Dyken, and S. E. Salisbuery, "Driving Performance and Obstructive Sleep Apnea: A Preliminary Look at the Manifestations of Impairment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2096, pp. 33–40, 2009.

[11] Advanced Brain Monitoring, "B-Alert x-10." 2011.

[12] R. Johnson, D. Popovic, R. Olmstead, M. Stikic, D. Levendowski, and C. Berka, "Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model," *Biological*

Psychology, vol. 87, no. 2, pp. 241–250, May 2011.

[13] E. Hoddes, V. Zarcone, H. Smythe, R. Phillips, and W. Dement, "Quantification of Sleepiness: A New Approach," *Psychophysiology*, vol. 10, no. 4, pp. 431–436, 1973.

[14] R. Wilkinson and D. Houghton, "Field test of arousal: A portable reaction timer with data storage," *Human Factors*, vol. 24, pp. 487–493, 1982.

[15] S. Loh, N. Lamond, J. Dorrian, G. Roach, and D. Dawson, "The validity of psychomotor vigilance tasks of less than 10-minute duration," *Behavior Research Methods, Instruments & Computers*, vol. 36, no. 2, pp. 339–346, 2004.

[16] W. Wierwille, L. Ellsworth, S. Wreggit, R. Fairbanks, and C. Kim, "Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness," National Highway Traffic Safety Administration, Washington, DC, DOT HS 808 247, 1994.

[17] R. Hanowski, D. Bowman, A. Alden, W. Wierwille, and R. Carroll, "PERCLOS+: Moving Beyond Single-Metric Drowsiness Monitors," presented at the Society of Automotive Engineers Commercial Vehicle Engineering Conference, 2008.

[18] D. King, D. Mumford, and G. Siegmund, "An Algorithm for Detecting Heavy-Truck Driver Fatigue from Steering wheel Motion," in *The 16th ESV Conference Proceedings*, Windsor, Ontario, 1998.

[19] T. Brown, J. Lee, C. Schwarz, D. Fiorentino, and A. McDonald, "Driver Monitoring of Inattention and Impairment Using Vehicle Equipment (DrIIVE), Phase 1," National Highway Traffic Safety Administration, Washington, DC, Under Review.

[20] J. Lee, J. Moeckli, T. Brown, S. Roberts, C. Schwarz, L. Yekshatyan, E. Nadler, T. Victor, D. Marshall, and C. Davis, "Distraction Detection and Mitigation through Driver Feedback," National Highway Traffic Safety Administration, Washington, DC, 2013.