U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

NHTSA

DOT HS 813 564b September 2026

# NHTSA's DrIIVE Program: Drowsy Driving Lane Departure Mitigation Countermeasures

*This page is intentionally left blank.*

## DISCLAIMER



Suggested APA Format Citation:

Brown, T., Gasper, J., Schwarz, C., Lee, J., Schmitt, R., Marshall, D., & Ahmad, O. (2026, September). *NHTSA's DrIIVE program: Drowsy driving lane departure mitigation countermeasures* (Second of three parts. Report No. DOT HS 813 564b). National Highway Traffic Safety Administration. https://doi.org/10.21949/d7af-zg83

The other two parts of this series:

Brown, T., Schwarz, C., Gasper, J., Lee, J., Marshall, D., & Ahmad, O. (2026, September). *NHTSA's DrIIVE program: Phase 2 summary* (First of three parts. Report No. DOT HS 813 564a). National Highway Traffic Safety Administration. https://doi.org/10.21949/2xfn-fa85

Brown, T., Schwarz, C., Lee, J., Gaspar, J., Marshall, D., & Ahmad, O. (2026, September). *NHTSA's DrIIVE program: Alcohol-impaired, drowsy, and distracted driving detection algorithms* (Third of three parts. Report No. DOT HS 813 564c). National Highway Traffic Safety Administration. https://doi.org/10.21949/2b5z-0398

*This page is intentionally left blank.*

# Technical Report Documentation Page

| 1. Report No. | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| DOT HS 813 564b | | |

| | |
|---|---|
| **4. Title and Subtitle**<br>NHTSA's DrIIVE Program: Drowsy Driving Lane Departure Mitigation Countermeasures | **5. Report Date**<br>September 2026<br>**6. Performing Organization Code** |
| **7. Authors**<br>Timothy Brown, John Gaspar, Chris Schwarz, John Lee, Rose Schmitt, Dawn Marshall, Omar Ahmad | **8. Performing Organization Report No.**<br>DOT-VNTSC-NHTSA |
| **9. Performing Organization Name and Address**<br>National Advanced Driving Simulator<br>University of Iowa<br>2501 Oakdale Blvd<br>Iowa City, IA 52242<br><br>University of Wisconsin-Madison<br>1513 University Avenue, Room 3217<br>Madison, WI 53706 | **10. Work Unit No. (TRAIS)**<br>**11. Contract or Grant No.** |
| **12. Sponsoring Agency Name and Address**<br>National Highway Traffic Safety Administration<br>Office of Vehicle Safety Research<br>1200 New Jersey Avenue SE<br>Washington, DC 20590 | **13. Type of Report and Period Covered**<br>**14. Sponsoring Agency Code** |

**15. Supplementary Notes**

Digital Object Identifier: https://doi.org/10.21949/d7af-zg83

**16. Abstract**

***NOTE***: This document is a reformatted version of the report originally published in 2022 under docket NHTSA-2022-0079-0279. The technical content remains unchanged.

This is the second part of a three-part study. The goal of the Driver Monitoring of Inattention and Impairment Using Vehicle Equipment mitigation track was to study how drowsiness mitigations (alerts) affect driving performance. Forty-eight participants completed two drives in the National Advanced Driving Simulator during two visits. One drive occurred between 10 p.m. and 2 a.m. (early night) and one occurred between 2 a.m. and 6 a.m. (late night). During both drives, participants experienced one of six drowsiness alert interfaces. Drivers with mitigation committed a marginally lower rate of drowsiness-related lane departures and demonstrated significantly reduced variability in both lateral lane position and speed compared to drivers who did not receive alerts. Furthermore, staged alerts resulted in a lower rate of drowsy lane departures and lower lane keeping variability relative to the baseline than did discrete alerts.

| | |
|---|---|
| **17. Key Words**<br>drowsy driving, lane departure, alerts, drowsiness detection, driving simulator, DrIIVE, mitigation, NADS-1 | **18. Distribution Statement**<br>Document is available to the public from the DOT, BTS, National Transportation Library, Repository & Open Science Access Portal, https://rosap.ntl.bts.gov. |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Unclassified | Unclassified | 74 | |

**Form DOT F 1700.7** (8-72) Reproduction of completed page authorized

*This page is intentionally left blank.*

# Table of Contents

# List of Figures

# List of Tables

## Terms and Definitions

| | |
|---|---|
| accuracy | ratio of true positives plus true negatives to all positives plus all negatives |
| alert | stimulus from a DVI intended to warn the driver; see also *warning* |
| alert modality | mode of delivery of an alert including audio, visual and haptic modalities |
| algorithm | process to be followed for solving a problem; a computerized process to detect driving impairment |
| ANOVA | analysis of variance |
| AUC | area under an ROC curve; an algorithm evaluation metric that accounts for linked pairs of metrics such as an algorithms family of potential true and false positive rates |
| auditory alert | single beep, chime, or repeating beep issued by the DVI as part of a drowsiness mitigation system |
| BAC | blood alcohol concentration, measured in grams per deciliter (g/dL) |
| Bayesian network | probabilistic graphical model that represents a set of random variables and their conditional dependencies; a type of machine learning algorithm |
| caret[1] | Classification And REgression Training; a statistical toolbox in R used to train and test machine learning algorithms |
| circadian rhythm | physical, mental and behavioral changes that follow a roughly 24-hour cycle and responding to light and darkness |
| confusion matrix | simple way to represent an algorithm's predictions as true positives, true negatives, false positives and false negatives |
| decision tree | decision-making algorithm that consists of a number of sequential choices leading to a binary classification; a component of a random forest |
| discrete warning | binary classification system that issues an alert when the warning becomes active |
| DrIIVE | Driver monitoring of Inattention and Impairment Using Vehicle Equipment |
| driver-based | using sensors that monitor the driver, such as eye and head trackers, electroencephalogram or other physiological sensors |
| driver state | impairment state of the driver including normal, drowsy, distracted, or intoxicated |

[1] Although an acronym, the lowercase name is the official package identifier and distinguishes the software package from the mathematical caret symbol (^) used for exponents.

| | |
|---|---|
| DVI | driver-vehicle interface |
| EEG | electroencephalogram; a measurement of electrical activity in the brain |
| endogenous | having an internal cause or origin |
| energy of a signal | sum of the squared value of a signal at each instant in time |
| exogenous | having an external cause or origin |
| false negative | algorithm classifies as negative (normal), but the ground truth is positive (impaired) |
| false positive | algorithm classifies as positive (impaired), but the ground truth is negative (normal) |
| ground truth | actual presence or absence of impairment against which algorithm performance may be judged |
| haptic alert | single, double, or repeated vibratory pulse on the driver's seat issued by the DVI as part of a drowsiness mitigation system |
| HMM | hidden Markov model; a statistical model with unobservable states and observable evidence inputs |
| IMPACT | Impairment Monitoring to Promote Avoidance of Crashes Using Technology |
| impairment detection | detection of a driving impairment through the use of sensors and an algorithm |
| impairment mitigation | application of countermeasures through a warning system and DVI alerts when impairment is detected. |
| kappa | metric that compares observed accuracy with expected accuracy; a more appropriate metric than accuracy when there is a small proportion of positive classifications |
| mitigation state | state of the mitigation system including normal, stage 1, stage 2, or stage 3 |
| NADS | National Advanced Driving Simulator |
| ORD | objective rating of drowsiness |
| PERCLOS | measurement of the percentage of time the pupils of the eyes are 80 percent or more occluded over a specified time interval |
| PERCLOS+ | combination measure based on PERCLOS and lane departures |
| PPV | positive predictive value; the ratio of true positives to true positives plus false positives |
| PVT | psychomotor-vigilance test |
| random forest | machine-learning algorithm that uses a group of decision trees to vote on a binary classification problem |
| R | open-source statistical software and programming language |

| | |
|---|---|
| ROC | receiver operator characteristic; a way to represent binary classifier performance for a family of algorithms that vary simultaneously on pairs of parameters. Also called ROC curve. |
| RSS | retrospective sleepiness scale; a driver's attempt to draw their moment-by-moment SSS score over the length of their drive |
| SD | standard deviation |
| SDLP | standard deviation of lane position |
| sensitivity | ratio of true positives to true positives plus false negatives |
| specificity | ratio of true negatives to true negatives plus false positives |
| SSS | Stanford sleepiness scale |
| static algorithm | algorithm that operates on only current data, not on previous algorithm outputs |
| TA | truly awake; corresponds to daytime driving segments that either match both the driver and location of a TD, or the location of a TD by a different driver |
| TD | truly drowsy; corresponds to a lane departure that has been video-coded as drowsy. |
| timeliness | The ability of an algorithm to predict impairment soon after it appears. |
| TLC | time to lane crossing |
| true negative | algorithm classifies as negative (normal), and the ground truth is negative (normal) |
| true positive | algorithm classifies as positive (impaired), and the ground truth is positive (impaired) |
| visual alert | white, yellow, or red icon issued by the DVI as part of a drowsiness mitigation system and requiring the driver to press a button to dismiss. |
| warning | stimulus from a DVI intended to warn the driver; see also *alert* |
| warning type | discrete or staged; a discrete warning is a binary classification; a staged warning is a series of binary classifications that have increasing urgency. |

## Executive Summary

This study demonstrated the potential short-term impact of vehicle-based drowsiness mitigation. Results show that drowsy drivers who received mitigation alerts maintained better vehicle control and had fewer drowsy lane departures than drowsy drivers without this mitigation. Additionally, drowsy drivers with mitigation showed less variability in speed maintenance. Furthermore, staged alerts may improve driver lane keeping as compared to discrete alerts for very drowsy drivers. Finally, alert modality did not affect driving performance, nor did the alerts significantly lower subjective drowsiness.

This report is part of a larger Driver Monitoring of Inattention and Impairment Using Vehicle Equipment research program known as DrIIVE aimed at examining driver impairment. Phase 1 of the program developed and tested a set of algorithms to detect drowsy driving (Brown et al., 2014). Phase 2 focused on improving algorithm performance and assessing its impacts. Specifically, the DrIIVE detection track aims to develop and evaluate a system of algorithms to identify signatures of impairment caused by alcohol-impaired, drowsy, and distracted driving. This report explores potential mitigations to those impairments.

To evaluate the potential mitigation strategies, 48 participants 21 to 34 years old completed two drives in the National Advanced Driving Simulator during two visits. One drive occurred between 10 p.m. and 2 a.m. (early night) and one occurred between 2 a.m. and 6 a.m. (late night). During both drives, participants experienced one of six drowsiness alert interfaces, which varied along two dimensions, mitigation type (discrete, staged) and alert modality (audio/visual, haptic, combined audio/visual and haptic). All drowsiness mitigation interfaces were based on the same drowsiness detection algorithm, which was developed using data from Phase 1. Participants drove a nighttime drive scenario for a total drive time of approximately 35 minutes. The drives started with an urban segment including a two-lane roadway with posted speed limits of 25 to 45 mph and signal-controlled and uncontrolled intersections. An interstate segment followed that included a four-lane divided expressway with a 70-mph posted speed limit. The drive continued with a rural segment that included a two-lane undivided road with curves, ending in a 10-minute section of straight rural roadway. This drive was identical to the phase 1 drive, with the exception that drowsiness alerts were presented to the driver.

*This page is intentionally left blank.*

# Introduction

## Background

The National Highway Traffic Safety Administration reports that 684 fatalities, or 1.6 percent of total fatalities, in 2021 involved drowsy drivers (Stewart, 2023). Vehicle sensors may be able to help detect and support warnings to drowsy drivers. One challenge in using vehicle-based sensors for drowsiness detection lies in the creation of algorithms that can identify signatures of drowsy driving across a variety of individual driving styles and a multitude of roadway environments.

Two main criteria for algorithm performance were used throughout this project. The first compared a whole family of algorithms obtained by varying one of its parameters, such as its correct detections, and determining performance on the other, such as its false positive results. The family of algorithms consists of all pairs of these parameters. The receiver operator characteristic is a way of analyzing such a family of algorithms and provides a simple visualization of performance (Brown & Davis, 2006). The second criteria described the timeliness of the algorithm, or its ability to issue a prediction of impairment soon before driving performance is adversely affected.

This is part of a project study where each phase builds on the findings of previous projects, one that focused on the detection and mitigation of distracted driving, one that focused on detecting alcohol-impaired driving, one that focused on detecting drowsy driving (phase 1), and the current project aims to detect all three driving impairments using a system of algorithms and explores the effect of alerts for drowsy drivers. The phase 1 drowsiness simulator drives occurred during the day, early night, and late night (Brown et al., 2014). Using drowsiness data collected in a simulator during phase 1, several metrics for drowsiness ground truth were explored and a variety of vehicle-based algorithms were developed. Ultimately, a random forest algorithm proved useful for predicting drowsiness 6 seconds in advance of drowsy lane departures (Brown et al., 2014). These results motivated the phase 2 project to improve the drowsiness algorithm, implement a real-time version along with countermeasures, and collect new drowsy-driving data in the simulator to compare different types of mitigation alerts.

## Roadway Environment

Driver drowsiness is affected by both endogenous (internal) and exogenous (external) factors. Circadian variation, time-on-task, and lack of sleep are considered endogenous factors, while characteristics of the vehicle and driving environment represent exogenous factors (Thiffault & Bergeron, 2003). Monotonous scenery in theory promotes drowsiness through under-stimulation, resulting in low levels of arousal. Similarly, straight road conditions are more likely to induce driver drowsiness than are curved roads (Oron-Gilad & Ronen, 2007).

Overall, research suggests that the driving situations most likely to induce drowsiness and its harmful effects are long stretches of low demand driving in predictable driving environments that require few vehicle control actions. A scenario with a long rural straightaway, little interaction with other traffic, no curves, and few changes in the surrounding environment (e.g., minimal signage) is consistent with these criteria. Such situations that occur toward the end of a drive are also likely to induce higher levels of driver drowsiness and fatigue because of sleep deprivation and time-on-task (Åkerstedt et al., 2010).

## Drowsy Driving Performance

A review by Liu et al. (2009) summarized the effects of drowsiness and fatigue on driving performance measures. The review included 17 studies published in peer-reviewed journals in which at least one objective vehicle-based measure was reported. Overall, the review showed an increase in lane departures with higher levels of drowsiness. Moreover, the average standard deviation of lane position, mean absolute value of steering wheel angle, and standard deviation of steering wheel movements increased with higher levels of drowsiness. Sleep deprivation was also associated with greater variability in driving speed.

Many researchers have found an increase in SDLP with increased drowsiness. Arnedtet al. (2001) found that hours of wakefulness predicted changes in SDLP. Specifically, 22 hours of wakefulness resulted in SDLP that was consistent with impaired performance at .08 g/dL blood alcohol concentrations. Using a time-on-task approach and partial sleep deprivation, Otmani et al. (2005) found that right-lane line crossings increased with partial sleep deprivation compared to normal sleep, and SDLP increased over the course of a 90-minute drive. De Valck and Cluydts (2001) showed that SDLP was sensitive to both time in bed and caffeine: SDLP increased with less time in bed and decreased after using caffeine. Importantly, SDLP has also been affected by alcohol intoxication and distraction, as documented in NHTSA's Impairment Monitoring to Promote Avoidance of Crashes using Technology program (Lee et al., 2010) and NHTSA's Distraction Detection and Mitigation through Driver Feedback program (Lee et al., 2013).

Research has also documented an increase in lane departures (i.e., inappropriate line crossings) with drowsy driving. Philip et al. (2005) found that the number of inappropriate line crossings, defined as crossing one of the lateral highway lane markers, increased for sleep-deprived drivers compared to well-rested drivers. Drowsiness-related differences have also been shown for speed control. Arnedt et al. (2001) found greater speed variability after 20 hours of wakefulness compared to after 16 hours. Importantly, these drowsiness-related performance decrements have been associated with increased crash risk.

These performance decrements, in combination with the rates of drowsy driving and association with crashes, injuries, and fatalities, demonstrate the potential to develop technological approaches to detect and mitigate driver drowsiness.

## Drowsiness Detection and Mitigation

Drowsiness-warning systems consist of a drowsiness-detection system (algorithm) and a driver-feedback system (mitigation). The drowsiness-detection algorithm collects data from the driver and/or vehicle, processes this data with a detection algorithm, and makes predictions about the wakefulness of the driver. The system then activates the drowsiness mitigation when the detection system algorithm predicts that the driver is drowsy.

Drowsiness detection algorithms can be characterized by their input data, detection algorithms, and ground truth definitions. Input data may consist of camera-based eye measures (e.g., Dinges & Grace, 1998), electric potential measures from the brain (e.g., Lin et al., 2005), or driver input to the vehicle such as steering wheel input (McDonald et al., 2014). Several types of algorithms may be used including many machine learning techniques. One development in the detection of drowsiness is a transition from static prediction algorithms to time-based prediction algorithm (Ji et al., 2006; Yang et al., 2010; Yang et al., 2009). These time-based algorithms allow predictions to account for well-understood temporal effects of drowsiness, for example a drowsy driver is more likely to stay drowsy and a wakeful driver is more likely to stay awake. Clinical electroencephalogram records scored by a sleep expert are highly accurate measurements of drowsiness that are sometime argued to be ground truth. However, ground truth is sometimes specified less stringently with measures such as hours of wakefulness, time of day and Stanford Sleepiness Scale scores.

The mitigation system links the detection system and the driver. While the detection system aims to accurately assess driver state, the aim of a mitigation system may be to present driver state information to the driver in a way that changes behavior, either in the short term (e.g., less likelihood of departing the lane) or the long term (e.g., pull over to rest). Mitigation systems, or countermeasures, can take many forms, from a simple audible chime or visual icon to more complex displays that relay more elaborate information to the driver. Although the same algorithm might be used across systems, the type of interface may influence driver response.

There are several drowsiness warning systems on the market currently. These systems vary along several dimensions; among these are mitigation type (continuous/staged or discrete) and alert modality. Some original equipment manufacturers' drowsiness-mitigation systems use binary alerts that display a coffee cup icon and play a chime when the alert is triggered. Throughout this report, the term discrete is used to refer to binary warning systems. Other systems provide a more continuous, or at least multi-level (staged), scale of drowsiness to the driver. The impact of these different interfaces is largely unknown.

## Research Questions

The goal of the present study examined the impact of drowsiness mitigations. Specific research questions included these.

- Does drowsiness mitigation reduce drowsy lane departures and reduce SDLP relative to a no-mitigation baseline?
- How do warning type and alert modality influence the effect of drowsiness mitigation relative to a no-mitigation baseline?
- Does drowsiness mitigation reduce subjective drowsiness?

*This page is intentionally left blank.*

# Drowsiness Algorithm Augmentation

## Introduction

Phase 1 of the DrIIVE program developed and tested a set of algorithms to detect drowsy driving (Brown et al., 2014). Phase 2 focused on improving algorithm performance by improving specificity (i.e., reducing the occurrence of false positives), implementing a real-time working version, and designing and implementing a mitigation system. The final algorithm and mitigation were integrated into the NADS simulator as part of a simulator study to evaluate drowsiness mitigation and alert types.

## Phase 1 Review

DrIIVE phase 1 work included several analyses of data collected at three time points (9 a.m. to 1 p.m., 10 p.m. to 2 a.m., 2 a.m. to 6.a.m) for each participant to address research questions concerning drowsiness, specifically these.

1. What kinds of ground truth definitions, performance measures, and algorithms accurately classify drowsiness?
2. Can intoxication and drowsiness be differentiated in the alcohol dataset?
3. Are distraction or alcohol algorithms sensitive to drowsiness?

Algorithm performance is most assessed by comparing its output classification to the actual state of the driver. The driver's actual state is sometimes referred to as ground truth and is ideally shown by a "gold standard" measure that provides an unambiguous gauge of driver state. Such a standard is difficult to define for drowsiness; but arguably a clinical EEG recording scored by a sleep expert would be a potential standard for drowsiness. Since it was not possible to obtain this level of ground truth, several other measures were explored in phase 1. These included pre- and post-drive subjective scores on the SSS, pre- and post-drive psychomotor-vigilance test scores, and retrospective sleepiness scale ratings (Brown et al., 2014).

Fine-grain drowsiness ground truth within a drive was obtained by video-coding lane departures to show whether the driver was or was not drowsy based on an objective rating of drowsiness using the method developed by Wierwille and Ellsworth (1994), and location-matching periods of awake driving. These two types of ground truth markers were denoted as "truly drowsy" cases and "truly awake" cases. Thus, there are TA and TD ground truth markers at common locations throughout the scenarios. The final evaluation dataset consisted of 578 total cases, with 162 drowsy instances, 80 awake cases matched to driver and road segment, and 336 awake cases matched to only road segment. The time periods in between TD and TA cases were unallocated to either state or might reasonably have been either one.

Additionally, several types of algorithm and measures including EEG, PERCLOS, PERCLOS+, Bayesian networks, and random forests were tested. The best-performing algorithm was a random forest trained on raw steering signals collected over a rolling 1-minute interval that was able to issue drowsy predictions 6 seconds in advance of drowsy lane departures.

As might be expected, algorithms that were designed for distraction and intoxication were not sensitive to drowsiness. A simple Bayesian network algorithm could distinguish intoxicated driving from a combination of intoxicated and drowsy driving.

## Algorithm Framework

There are several ways a system of algorithms might detect, differentiate, and mitigate driver impairment. One approach is a layered framework, shown in Figure 1. The bottom three layers relate to the sensors that collect data from the vehicle, and potentially from the driver and environment, that are inputs to the algorithm. Next, the algorithm layer combines these inputs to estimate the driver state. With an algorithm for each type of impairment, it is possible that two or even all three algorithms could show impairment at the same time. Therefore, an arbitration layer may determine the best way to mitigate the detected impairments.

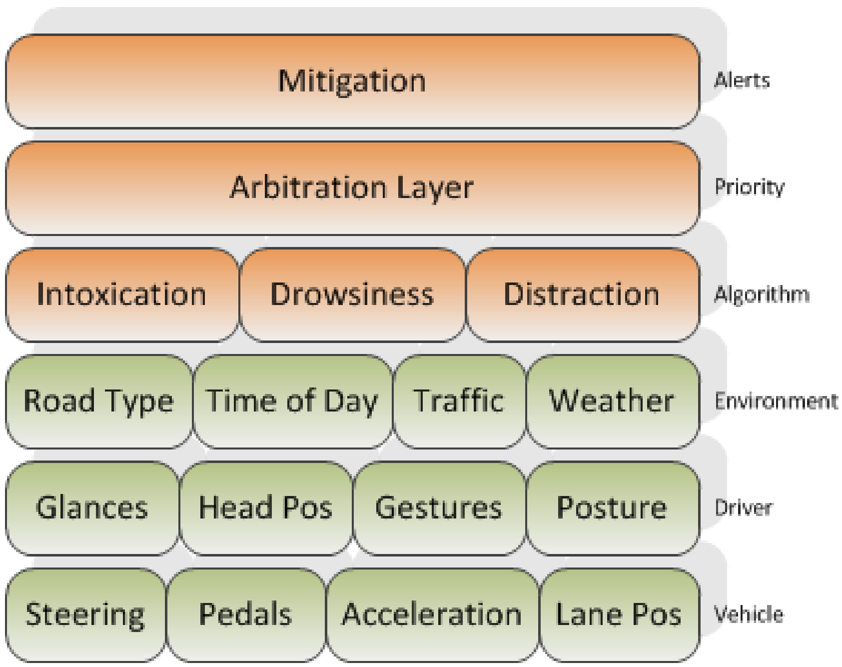



*Figure 1. A layered framework for detecting, differentiating and mitigating impairments*

The impairment detection algorithms had two components. First, a random forest algorithm (Breiman, 2001) classified time windows of input data as impaired or not impaired. This ensemble method works by training many individual decision trees, each with a different sample of data, different subsets of features, thresholds, and different branching conditions.

Each decision tree predicts the driver state, and the ensemble of these predictions result in a driver state output according to a majority vote (see Figure 2).

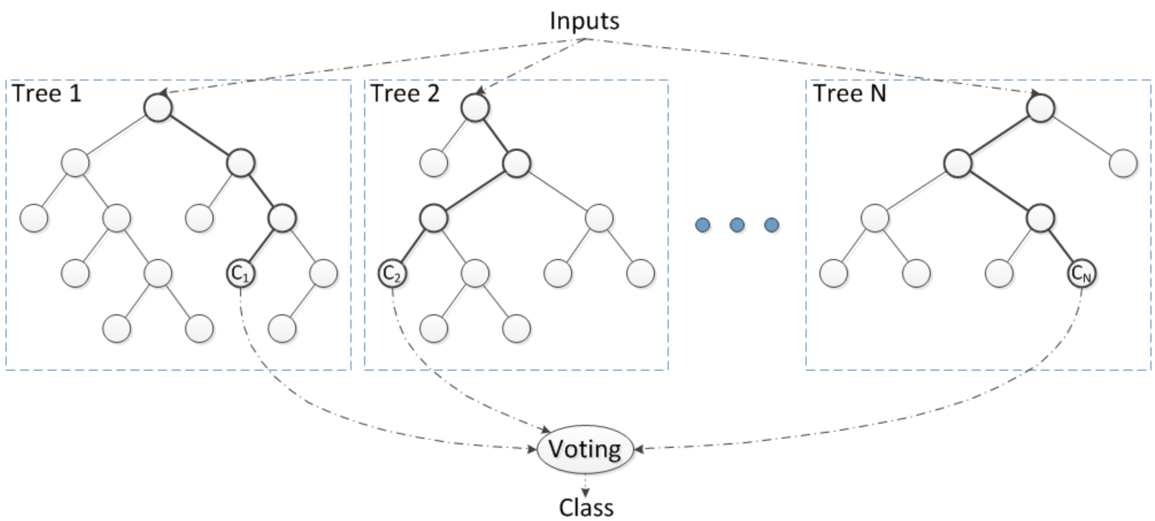



*Figure 2. Random forest machine learning model*

The second component of the algorithm used a hidden Markov model that combined the random forest observations over a brief time to augment the driver state classification by assigning probabilities regarding whether the system remained in a given state or transitioned to a new state at each moment in time (Rabiner, 1989). The HMM uses state history to predict the current driver's state. This history is particularly important in predicting driver state assuming that a driver who was drowsy, for example, is likely to be drowsy in the near future. Distraction depends on the shortest history and alcohol depending on the longest history. Figure 3 shows how the HMM combines a series of inputs, (the classifications of the random forest) to better predict driver state in a way that reflects how the current state depends on past states.

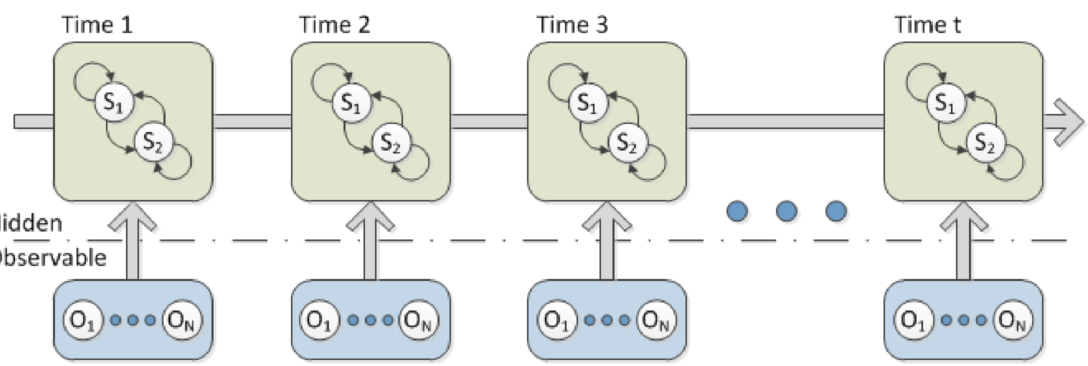



*Figure 3. Hidden Markov model*

## Algorithm Refinement

The phase 1 drowsiness-detection algorithm used continuous steering data as a baseline for refinement and augmentation in phase 2. The ROC curve for this algorithm is shown in Figure 4. It had an area under an ROC curve of 0.67 with 95 percent CI of [0.62, 0.72] when judged by its ability to estimate drowsiness 6 seconds before a drowsy lane departure.

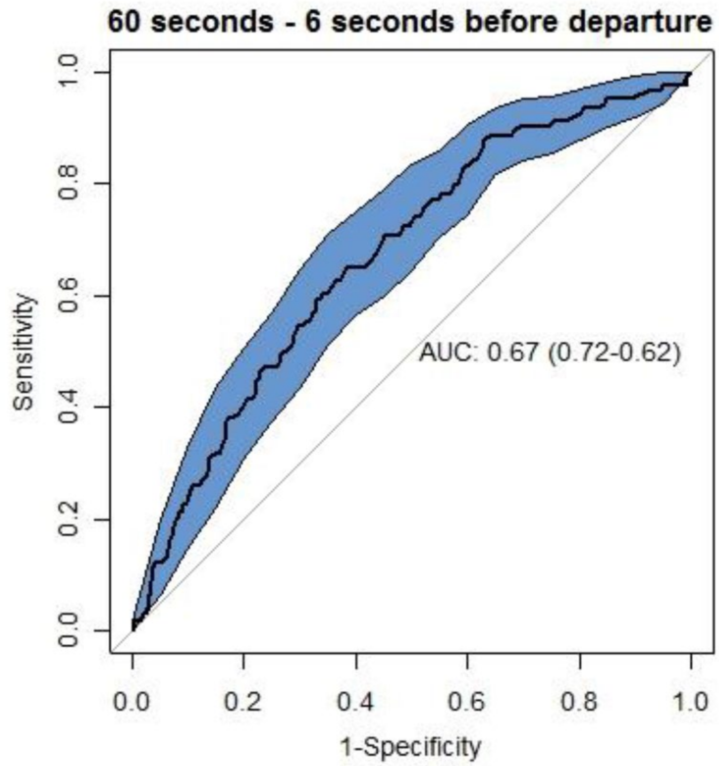



*Figure 4. ROC curve for predictive model of drowsiness related lane departure, using only continuous steering data with a moving window of 60 seconds (Brown et al., 2014)*

In the current effort, several methods of augmenting the random forest algorithm from phase 1 were explored (Brown et al., 2014). That algorithm used 54 sequential samples of the steering wheel angle. The steering segment would advance as time progressed in the drive. That method differentiated between different trends in the steering signal but was vulnerable to steering trends that were the result of geographic features like curves and turns. In phase 2 the use of descriptive statistics on the steering time window (mean, median, standard deviation, max, and min) rather than using raw sequential samples were evaluated. This added the benefit of reducing the number of inputs from 54 to five. Additionally, geographically biased steering trends were filtered out of the steering signal.

Phase 2 development included testing other inputs to the algorithm beyond steering wheel angle. Other low-cost sensors provided inputs from steering and pedals, direct measures of driver behavior. A slightly more advanced option is a forward-facing camera that provided information on lateral lane position. This type of sensor could be used to detect lane departures and calculate measures such as time-to-lane crossing. In place of the forward-facing camera, these measures were calculated from the simulator data. Finally, some driver-based sensor signals from an eye tracker were added for the purpose of comparison. PERCLOS calculated in a 3-minute interval was included to provide the most direct gauge of drowsiness.

Algorithmically, the greatest innovation in phase 2 was the addition of an HMM stage. Random forests were used as inputs to an HMM that inferred driver state. This approach benefitted from melding the predictions of algorithms and considering previous state values. From an exploration of many options, a final algorithm was selected and implemented in real-time using a combination of C++ and Mathworks Simulink. This section reports more details of these refinements to the drowsiness-detection algorithm.

### *Steering Wheel Angle Bias*

Common driving maneuvers like turning, changing lanes, and driving on curves raise a concern with using steering wheel angle as an input to the algorithm. A trend in the steering signal may be caused by an artifact from the driving environment, such as the gradual appearance of steering adjustments (a signal bias) on a curve (Eskandarian et al., 2007). These trends can be characterized by their lower signal frequency content. On one hand, the presence of these steering trends might confound the training of an algorithm, causing it to detect biased variations in the trend rather than informative steering signal content. On the other hand, the process of removing trends from the steering signal could potentially strip vital information from the signal, causing driver state classification to suffer.

A subtraction of the mean steering angle over the length of a curve was the method employed by Eskandarian et al. (2007). This approach is not appropriate for real-time implementation because it uses steering samples that would be collected in the future. Alternatively, road curvature from a geographic information system database, such as from a navigation system, might be used. However, drivers do not always follow the curvature of the road, especially when entering or exiting a curve, so GIS data also does not offer a perfect solution.

An approach was used that filtered out low steering frequencies at the scale of geographic features. An attractive choice for this purpose would have an easily adaptable cut-off frequency to match curves and turns with varying radii of curvature. Brown's method of double exponential smoothing (DES) (LaViola, 2003) provided one such filter, as the frequency cutoff could be easily set (and varied) using a single parameter. A scheme to adapt this parameter for gentle curves as well as tight turns was developed and tested on phase 1 simulator data (Brown et al., 2014). The steering signal was first low pass filtered to remove noise. This filtered signal, as well as steering wheel angle de-trended using the DES filter are shown in Figure 5. The steering signals for four types of curves in the figure show the effect of the DES filter in removing steering bias, though the short duration and low radius of curvature in the left turn make that case particularly challenging.

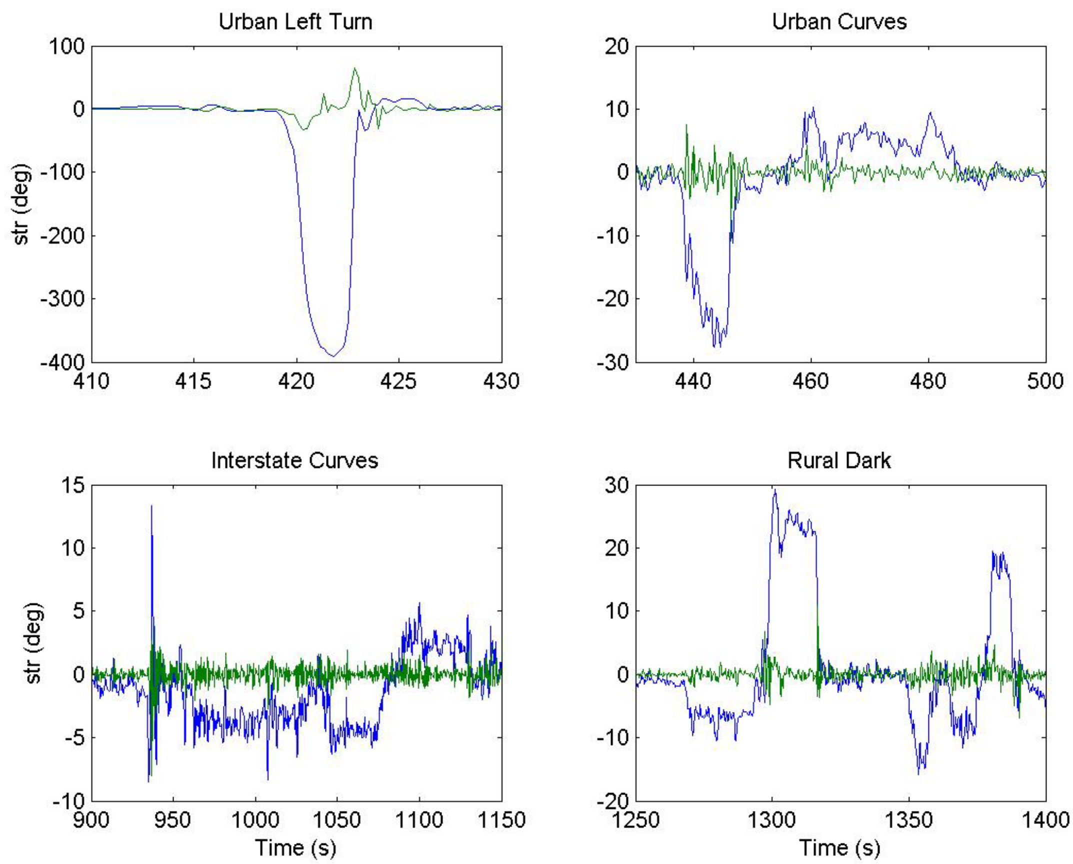



*Figure 5. Original (blue) and de-trended (green) steering wheel angle for four roadway types: A left turn (upper left), curved roads in urban setting (upper right), curved roads on highway (lower left) and curved roads in rural setting (lower right)*

### *Steering Wheel Angle Energy*

A measure that can be called energy is calculated from the de-trended steering wheel angle over a limited window of time. While not the same as the concept of energy in physics, the notion of signal energy is expressed similarly (Gabel & Roberts, 1980). The mathematical notion of energy is simply the sum of the squares of each time instant of a signal. The formula for energy, *E*, over a fixed time window of length *T* for a signal $u_t$ is given by

$$E = \sum_{1}^{T} (u_t)^2$$

Energy was also calculated for steering wheel angle rate, the derivative of steering angle.

### *Real-Time Implementation*

The ultimate deployment scenario for a drowsy-driving algorithm involves integration into a vehicle's hardware and software, and an implementation that can process sensor inputs and make driver state estimates in real-time. In the present study, the framework of random forest plus HMM for a drowsiness-detection algorithm was implemented in a combination of C++ and

Simulink and used to feed a mitigation system for a new drowsy driving data collection. Some general issues regarding real-time implementation of this framework are discussed here.

Running the entire random forest at each time step was too computationally intensive to consider, instead the driver state was updated only every 6 seconds. A C++ implementation of the algorithm called by a Simulink model easily met this requirement. In fact, the update rate could have been reduced to around once every half a second as implemented on a desktop computer.

The HMM implementation was quite simple. Indeed, one attraction of the method is that it can be realized as a system of linear equations. The complexity of computing HMM outputs relate to the "prediction" method used and the time history of interest. The hardest question to answer regarding HMMs is: what is the most likely state at *each* time in an interval of some length, $T$? This question is answered using the Viterbi algorithm and involves processing over the interval both forward and backwards in time. The computational load increases with the interval length, $T$.

Typically, only the value of the most recent state is of interest. Thus, a more suitable approach for real-time implementation is to apply smoothing or filtering to the HMM to estimate the posterior probability of the impaired state at time $T$. As a practical consideration, we are unable to consider an infinite time horizon, so a fixed interval size is selected and the HMM is computed on the running interval at the same update rate as the random forest. The authors fixed the time horizon at 5 minutes for the HMM algorithm.

## Algorithm Results

### *Episode Resolution*

Is there a natural resolution in the data that matches episodes of drowsy driving? The aggregation time period of the training samples was varied to explore this question. Four time periods were tested: 30, 60, 90, and 120 seconds. Liang, Reyes, and Lee (2007) considered similar window sizes in an investigation of cognitive distraction.

A 60-second time segment performed as well as, or better than, other time segment choices. It aligned well with the procedure for calculating ground truth instances by reviewing 60 seconds of video prior to each lane departure. A 60-second time window was small enough to pick up the effects of acute drowsiness episodes such as micro sleeps, yet large enough to smooth out micro-variations in driving performance across time segments and between drivers. Aggregation windows of 60 seconds were selected for use in the mitigation track algorithms.

### *Random Forest Algorithms*

After settling on aggregation over a 60-second running window, the next step in algorithm development was to develop a group of candidate algorithms that used a variety of inputs. The candidate algorithms (see Table 1) were composed of static algorithms that operated on 60 seconds of data (54 for model A, 3 minutes for PERCLOS) but did not depend on past windows.

*Table 1. Candidate drowsiness algorithms*

| Model | Type | Composition |
|---|---|---|
| **A** | Random Forest | Consecutive steering wheel samples, aggregated each second |
| **B** | Random Forest | Statistics on de-trended steer, de-trended steer rate, de-trended steer energy, de-trended steer rate energy (min, max, mean, median, SD) |
| **C** | Random Forest | Statistics on throttle, brake (min, max, mean, median, SD) |
| **D** | Random Forest | Statistics on TLC (min, SD), SDLP (SD), lane departure period (max) |
| **E** | Random Forest | Statistics on Seeing Machines PERCLOS and NADS PERCLOS (min, max, mean, median, SD) |
| **F** | PERCLOS | Seeing Machines PERCLOS |

Algorithms A to E are based on random forests. Algorithms A to C use vehicle-based, low-cost sensor packages, with algorithms A and B focused particularly on steering and algorithm C focused on throttle and brake inputs. Algorithm D requires a forward-facing camera. Algorithms E and F are driver-based, requiring a driver-facing camera system.

Algorithm A is the only one that uses consecutive samples culled from the sensor stream. Algorithms B through E use descriptive statistics calculated over the desired time segment. Algorithm E applies descriptive statistics to the PERCLOS variables and feeds those into a random forest. Algorithm F is a simple threshold on the value of PERCLOS (Dinges & Grace, 1998) reported by the eye tracker system. An alternate version of PERCLOS, NADS PERCLOS, is included due to concerns about the quality of the FaceLab measure based on its poor phase 1 performance in predicting SSS scores (Brown et al., 2014).

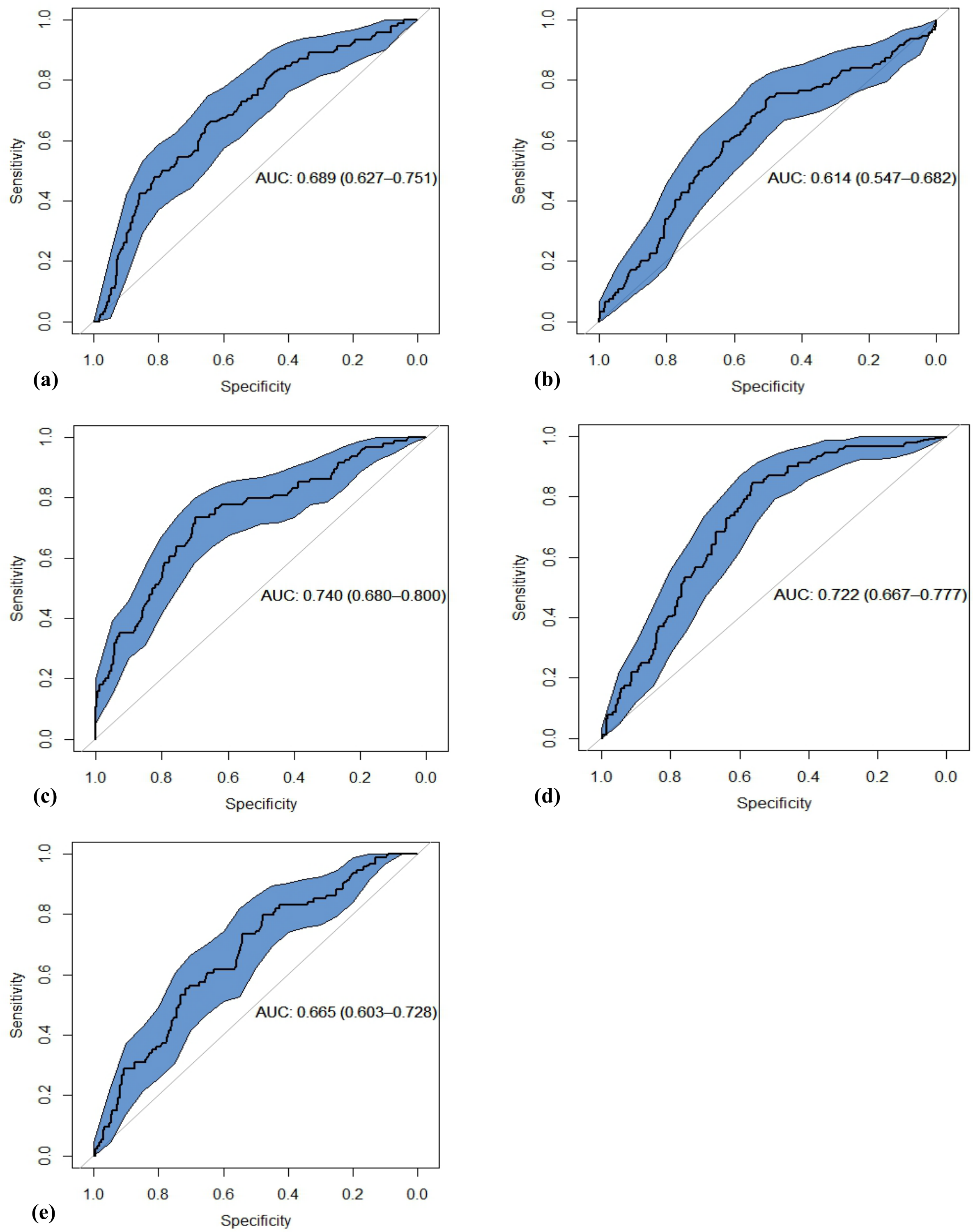



*Figure 6. Random forest ROC curves using vote threshold (1–500) as the continuous predictor. Subplot: (a) Algorithm A, (b) algorithm B, (c) algorithm C, (d) algorithm D, (e) algorithm E*

The caret package (Kuhn, 2008) in the statistical software programming language R is a way to train and test random forest algorithms. This package unifies many different machine-learning approaches into a common framework. The random forest algorithm in caret trains 500 decision trees and the raw vote counts are available in addition to the ultimate classification. Normally, a classification is inferred from a random forest by running all the decision trees and using the majority vote classification as the output. However, if one keeps track of the vote count, then it can be used as the continuous predictor in a ROC analysis. Then, an optimal threshold on the vote count may be computed from the ROC curve, referred to as the operating point. The ROC curves for each random forest are shown in Figure 6. Note that algorithm F does not have an ROC curve since PERCLOS is defined as a single algorithm and not a parameterized family.

The area under the ROC curve and common statistics calculated from the confusion matrix are summarized in Table 2. The operating point on the ROC curve may be selected using the Youden index, which is calculated as sensitivity plus specificity minus one. The table shows the values for votes, accuracy, sensitivity, specificity, and positive predictive value at the operating point. A consequence of selecting a vote threshold of less than 250 of 500 decision trees is that the algorithm might classify a driving interval as drowsy even if most decision trees show awake. With majority rule, one may stop computing decision trees after the majority is found. Here, it is important to calculate all the trees as relatively few are needed for a drowsy classification and tree number 500 might cast the deciding vote.

Intuitively, a higher vote threshold for classifying drowsiness requires greater agreement among the individual trees and therefore reduces the number of drowsy predictions and thus, the sensitivity. Observe from Table 2 that all five models have optimal vote thresholds smaller than the majority requirement of 250 showing agreement that the interval did not demonstrate drowsy driving. The 95 percent confidence interval on the AUC is obtained by bootstrapping 500 times using the built-in caret functionality; and the 95 percent confidence interval for the confusion matrix statistics is obtained using the Wilson confidence score interval (Wallis, 2013).

*Table 2. Random forest algorithm statistics*

| | **A** | **B** | **C** | **D** | **E** |
|---|---|---|---|---|---|
| **Votes** | 136 | 116 | 121 | 100 | 85 |
| **AUC** | 0.69 (0.63,0.75) | 0.61 (0.55,0.68) | 0.74 (0.68,0.80) | 0.72 (0.67,0.78) | 0.67 (0.60,0.73) |
| **Accuracy** | 0.62 (0.61,0.63) | 0.63 (0.62,0.64) | 0.69 (0.67,0.70) | 0.60 (0.58,0.61) | 0.63 (0.62,0.65) |
| **Sensitivity** | 0.66 (0.63,0.69) | 0.84 (0.82,0.87) | 0.78 (0.76,0.81) | 0.83 (0.81,0.86) | 0.91 (0.90,0.91) |
| **Specificity** | 0.61 (0.59,0.62) | 0.56 (0.55,0.58) | 0.65 (0.64,0.67) | 0.52 (0.50,0.54) | 0.55 (0.53,0.56) |
| **PPV** | 0.35 (0.33,0.37) | 0.38 (0.36,0.40) | 0.42 (0.40,0.44) | 0.36 (0.34,0.37) | 0.39 (0.37,0.41) |

### *Hidden Markov Model Algorithms*

The next step in the algorithm development process was to select a model or models from the candidate list in Table 1 and augment the algorithm with an HMM. The progression from top to bottom in Table 1 and left to right in Table 2 represents a shift from low-cost, vehicle-based sensors to a higher-cost forward-facing camera and a driver-facing camera. Since the DrIIVE program focused on vehicle-based sensor technology, models D, E and F were eliminated from consideration and carried forward models A, B and C for further development.

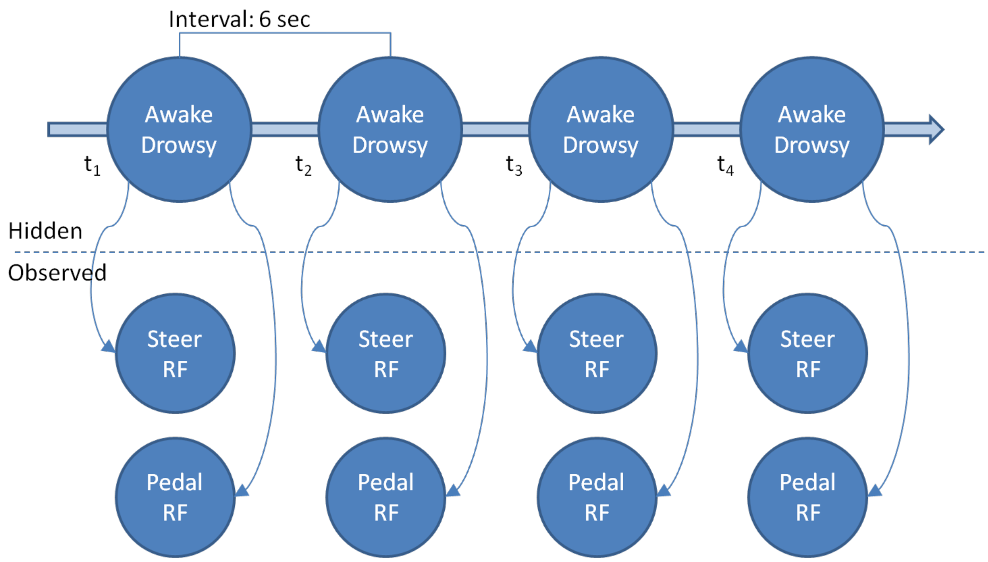



*Figure 7. HMM states and evidence at four time samples using steering and pedal-based random forests as evidence*

A HMM adds an element of driver state history to the algorithm. The hidden state of the HMM is drowsiness. The HMM inputs come from the component random forests. The parameters of the HMM algorithm include the probabilities that it remains in its current state or transitions to a new state given each possible combination of drowsiness classifications from the random forests. The general diagram of an HMM is shown in Figure 3; and a specific version that uses the steering and pedal random forests as inputs is shown in Figure 7.

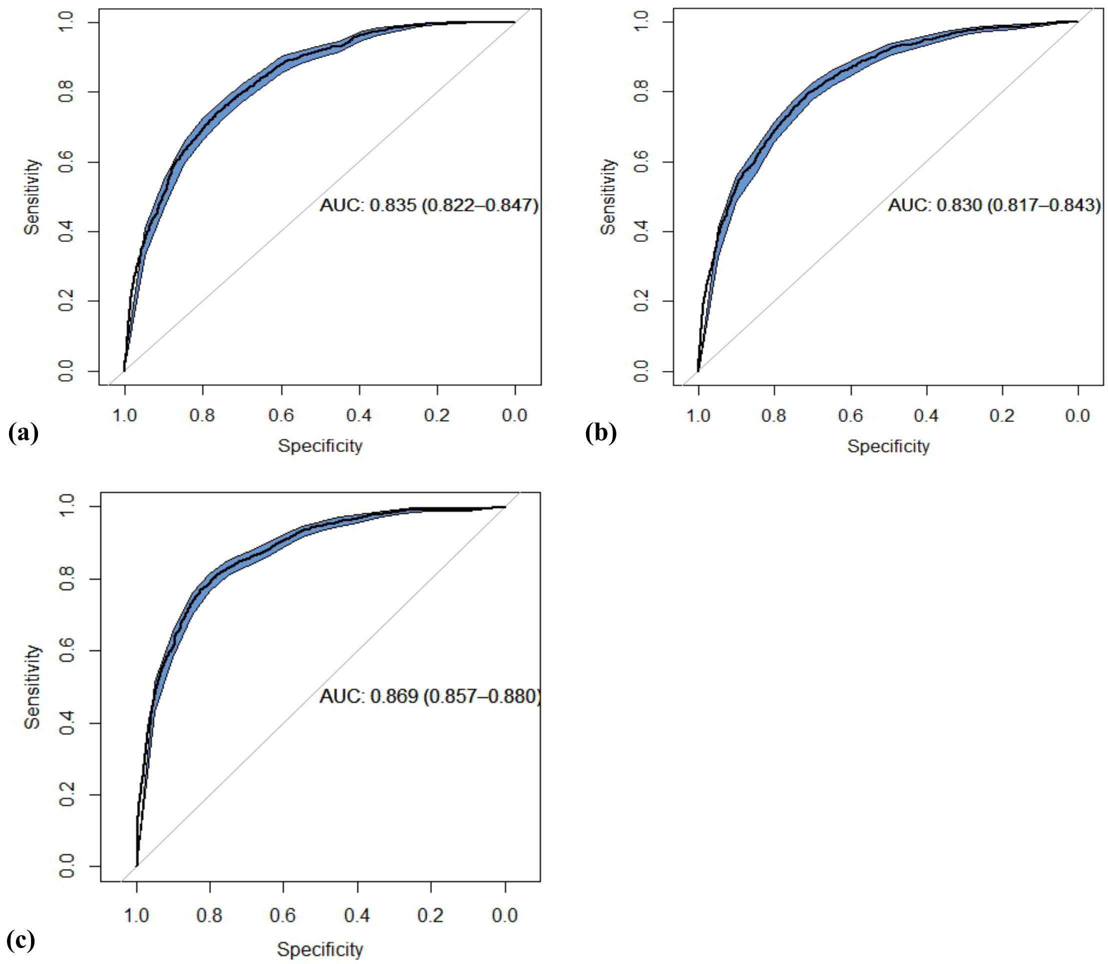



*Figure 8. HMM ROC curves using drowsy probability as the continuous predictor. Subplot: (a) Algorithms A and B, (b) algorithms A and C, (c) algorithms B and C*

Three different HMM-based algorithms were created and tested on the dataset. The three algorithms used different combinations of random forests as their inputs. The three combinations were A and B, A and C, and B and C. It was possible to combine all three models; but some of the nine possible combinations of random forest classifications did not occur, so their likelihood was difficult to estimate. For this reason, only pair combinations were assessed. All time segments in the interstate and rural segments of the drives were used in evaluating the HMMs, not just the ground truth segments. The urban section was excluded due to a lack of available ground truth intervals, due to the relatively small number of urban drowsy lane departures. The final algorithm was set to work only at speeds greater than 40 mph as a method to ensure operation only in the highway and rural segments.

The three time-of-day conditions (day, early night, and late night) were separated and segments from all subjects in each condition were concatenated to make three large driving samples. The random forest algorithms processed steering and pedal statistics in each 6-second time segment and generated inputs for the HMM. The HMM smoothed posterior probability of drowsiness and updated the state prediction every 6 seconds. The delay in running this computation in a real-time setting on a desktop workstation is around 10 milliseconds.

The posterior probability output of the HMM could vary continuously between zero and one and was used as the predictor, or threshold variable, in the family of algorithms for which ROC curves were generated (see Figure 8). Moreover, variations on the state transition probabilities were investigated to learn about the sensitivity of the HMM performance to these values.

*Table 3. HMM using algorithms A and B. Parameter variation on state transition probabilities*

| $p_{D \to A}$ | **0.001** | **0.01** | **0.1** |
|---|---|---|---|
| $p_{A \to D}$ | | **0.001** | |
| **Threshold** | 0.94 | 0.59 | 0.11 |
| **AUC** | 0.83 (0.82,0.84) | 0.83 (0.82,0.75) | 0.84 (0.82,0.85) |
| **Accuracy** | 0.76 (0.75,0.77) | 0.75 (0.74,0.76) | 0.74 (0.73,0.75) |
| **Sensitivity** | 0.72 (0.69,0.74) | 0.76 (0.74,0.79) | 0.80 (0.78,0.82) |
| **Specificity** | 0.77 (0.76,0.79) | 0.74 (0.73,0.76) | 0.72 (0.70,0.73) |
| **PPV** | 0.50 (0.48,0.53) | 0.49 (0.46,0.51) | 0.47 (0.45,0.50) |
| $p_{A \to D}$ | | **0.01** | |
| **Threshold** | 0.97 | 0.83 | 0.38 |
| **AUC** | 0.83 (0.82,0.85) | 0.83 (0.82,0.85) | 0.83 (0.82,0.84) |
| **Accuracy** | 0.76 (0.75,0.77) | 0.75 (0.74,0.76) | 0.73 (0.72,0.75) |
| **Sensitivity** | 0.74 (0.72,0.77) | 0.77 (0.75,0.80) | 0.80 (0.78,0.82) |
| **Specificity** | 0.76 (0.75,0.78) | 0.75 (0.73,0.76) | 0.71 (0.70,0.73) |
| **PPV** | 0.50 (0.48,0.52) | 0.49 (0.47,0.52) | 0.47 (0.45,0.49) |

The state transitions are fully specified by a 2x2 matrix, $p_T$, given by

$$p_T = \begin{bmatrix} p_{A \to A} & p_{A \to D} \\ p_{D \to A} & p_{D \to D} \end{bmatrix},$$

where $p_{A \to A}$ is the probability of remaining in the awake state, $p_{D \to D}$ is the probability of remaining in the drowsy state, and $p_{A \to D}$ and $p_{D \to A}$ are the probabilities of transition from awake to drowsy and drowsy to awake, respectively. Each row of the matrix must sum to one, so if only one probability is specified in each row of the matrix, then the other one is known as well.

*Table 4. HMM using algorithms A and C. Parameter variation on state transition probabilities*

| $p_{D\to A}$ | **0.001** | **0.01** | **0.1** |
|---|---|---|---|
| $p_{A\to D}$ | | **0.001** | |
| **AUC** | 0.83 (0.81,0.84) | 0.83 (0.82,0.84) | 0.83 (0.82,0.84) |
| **Threshold** | 0.43 | 0.18 | 0.05 |
| **Accuracy** | 0.73 (0.72,0.74) | 0.73 (0.72,0.75) | 0.73 (0.72,0.75) |
| **Sensitivity** | 0.78 (0.76,0.81) | 0.80 (0.78,0.82) | 0.80 (0.78,0.82) |
| **Specificity** | 0.72 (0.71,0.73) | 0.71 (0.70,0.73) | 0.71 (0.70,0.73) |
| **PPV** | 0.47 (0.45,0.49) | 0.47 (0.45,0.49) | 0.47 (0.45,0.49) |
| $p_{A\to D}$ | | **0.01** | |
| **AUC** | 0.83 (0.82,0.84) | 0.83 (0.82,0.84) | 0.83 (0.81,0.84) |
| **Threshold** | 0.72 | 0.72 | 0.4 |
| **Accuracy** | 0.73 (0.72,0.75) | 0.75 (0.74,0.77) | 0.76 (0.75,0.77) |
| **Sensitivity** | 0.80 (0.77,0.82) | 0.76 (0.74,0.78) | 0.75 (0.73,0.77) |
| **Specificity** | 0.71 (0.70,0.73) | 0.75 (0.74,0.77) | 0.76 (0.75,0.78) |
| **PPV** | 0.47 (0.45,0.49) | 0.49 (0.47,0.52) | 0.50 (0.48,0.52) |

Table 3 presents statistics from the combination of random forest models A and B. Note that this pairing considers only steering inputs and not pedal inputs. Table 4 summarizes the performance from the combination of models A and C. This combination considers both steering and pedals but uses raw data from the one and summary statistics from the other. Table 5 shows statistics from the combination of models B and C. Aesthetically, this combination is attractive because it combines two models of a similar type and includes both steering and pedal inputs. Based on the results obtained through the combinations of different random forest models and state transition parameter variations, the HMM that was selected that which used models B and C with the state transition matrix given by

$$p_T = \begin{bmatrix} p_{A\to A} & p_{A\to D} \\ p_{D\to A} & p_{D\to D} \end{bmatrix} = \begin{bmatrix} 0.999 & 0.001 \\ 0.01 & 0.99 \end{bmatrix}.$$

As described below, other transmission matrix options performed similarly.

The corresponding statistics from Table 5 are highlighted in green and surrounded with a heavy border. This result is a clear improvement over the random forest from phase 1 that operated on raw steering samples (Brown et al., 2014).

The low PPV is a cause for some concern. PPV gives the probability that drowsy algorithm estimates coincide with drowsy lane departures. There are two main reasons why PPV is so low throughout the results. First, the algorithm is expected to register drowsiness in advance of the drowsy lane departure, but PPV penalizes that kind of prediction. Second, there are several periods of detected drowsiness outside of TD and TA cases in Figure 9 (e.g., the first 100 seconds of the early night subplot). It is not possible to fully judge the correctness of such classifications in the absence of reliable ground truth; but, once again, PPV penalizes all such periods of drowsiness classification.

*Table 5. HMM using algorithms B and C. Parameter variation on state transition probabilities*

| $p_{D \to A}$ | 0.001 | 0.01 | 0.1 |
|---|---|---|---|
| $p_{A \to D}$ | | 0.001 | |
| **AUC** | 0.87 (0.86,0.88) | 0.87 (0.86,0.88) | 0.87 (0.86,0.88) |
| **Threshold** | 0.88 | 0.74 | 0.26 |
| **Accuracy** | 0.80 (0.79,0.82) | 0.81 (0.80,0.82) | 0.81 (0.79,0.82) |
| **Sensitivity** | 0.78 (0.76,0.81) | 0.77 (0.75,0.80) | 0.78 (0.75,0.80) |
| **Specificity** | 0.81 (0.80,0.83) | 0.82 (0.81,0.84) | 0.81 (0.80,0.83) |
| **PPV** | 0.57 (0.55,0.60) | 0.58 (0.56,0.61) | 0.57 (0.55,0.60) |
| $p_{A \to D}$ | | 0.01 | |
| **AUC** | 0.87 (0.86,0.88) | 0.87 (0.86,0.88) | 0.86 (0.85,0.87) |
| **Threshold** | 0.97 | 0.76 | 0.24 |
| **Accuracy** | 0.81 (0.80,0.82) | 0.79 (0.78,0.81) | 0.77 (0.76,0.78) |
| **Sensitivity** | 0.77 (0.75,0.80) | 0.80 (0.77,0.82) | 0.81 (0.79,0.83) |
| **Specificity** | 0.82 (0.81,0.84) | 0.79 (0.78,0.81) | 0.76 (0.74,0.77) |
| **PPV** | 0.58 (0.56,0.61) | 0.55 (0.53,0.58) | 0.52 (0.49,0.54) |

Raw output from the selected HMM algorithm is shown in Figure 9. The figure shows 600 seconds of data for each time-of-day condition. The raw data is obtained by combining drives from all participants; however, the plots are limited to only the first 600 seconds of this data. Ground truth values are displayed as dark points on each plot where TA instances have the value 0.0 and TD instances have the value 1.0. Unallocated time frames are not marked in black since those segments have no video coded value. The continuous line is the posterior probability of a "drowsy" classification from the HMM.

Most TA ground truth instances were drawn from the daytime condition in the phase 1 dataset (Brown et al., 2014). It seemed that when TA and TD instances were closely spaced, it was difficult for the model to correctly classify both. Indeed, the model was severely challenged to detect all TD instances while simultaneously *not* misclassifying any TA instances. On the other hand, rising levels of drowsiness prior to TD instances could be observed at several points throughout a drive. This behavior is beneficial and lines up with the previous result of predicting drowsy lane departures 6 seconds ahead of time.

The early night and late-night conditions from phase 1 showed several periods classified as drowsy (Brown et al., 2014). There were very few TA ground truth segments in these two experimental conditions; yet periods of wakefulness between the TD instances may have been present. Given the spacing and density of TD instances in the drives, it was concluded that periods of drowsiness precede and follow dense collections of TD instances and that lengthy periods of unallocated segments may contain periods of awake and drowsy driving that did not result in lane departures, but that may warrant mitigation.

The late-night drives (see Figure 9), showed many more TD instances and more sustained episodes of drowsiness, according to the algorithm. There were also several segments that clearly showed high probabilities of drowsiness in the minutes leading up to a TD instance, showing good timeliness.

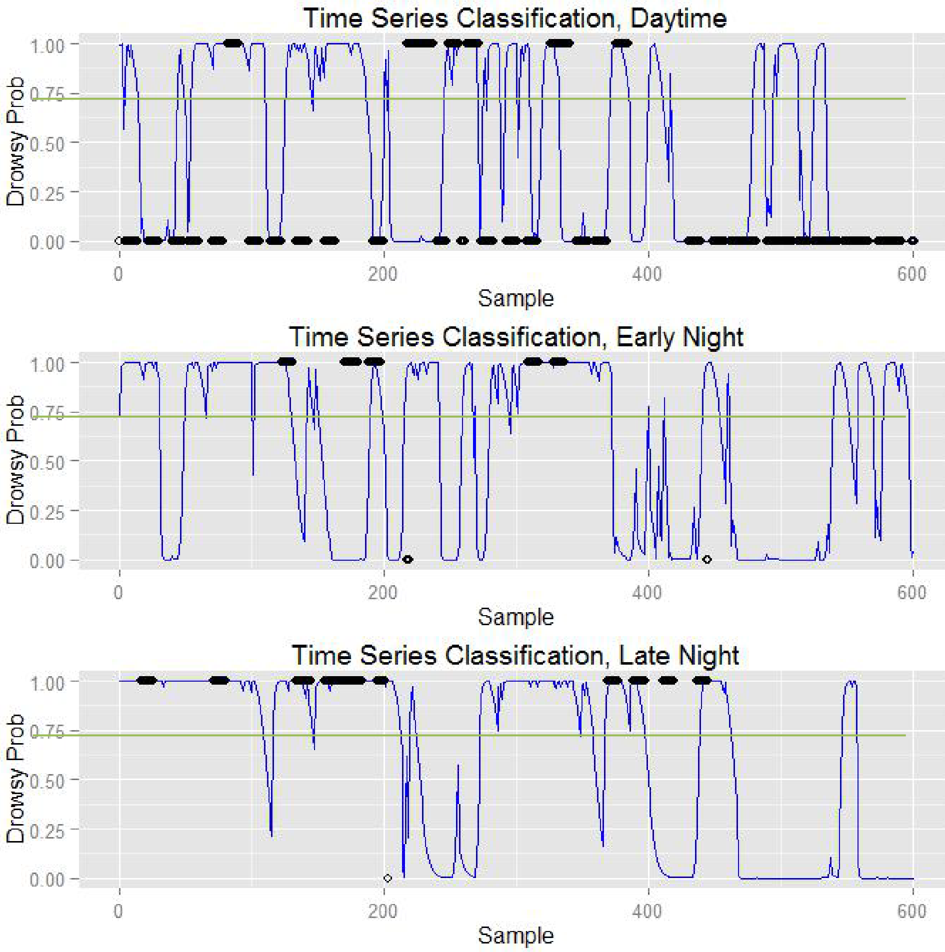



*Figure 9. HMM classification of one participant for HMM using algorithms B and C. The blue line is the algorithm output for one participant. The black points are coded ground truth: Drowsy lane departures or awake driving. A threshold of 0.74 is marked by a horizontal line. There is one sample per second in the figure.*

The predictive property of the phase 1 random forest algorithm falls under the idea of timeliness. Timeliness is the degree to which an algorithm can generate accurate impairment classifications soon after onset of the impairment. A timely drowsiness-detection algorithm can generate an estimate of drowsiness in advance of an adverse event, such as a lane departure. Good timeliness might be expected from an algorithm that acts on data averaged over the previous rolling minute, as is done with the drowsiness algorithm, while one that accumulates data from the previous 30- to 60 minutes would not be as timely because it may take a long time before it can issue an accurate impairment classification.

The basic timeliness of the drowsiness-detection algorithm can be quantified by the number of seconds prior to a ground truth drowsy lane departure (TD) that a drowsy classification is made. This is estimated by counting samples from the classification to the TD case and multiplying by six (since classifications are updated only every 6 seconds). If additional TD annotations are present within the same classification period, they are not counted in the timeliness metric. Additionally, TDs may be present during awake classifications; and these are false negatives. In these cases, the number of samples from the beginning of the previous TD to the beginning of the next drowsiness classification is counted, multiplied by six and negated. Periods of drowsy classification without any TDs nearby are not considered since the actual state of drowsiness is unknown. This may happen if all the adjacent TD cases are already associated with other periods of drowsiness classification. For example, there is such a period in the late night (bottom) subplot of Figure 9 at around 550 seconds (at one sample per second).

Table 6 summarizes statistics of the final algorithm classifications and their timeliness. The percentage of time in all drives where the vehicle is traveling greater than 40 mph that is classified as drowsy is, as expected, larger in the late-night condition, but not that much smaller in the early night condition. That it is as large as 26 percent in the daytime condition is surprising, especially since the daytime data comes from verifiably awake participants. The median duration of the drowsy episodes is also shown. The trend again follows the expected increase from daytime to early night to late night. Finally, the median timeliness shows a trend opposite that of the previous two measures. The set of timeliness measurements for each condition has a distribution that is centered around a low, positive timeliness, but with some large outliers on both the positive and negative sides. The late-night condition shows that the 6 second timeliness is still achieved; but higher values are seen in the early night (12 seconds) and day (24 seconds) conditions. This shows that the algorithm has a propensity to over-classify drowsiness in the day and, to a lesser degree, in the early night.

*Table 6. HMM classification statistics*

| Time of day | % Time classified as drowsy (%) | Median duration of drowsy episodes (sec) | Median timeliness to drowsy lane departure (sec) |
|---|---|---|---|
| **Day** | 26 | 48 | 24 |
| **Early night** | 40 | 54 | 12 |
| **Late night** | 43 | 75 | 6 |

## Conclusions

Several new techniques were applied to improve a steering-based random forest algorithm developed in phase 1 (Brown et al., 2014). By combining classifications from pairs of algorithms that used various types of vehicle sensor input and considering time sequences of those classifications, combinations of random forests and HMM were able to significantly improve classification performance, raising the AUC of the ROC curve from 0.67 to 0.87.

The phase 1 algorithm used raw steering samples performed well when compared to other static models and was the only one that could be classified based on steering trends. On the other hand, algorithms based on steering summary statistics and a de-trended steering signal also performed well, implying that steering trend differences between awake and drowsy drivers may not have been an important feature. Interestingly, the random forest based on throttle and brake performed just as well as the ones based on steering. The family of algorithms based on basic vehicle inputs could provide a good set of classifiers using relatively low-cost sensors.

It is important to consider the pitfalls in the approach taken here, given the limitations of the dataset. Ground truth measures that were taken once or twice per drive, like SSS and PVT, did not result in effective algorithms in phase 1. Video coded TD and TA cases gave a more fine-grained estimate of ground truth but left most time frames of a drive unallocated to either TD or TA. Another potential avenue of investigation is the customization of the algorithm to each driver. Training of Bayesian filters can occur in real-time and benefit from new input from the driver, as is done in some email spam filters. Algorithm individualization is considered in the detection track (Brown et al., 2024).

# Methodology

## Introduction

The goal of the phase 2 data collection was to investigate the effect of drowsiness mitigation across a range of driver-vehicle interface (DVI) characteristics. To that end, six drowsiness mitigation alerts were compared that varied along two dimensions: the type of the mitigation (discrete and staged) and the modality of the warning (audio-visual, haptic, and combined audio-visual plus haptic).

## Data Collection Methods

System evaluation needed detailed data from drowsy drivers in a controlled, yet realistic, situation. Participants drove through representative situations on three types of roadways (urban, freeway, and rural) at two times of day (early night and late night) in a high-fidelity driving simulator. The only difference between this study and phase 1 was the presence of the mitigation system. This allowed data from the 24 phase 1 participants (21 to 34 years old) to be included as a no-mitigation comparison group.

No significant statistical interaction effects on driving performance between age group and drowsiness condition were found in phase 1, phase 2 data collection included only the youngest age group from phase 1 (21 to 34-year-olds). The following sections summarize the data collection methods.

### *Participants*

Forty-eight participants completed the study, recruited from the NADS participant database from a University of Iowa mass email, internet postings, and referrals (see Schmitt & Brown, 2015). They were healthy men and women 21 to 34 (median age 24, mean age 25.3, SD 3.1) with a valid driver's license. They were paid $250 for completing all study sessions. Pro-rated compensation was provided for those who did not complete the study. An initial telephone interview determined eligibility for the study. Applicants were screened for health history and current health status (Schmitt & Brown, 2015) and whether they were a morning or evening person (Adan & Almirall, 1991) (Schmitt & Brown, 2015). Applicants with scores on the morning/evening scale less than 12 out of 30 were not eligible for participation. Additional telephone screening criteria included:

- possession of a valid U.S. driver's license,
- licensed for 2 or more years,
- drove at least 10,000 miles per year,
- no driver's license restrictions except for vision,
- lived within a 30-minute drive to the facility,
- ability to meet study schedule and follow study protocol,
- normal sleep pattern, and
- good general health.

Potential participants who passed the telephone screening scheduled an initial visit to verify eligibility. Participants were screened for drug and alcohol use and for females, pregnancy; cardiovascular health; caffeinated food or beverages use; sleep patterns and use of drugs that effect sleep; and potential simulator sickness. If participants' drug screen was positive, blood pressure (systolic blood pressure = 120 ± 30 mm Hg, diastolic blood pressure = 80 ± 20 mm Hg) and/or heart rate (heart rate = 70 ± 20 bpm) did not meet study criteria, tested positive for alcohol, or a female screened positive for pregnancy, the participants would be excluded from continuing in the study. If following a practice drive in the simulator, the participant reported high scores (Nausea ≥ 21, Oculomoter ≥ 32, Disorientation ≥ 15, or Total ≥ 32) on the simulator sickness questionnaire (Kennedy et al., 1993; Schmitt & Brown, 2015), that person was excluded from the study.

### *Simulator and Apparatus*

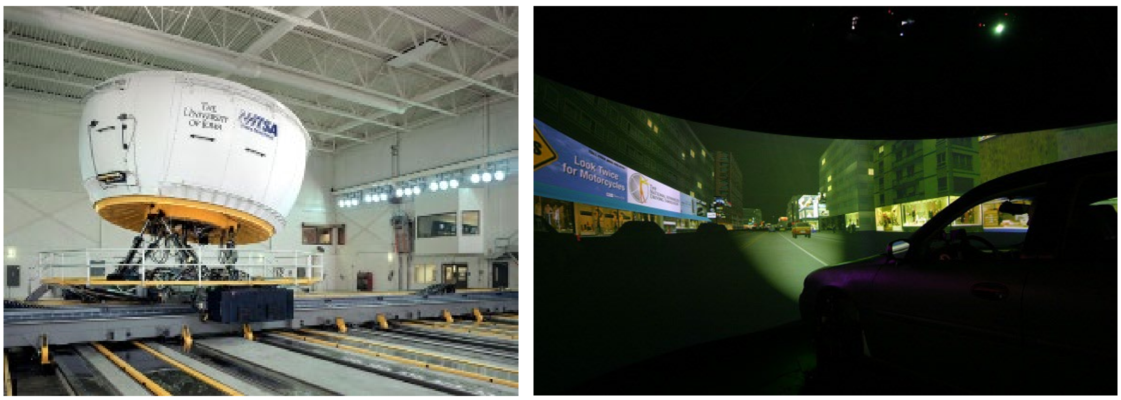


*Figure 10. Images of the NADS-1 driving simulator (left) with a driving scene from inside the dome (right).*

The NADS is located at the University of Iowa's Research Park. It consists of a 24-foot dome in which an entire vehicle is mounted. All participants "drove" the same vehicle, a 1996 Malibu sedan. The motion system on which the dome is mounted provides 400 $m^2$ of horizontal and longitudinal travel and ±330° of rotation. The driver feels acceleration, braking, and steering cues much as if driving a real vehicle. The three front projectors each have a resolution of 1,600 x 1,200; the five rear projectors have a resolution of 1,024 x 768. The edge blending between projectors is 5°horizontal. The NADS produces a thorough record of vehicle state (e.g., lane position) and driver inputs (e.g., steering wheel position), sampled at 240 Hz.

The cab is equipped with a FaceLab 5.0[2] eye tracking system that is mounted on the dash in front of the driver's seat above the steering wheel. The worst-case head-pose accuracy is estimated to be about 5°. In the best case, where the head is motionless and both eyes are visible, a fixated gaze may be measured with a root mean square deviation error of 2°. The eye tracker samples at a rate of 60 Hz. The installation of the cameras in a NADS cab is shown in Figure 11.

[2] FaceLab 5.0, eye tracking system by Seeing Machines, Canerra, Australia.

*Figure 11. FaceLab cameras mounted in the Malibu cab with a separate head tracking system mounted between them*

An Alco-Sensor IV[3] breath alcohol-testing instrument was used to measure participants' BAC. The hand-held sensor uses a fuel cell to determine BAC. The system was approved by the U.S. Department of Transportation for evidential use and exceeds the Federal model specification for traffic enforcement and breath alcohol testing specified by the Omnibus Transportation Employee Testing Act of 1991. The system is designed to measure BACs from .00 g/dL to .40 g/dL with drift of less than .005 g/dL BAC over several months. The system was recalibrated using an approved dry gas standard prior to the start of the study.

A Motionlogger Actigraph[4] measured participants' activity levels to determine when participants last slept. Participants were instructed to wear the monitor, which was like a wristwatch, on their non-dominant wrist for 2 days prior to their overnight visit through check-in at the overnight visit. Participants were given the Motionlogger no less than 24 hours prior to their overnight visit check-in time. The single axis Motionlogger uses a precision piezoelectric bimorph-ceramic cantilevered beam, which generates a voltage each time the actigraph is moved. The Motionlogger's analog circuitry amplifies the voltage and filters it to between 2 and 3 Hz. It stores the accumulated motion data each minute (Schmitt & Brown, 2015).

Using an iPad-based cognitive assessment from Digital Artefacts[5] provided an objective assessment of driver vigilance. This test is based on the psychomotor vigilance task by Wilkinson and Houghton (1982). It presents an infrequent target that the subject must respond to as quickly as possible via a button press. Although the duration of the PVT task is generally 10 minutes, recent research demonstrated sensitivity at shorter durations (Loh et al., 2004). The

[3] Alco-Sensor IV, handheld breath-alcohol testing instrument by Intoximeters, Inc., St. Louis, MO.

[4] Motionlogger Actigraph, wrist monitor for measuring activity levels by Ambulatory Monitoring Inc., Ardsley, NY.

[5] Digital Artefacts, iPad-based cognitive assessment, by Digital Artefacts, Iowa City, IA.

Digital Artifacts PVT duration is 5 minutes. Accuracy and reaction time were measured. This version of the PVT was also used in phase 1 (Brown et al., 2014).

### *Driving Scenarios*

Each drive was composed of three nighttime driving segments. The drive started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph with signal-controlled and uncontrolled intersections. An interstate segment followed and consisted of a four-lane divided expressway with a 70-mph posted speed limit. After merging onto the interstate segment, drivers made lane changes to pass several slower-moving trucks. The drives concluded with a rural segment composed of a two-lane undivided road with curves. These three segments mimicked a drive home from an urban parking spot to a rural house via an interstate. The three segments combined to provide a representative trip home in which drivers encountered situations that might be present in a real drive. Participants received recorded audio navigational instructions to guide them through the route. The drive was segmented into “events” that were used to characterize driving performance in various road environments such as urban curves, yellow light dilemma, interstate merge, rural dark road, etc. Event durations varied from around 30 seconds to several minutes. Scenario events are summarized in Table 7.

To minimize learning effects, three versions of the drives were created, and each participant was assigned to two of the three drives in a counterbalanced order. These three versions of the drive were also used in phase 1. Each included the same number of curves and turns, but their order varied within a scenario segment. For example, the position of the left turn in the urban section varied so that it was located at a different position in each version. Additionally, the order of the left and right rural curves varied between versions.

*Table 7. Summary of events for each of the three segments of the drive*

| Segment | Event Name (number) | Description | Approximate Duration (seconds) |
|---|---|---|---|
| Urban | Pull out (101) | Pull out of parallel parking spot into traffic | 30 |
| | Urban Drive (102) | Drive on a narrow two-lane road with traffic and parked vehicles | 45 |
| | Green Light (103) | Navigate green traffic light on urban two-lane road with parked vehicles along the road, oncoming traffic, traffic behind driver | 30 |
| | Yellow Light Dilemma (104) | Navigate yellow light dilemma on urban two-lane with parked vehicle, oncoming traffic, traffic behind driver | 75 |
| | Left Turn (105) | Left turn at signalized intersection (no green arrow, no dedicated turn lane), oncoming traffic, variety of gaps | 80 |
| | Urban Curves (106) | Three curves of mixed radius | 180 |

| Segment | Event Name (number) | Description | Approximate Duration (seconds) |
|---|---|---|---|
| Freeway | Turn on Ramp (201) | Turn right onto interstate on-ramp | 30 |
| | Merge on (202) | Merge onto interstate | 50 |
| | Following (203) | Intermittent slower-moving truck traffic in the driving lane and a single slow moving passenger vehicle in the passing lane, interleaved with a CD-change task | 180 |
| | Merging Traffic (204) | Approach second interchange, interact with traffic merging into interstate | 60 |
| | Interstate Curves (205) | Navigate three curves on interstate | 185 |
| | Exit Ramp (206) | Take exit ramp off interstate | 30 |
| Rural | Turn off Ramp (301) | Turn right from ramp onto rural two-lane road | 30 |
| | Lighted Rural (302) | Lighted two-lane rural road, 55 mph | 90 |
| | Transition to Dark Rural (303) | Straight roadway that transitions from lighted to unlit | 20 |
| | Dark Rural (304) | Unlit straight and curved road, segments, center and road edge marking is faded, and the road surface is grayish | 270 |
| | Dark Rural Hairpin Curve (304HP) | A hairpin turn and a vertical curve located within the Dark Rural (304) Hairpin Curve data were not included in the Dark Rural event data | 30 |
| | Transition to Straight Segment (305) | Transition to gravel rural | 30 |
| | Gravel Rural (306) | Dark, gravel rural road | 90 |
| | Gravel extension (308) | Gravel rural extension after IMPACT driveway | 160 |
| | Gravel to Straight Transition (309) | Gravel rural extension that transitions to straight segment | 25 |
| | Straight Segment (311) | Unlit two-lane rural road, straight, 55 mph | 600 |

## Experimental Design and Independent Variables

A 2 x 2 x 3 x 2 mixed design was used for the collection of data in this study. Between-subject independent variables were gender, mitigation type (staged and discrete), and alert modality (audio/visual-manual, haptic, and combined audio/visual-manual + haptic). The within-subject independent variable was level of drowsiness, manipulated by varying drive time (early and late). The following sections describe these factors in detail.

### *Mitigation System*

Two types of mitigations were designed and implemented. One was a single-stage discrete warning and the other was a three-stage warning meant to convey a sense of increasing urgency. Three alert modalities were crossed with the two types of mitigation. They were audio/visual-manual, haptic, and combined audio/visual-manual plus haptic. Interactivity was built into the visual alerts by instructing the driver to press a button to clear the alert after it appeared. Participants were not told how quickly they needed to respond nor what would happen if they failed to respond. The mitigation types and modalities combined to make up the six study alerts (see Table 8). Participants were randomly assigned to one of the six mitigation types by alert modality combinations.

The discrete warning was a binary alert that was triggered when drowsiness was detected. The visual component of the alert is shown as the middle icon in Figure 12 (Schwarz et al., 2015). The audio component was a two-tone chime that coincided with the appearance of the icon. The haptic component of the discrete warning was a double vibration pulse on the left and right sides of the driver's seat bottom. All three alert modes fired together in the combined condition.

*Table 8. Experimental design: Mitigation type and alert modality.*

| | Audio/Visual | Haptic | Audio/Visual + Haptic |
|---|---|---|---|
| **Staged** | N=8 | N=8 | N=8 |
| **Discrete** | N=8 | N=8 | N=8 |

*Figure 12. The visual icons presented during audio/visual and combined alerts*

The staged warning consisted of three binary alerts that were triggered in stages to reflect increased urgency the longer the driver remained drowsy, according to the algorithm. The visual component of the staged warning consisted of the three icons shown in Figure 12. The changing colors, from white to yellow to red, showed the increasing urgency of each stage. The audio component had three sounds, one for each icon. The first stage, equivalent timing with the discrete alert, had a single beep. The second stage had a two-tone chime. The third stage had a loud repeating beep that lasted about 3 seconds. The haptic component of the staged warning consisted of three types of vibration patterns. The first stage had a single vibration pulse. The second stage had a double vibration pulse, aligned with the two-tone chime. The third stage of the haptic alert was a repeating vibration pulse that aligned with the repeating beep of the audio alert. The visual, auditory, and haptic components of the second stage of the staged alert defined the discrete alert.

The drowsiness-detection algorithm determined when the alerts initially occurred. Figure 13 shows the stages of the staged alert; and the discrete alert occurs when the driver departs from the "awake" state. Each stage persisted for at least 60 seconds to prevent alerts from firing too often.

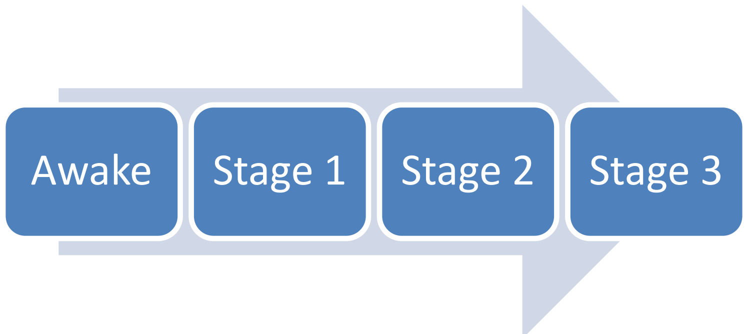



*Figure 13. Mitigation states escalate from normal to stage three.*

The drowsiness algorithm issued assessments of "awake" or "drowsy" every 6 seconds. The mitigation stage escalated from awake to stage one, triggering the appropriate alert, when the algorithm exceeded the drowsiness threshold. If the algorithm output was "awake" when the mitigation state expired after 60 seconds, then the output returned to the previous stage. If drowsiness continued, then the mitigation stage escalated to the next stage (for the staged alert) triggering the next alert. If the mitigation system was already in Stage 3, then an escalation simply re-triggered the alert for that stage. This results in a system that does not directly link the DVI to the algorithm, but instead manages the alerts to balance immediate changes in performance detected by the algorithm and the logical progression of how drowsiness occurs and the desire to influence driver behavior.

The description of the mitigation layer shows that there are many possible ways of linking an impairment algorithm to a mitigation system. Some alerts may under-stimulate the driver and have little short-term or long-term effect on driving performance, while others may be overbearing and be perceived as annoying. The warning types and modalities used in this study were meant to assess a variety of possible DVI implementations.

### *Drowsiness Level*

The choice of the drowsy conditions was designed to provide data consistent with Phase 1, where drowsiness was manipulated via continuous hours awake. Drivers completed two drives, one early night and one late night.

### *No-Mitigation Baseline*

As noted above, data from a subset of participants from phase 1 were used as a baseline control for the mitigation approaches. This subset of data was drawn from the 24 participants (21 to 34 years old), who had completed two drives during early and late night in identical fashion to the procedure used in the present phase 2 study with the exception that they received no drowsiness mitigation.

## Procedure

### *Screening Visit (Visit 1)*

On the first study visit, each participant gave informed consent to participate in the study and received a copy of the signed consent form (Schmitt & Brown, 2015). A researcher verified that the participant held a valid driver's license. Participants then provided a urine sample for the drug screen and, for females, the pregnancy screen, followed by the blood pressure and breath-alcohol tests.

If participants met study criteria, they watched an orientation and training presentation (Marshall & Schmitt, 2015) that provided an overview of the simulator cab, driving task, and CD changing task. Participants completed a screening drive lasting approximately 8 minutes, which consisted of segments like the main drives. The drive included a left-hand turn, driving on two- and four-lane roads, and practicing the CD changing task. After the drive, participants completed a wellness survey with questions about how they felt (Schmitt & Brown, 2015). If the survey showed a propensity for simulator sickness, the participant was deemed ineligible to continue. If eligible, the date and time for visits 2 and 3 were arranged (see Schmitt & Brown, 2015 for and example of the reminder card used).

Prior to visit 2, participants received an activity monitor (an actigraph, see Simulator and Apparatus) that they were instructed to wear until they checked in at visit 2. The monitor recorded periods of activity and inactivity prior to study visits. Participants also were instructed to keep a written activity log (Schmitt & Brown, 2015) during this period to provide more details about activities that could affect their wakefulness.

### *Overnight Visits (Visits 2 and 3)*

Following the screening visit, each participant participated in two nighttime data collection sessions, on subsequent visits, once in the early night condition and once in the late-night condition. The two-night drives were separated by at least 1 week. The early night drives began between 10 p.m. and 1 a.m. and the late-night drives began between 2 a.m. and 5 a.m. Actigraph records ensured that the early night drives occurred after at least 14 hours of continuous wakefulness and that the late-night drives occurred after at least 18 hours of continuous wakefulness. To control learning effects, the order of the two nighttime drives was counterbalanced, and at least a 3-day gap was specified between the two sessions to facilitate recovery and reduce carryover effects.

On the days of each of their overnight visits, participants were instructed to restrict beverage consumption to water after 12 p.m. to eliminate caffeine intake. They were given a list of items containing caffeine to avoid, including coffee, tea, soda, energy bars, energy drinks, and foods with chocolate or candy. After eating dinner, a friend, family member, or taxi (if needed) drove them to NADS to arrive between 5 p.m. and 7:30 p.m. Upon arrival, the activity monitor and log (Schmitt & Brown, 2015) were collected and reviewed by a research assistant, while the participants completed a sleep and food intake survey. The activity monitor data were reviewed to ensure participants had a normal night's sleep (at least 6 hours) the preceding evening and did not take naps during the day of their session. If activity data were not recorded, but the participant claimed that the monitor was worn, only the written activity log was used to determine if the participant was eligible to continue. If the participants claimed that the monitor was taken off or not worn, the participant was dropped from the study. Each participant's breath alcohol level was checked to ensure that there was no alcohol in their system (BrAC = .00 g/dL). If a participant did not meet the sleep or breath alcohol concentrations criteria, they were dropped from the study and returned home. Each participant's caffeine intake was reviewed in the activity log and again in the sleep and intake log to confirm that they had not consumed caffeine since noon. Participants were then assigned to simulator drive times based on their waking times, such that the participant who had the earliest wake time drove first, followed by the participant with the second earliest wake time, and so forth.

While waiting to drive in a dark quiet conference room, participants were allowed to engage in activities such as watching TV or reading. Participants were monitored to ensure they remained awake and did not stand up and walk or converse with other participants. If participants began to fall asleep, a research assistant reminded them to stay awake. The participant completed a SSS survey (Figure 14) (see Schmitt & Brown, 2015) every 30 minutes until they drove. One hour prior to driving, participants were taken to a private secluded room to wait quietly without engaging in any activities. They completed SSS surveys after entering and leaving the private room. Participants were escorted to the simulator, eye tracking calibration procedures were performed, followed by a 5-minute PVT and SSS. After the drive, participants completed a SSS and a wellness survey.

| **Degree of Sleepiness** | **Scale Rating** |
|---|---|
| Feeling active, vital, alert, or wide awake | 1 |
| Functioning at high levels, but not at peak; able to concentrate | 2 |
| Awake, but relaxed; responsive but not fully alert | 3 |
| Somewhat foggy, let down | 4 |
| Foggy; losing interest in remaining awake; slowed down | 5 |
| Sleepy, woozy, fighting sleep; prefer to lie down | 6 |
| No longer fighting sleep, sleep onset soon; having dream-like thoughts | 7 |
| Asleep | X |

*Figure 14. Levels of the SSS*

After completing the surveys, participants were escorted to a private room where they completed a RSS (Schmitt & Brown, 2015), for comparison with phase 1 data, that asked them to plot their level of drowsiness across the drive. At the end of visit 3, participants were also given a debriefing statement (Schmitt & Brown, 2015). At the end of each overnight visit, a taxi was called for participants unless they had someone pick them up. No less than 24 hours prior to visit 3, the activity monitor and log were given to participants, and they were reminded of their next appointment. Participants who did not complete the study received pro-rated compensation, and those who completed it were paid $250.

## Hypotheses and Dependent Variables

Data were analyzed to test hypotheses regarding the impact of the mitigation interfaces. Specific hypotheses included:

- Drowsiness mitigation alerts will reduce the rate of drowsy lane departures and SDLP relative to the no-mitigation baseline.
- Staged alerts will reduce the rate of drowsy lane departures more relative to the mean no-mitigation baseline results for the corresponding drive time (early night or late night) than single-stage discrete alerts.
- Each alert modality will reduce the rate of drowsy lane departures and SDLP relative to the no-mitigation baseline, but they will not differ significantly from one another.
- Drowsiness mitigation will not significantly reduce subjective drowsiness relative to the no-mitigation baseline.

## Potential Intervening Variables

Several intervening variables may have influenced the dependent variables or mediated the influence of drowsiness on the dependent variables. Table 9 summarizes these variables and the approach to minimize any confounding effects they may have produced. The potential confounding of drowsiness by individual differences in sleep/wake times was minimized by scheduling the drives according to the typical wake times of participants and suggesting they maintained their typical sleep schedule during the study.

## Data Verification and Validation

The data reduction began with verification of the raw input data. The data were then aggregated as needed to support sensitivity analyses and algorithm development and testing. The process concluded with validation of metrics that summarized the data.

### *Verification of Input Data*

Verification concerned the process of ensuring that the raw data accurately reflected the state of the vehicle, driver, and roadway. Scenario event errors, database flaws, and measurement noise all may have contributed to false raw data that needed to be removed before they could be transformed into measures of driver behavior. Several automatic data checks combined with manual visualizations identified these issues.

The simplest approach was to verify that all the variables in the raw data contained values and that the file was of the expected size. These two simple strategies identified problems associated with data logging and were performed immediately after each drive to ensure that the basic data collection was successful.

A slightly more thorough analysis assessed the integrity of each variable. The integrity of the variable depended on three factors. The first consideration was whether the values lay within the expected range of the variable. For example, depending on the lane configuration, lane position values could not deviate from -8 to +8 feet. Values outside this range were suspect and merited investigation. The second consideration was whether the values varied in a meaningful manner. A possible failure mode was a sensor that stopped showing system state but continued to produce a valid value. For example, a steering wheel angle of zero was valid, but if this value did not change for more than 10 seconds, the data was investigated. Calculating the minimum, maximum, and standard deviation could show a problem if the minimum and maximum were the same and the standard deviation was zero. The third consideration was whether the variation in the values was continuous. With lane position, physics limited the change in lateral position over time if the vehicle suddenly moved horizontally at 100 mph, there was likely a problem with the data. Calculating the derivative of values (the difference between two sequential data points divided by the time between samples) and then calculating the minimum and maximum of this value could also show discontinuities in the data. For example, such discontinuities could occur when the driver moved between lanes and the lane position was measured with reference to the lane that the driver was in instead of using the new lane as a reference. This error could lead to spurious calculation of the SDLP. Unexpected range, lack of variability, and discontinuities were automatically assessed or revealed in a plot of the data.

### *Validation of Summary Measures*

Validation concerned the process of ensuring that the summary measures accurately reflected the driver behavior or vehicle performance of interest. Summary measures should represent the underlying data, but this can fail when standard measures, such as SDLP, do not characterize the lane-keeping behavior precisely. The capacity of summary variables to characterize behavior can fail at two levels. First, measures based on raw data might not reflect the actual time history of the raw data because of the need for a consistent window duration that might not provide sufficient context. Second, aggregating measures at the sample level (across events, drives, or people) might fail to reflect the underlying population differences in behavior due to such issues as differences in the distribution of the data, or the presence of data that differs in significant ways from the rest of the sample. Data visualization techniques provided a useful tool for addressing both challenges to the validity of summary measures.

Data were visualized by superimposing the summary measures over the raw data, so that a brief inspection could assess whether the summary measure reflected the data sufficiently. For example, the time history of lane position could be augmented with summary measures, such as the minimum time-to-lane crossing and the SDLP. Overlaying these summary measures and raw data along with a reference point such as the posted speed limit made it possible to roughly assess whether the underlying calculations were correct and whether the variables captured behavior. This was particularly critical if developing a new variable such as one that aimed to distinguish between different types of speeding behavior. A visual representation might have identified separate types of behavior that could otherwise have been combined.

*Table 9. Intervening variables that might moderate the effect of drowsiness*

| Variable | Rationale | Mitigation |
| --- | --- | --- |
| Daytime Drowsiness | People who suffer from daytime drowsiness may not have significant differences in performance between their daytime and nighttime drives, as both could be classified as drowsy. This may be associated with a medical condition such obstructive sleep apnea or narcolepsy or could be associated with abnormal or lack of sleep the night before the daytime drive. | Participants with reported sleep disorders were excluded from the study.<br>Participants were asked to track their sleep including the use of an activity monitor. For participants who have potentially confounding sleep, an attempt was made to reschedule. |
| Nighttime Wakefulness | People who will be overly awake at night may not exhibit drowsiness at the planned data collection times. | Participants were asked to maintain a normal sleep pattern and not to nap the day of their overnight drive. Participant sleep patterns were monitored through an activity monitor and activity logs. Participants who deviated from typical nighttime sleep patterns (shift work and extreme evening people) were dropped from the study. |
| Physical Fatigue | Excessive physical activity may confound the effects of drowsiness, particularly if present on only one of the data collection days. | Participants were asked to maintain the same pattern of activity on data collection days. |
| Ambient temperature | Can affect simulator sickness. | Temperature was set at 72° F. |
| Simulator sickness score on wellness survey (see Schmitt & Brown, 2015) | Higher levels of simulator sickness can result in potential changes in driver behavior to counteract its effects. | Participants took a practice drive during their screening. Participants with high scores (Nausea $\geq$ 21, Oculomotor $\geq$ 32, Disorientation $\geq$ 15, or Total $\geq$ 32) following the screening drive were excluded from the study. |
| Different levels of drowsiness between phase 1 and 2 | Any differences in procedure or driving task could have led to differing levels of drowsiness that would confound the comparison of mitigation and no-mitigation. | The procedure used to induce drowsiness and the driving task were kept as identical as possible between the two studies. |

# Results

## Overview

The primary goal of the analyses was to determine whether drowsiness mitigation reduced the risk of drowsiness-related lane departures and reduced SDLP relative to the driving performance of baseline drivers who did not receive mitigation (24 participants from phase 1). Secondary measures included deviations in speed and speed variability (standard deviation in speed) from the speed limit, as well as subjective drowsiness assessed by scores on the SSS, which were collected immediately before and after the drives. The availability of the objective dependent variables varied across the 22 distinct events that comprised the three segments of the drive.

## Verification of Drowsiness Manipulation

As stated in the Drowsiness Algorithm Augmentation section, drowsiness was manipulated by varying the time of day at which a drive occurred, thereby manipulating a participant's continuous hours awake. Drivers in both phase 1 and 2 completed early night and late-night drives, which were meant to induce moderate and severe drowsiness, respectively. To verify that the time-of-day manipulation was able to induce different levels of drowsiness, SSS scores collected pre- and post- drive for both groups were compared. As seen in Figure 15, SSS scores increased from early to late night drives for both groups, demonstrating that the manipulation induced subjective driver drowsiness.

## Identification of Drowsy Lane Departures

The first goal was to establish whether the drowsiness mitigation reduced the rate of drowsy lane departures and lane position variability relative to the phase 1 no-mitigation baseline groups. Lane departures were recorded each time a front wheel of the driver's vehicle crossed one of the lane lines. Raters performed video verification using a procedure identical to phase 1 because some of the lane departures may not have been associated with drowsiness. For each lane departure, a researcher scored the corresponding portion of the video showing the driver's face and determined whether the lane departure was associated with a drowsiness episode, assessed as a score greater than 2 on the objective rating of drowsiness scale (Wierwille & Ellsworth, 1994). Schmitt and Brown (2015) provide the description of each level of drowsiness.

The second goal was to establish that the participants in the early and late-night phase 1 data collections without mitigation were not significantly more or less drowsy than the phase 2 participants. As shown in Figure 26, there was no effect of mitigation condition on SSS scores, $F(1,133) = 2.79$, $p = 0.10$, showing that pre-drive, drivers in the phase 1 data set were not more or less drowsy than drivers in the phase 2 study.

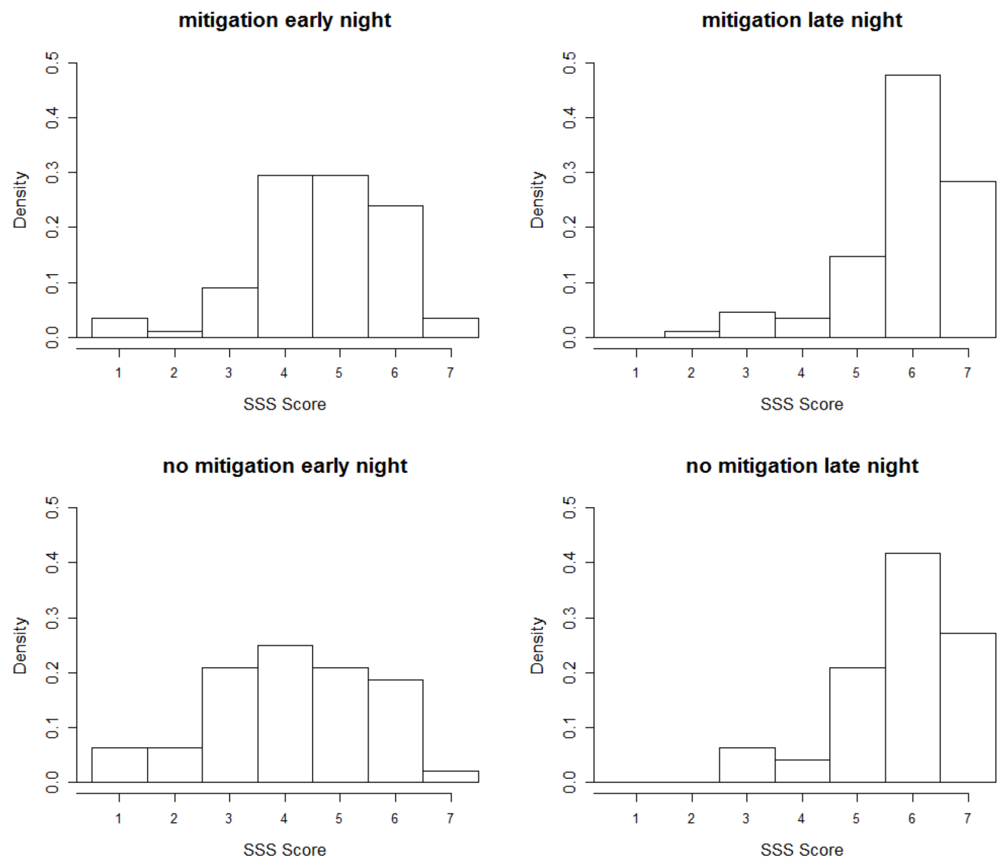



*Figure 15. Histogram of scores on the SSS for early and late-night drives in the mitigation and no-mitigation groups. Higher SSS scores show greater drowsiness.*

## Events Associated with Drowsy Lane Departures

Inspection of these drowsy lane departure data showed that nearly all the drowsy lane departures occurred in five events, as seen in Figure 16, on the interstate (events 203 and 205), on the dark rural roadway (events 303 and 304), and on the dark straight rural section at the end of the drive (event 311). Therefore, the following analyses were restricted to those events and data from them were combined. Consequently, the results do not represent performance on all roads; rather they represent performance on roads that are more likely to engender risk for drowsy drivers. Furthermore, two participants from phase 2 produced a much higher rate of lane departures than any other participant in phase 1 or 2. Video review revealed that they had fallen asleep for large sections of both the early and late-night drives. Data from these participants was excluded from all analyses.

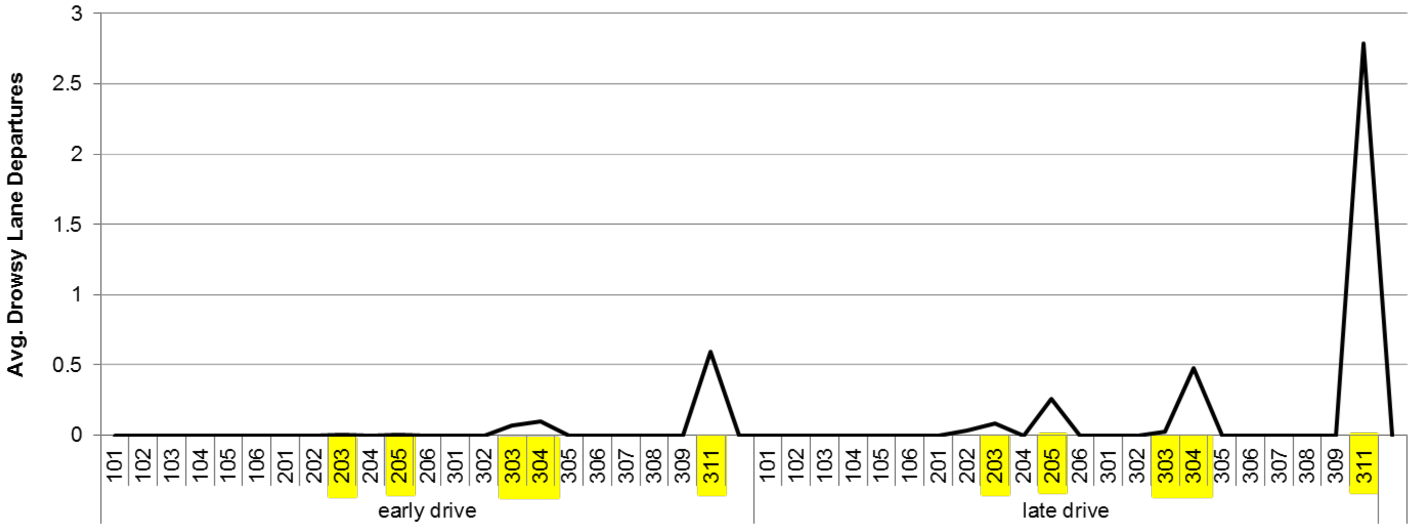



*Figure 16. Average number of drowsy lane departures in each event of early and late-night drives. Highlighted events show those included in the subsequent analysis.*

## Evaluation of Drowsiness Mitigation

### *Overview*

The goal of this set of analyses was to determine whether the presence of drowsiness mitigation improved driving performance relative to no mitigation. Data were collapsed across types and modalities of mitigation to provide more statistical power. Mixed factor analysis of variances were performed with mitigation (mitigation versus no mitigation) as a between-subjects factor and drive time (early versus late) as a within-subjects factor. Means and standard deviations for each of the following measures can be found in Appendix A.

Boxplots represent the median (dark bar), first and third quartiles (box edges), and min/max values. Circles represent individual participant means and red squares represent group means. Points beyond the horizontal bars represent outliers.

### ***Lane Departures per Minute***

#### **Drowsy lane departures**

To control for the varied length of events and varied speed among drivers, we calculated the rate of lane departures per minute. As seen in Figure 17, the drowsiness mitigation resulted in a marginal reduction in lane departures per minute compared to the no-mitigation baseline during both the early and late drives. The main effect of mitigation was marginally significant, $F(1,68) = 3.76$, $p = .057$. As expected, the frequency of lane departures increased in the late drives compared to the early drives, $F(1,68) = 17.59$, $p < .001$. The interaction between mitigation and drive time was not significant, $F(1,68) = 1.59$, $p = .21$.

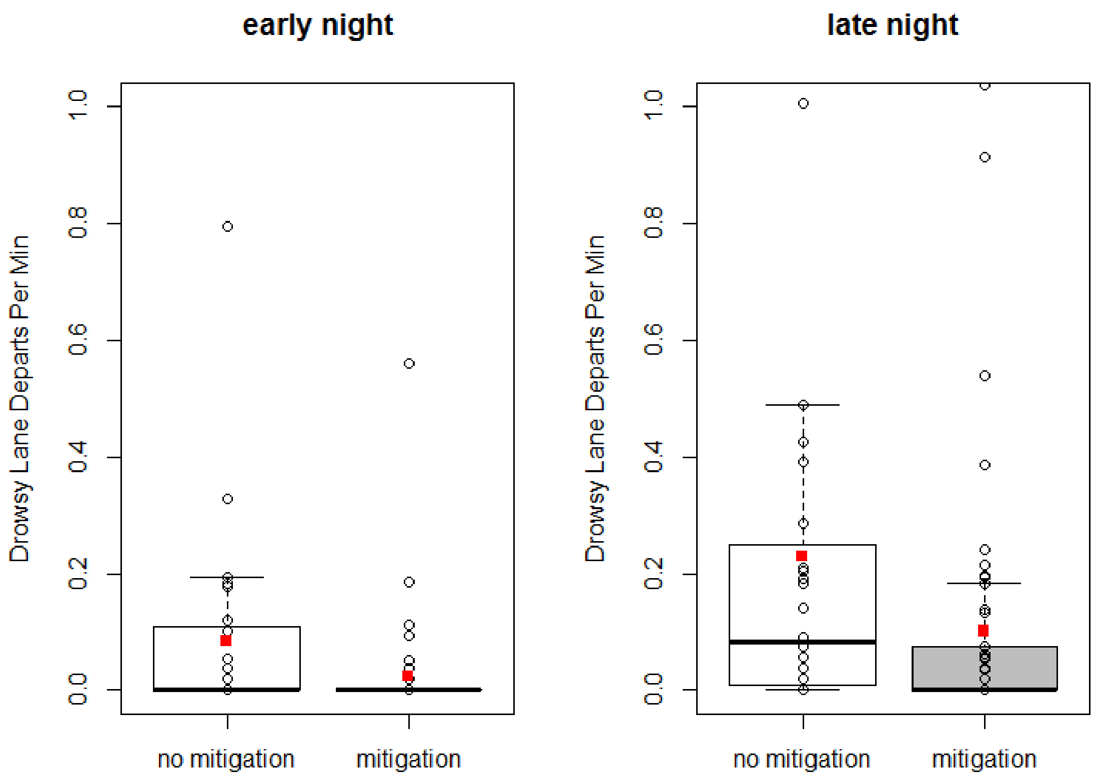



*Figure 17. Boxplots of drowsy lane departures per minute in early and late drives. Circles represent individual participant means. Red squares represent group means.*

### Awake lane departures

To determine whether the mitigation caused drivers to be more cautious overall (and did not have an impact directly on drowsiness), we calculated the average rate of lane departures that were coded as awake (i.e., non-drowsy) in the mitigation and no-mitigation groups. As Figure 17 shows, awake lane departures per minute did not vary as a function of either drive time, $F(1,68) = 2.06$, $p = .16$, or mitigation, $F(1,68) = 0.01$, $p = .95$. The interaction between drive time and mitigation was also not significant, $F(1,68) = 0.99$, $p = .32$. The rate of awake or "normal" lane departures was 8.8 times that of drowsy lane departures.

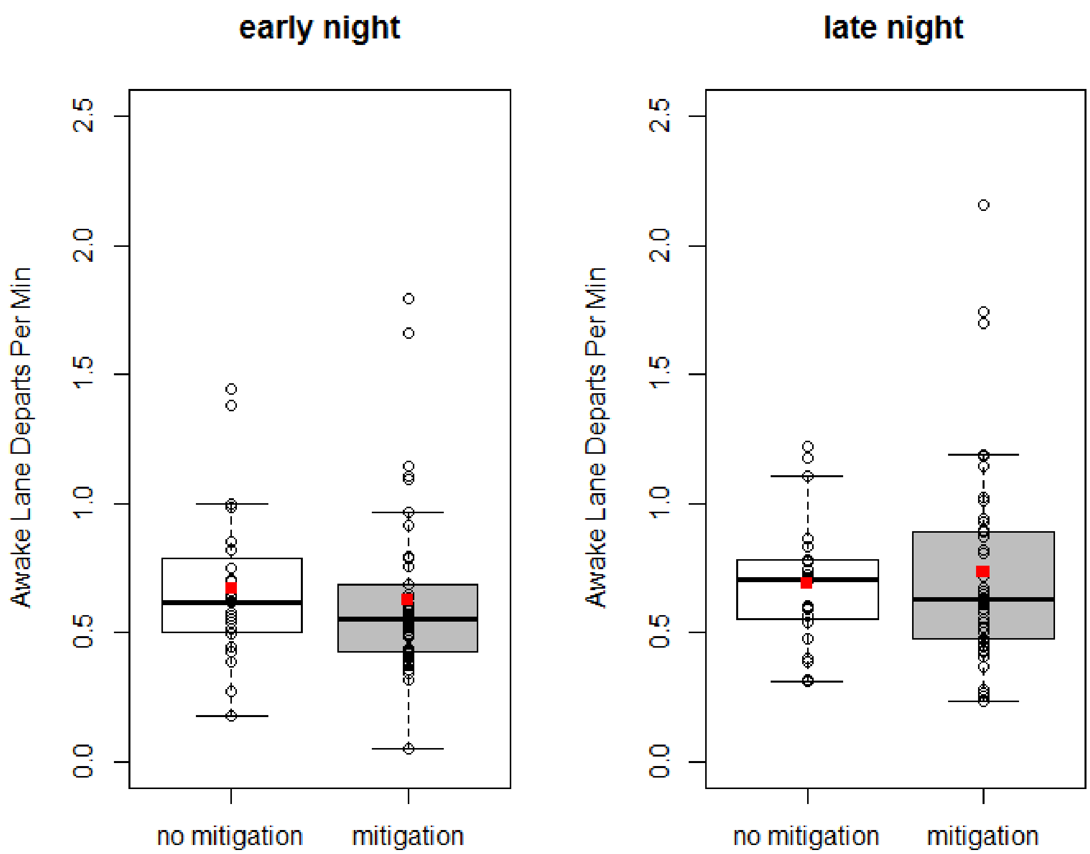



*Figure 18. Boxplots of awake lane departures per minute for early and late drives. Circles represent individual participant means. Red squares represent group means.*

### ***Standard Deviation of Lane Position***

SDLP provides a measure of the variability in lateral lane position. As seen in Figure 19, the mitigation significantly reduced SDLP compared to the no-mitigation baseline, $F(1,68) = 7.92$, $p = .006$. As expected, SDLP increased in the late drive relative to the early drive, $F(1,68) = 21.88$, $p < .001$. However, there was no interaction between mitigation group and drive time, $F(1,68) = 0.14$, $p = .70$.

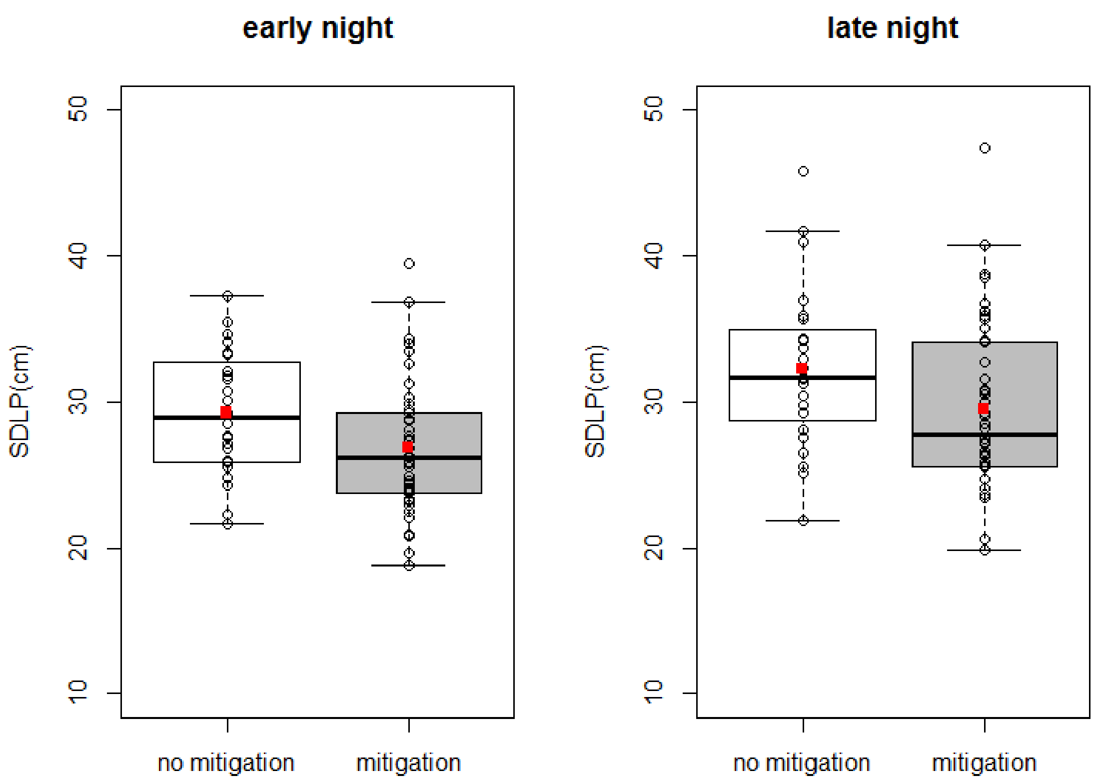



*Figure 19. Boxplots of SDLP for early and late drives. Circles represent individual participant means. Red squares represent group means.*

### ***Average Speed and Speed Variability Relative to the Speed Limit***

In addition to determining whether the drowsiness mitigation affected lateral control, we examined its effects on maintaining speed at the speed limit. To standardize speed across different events with different speed limits, we computed average speed relative to the speed limit by subtracting the speed limit from the participant's average speed across events. Thus, a negative value showed that participants were driving below the speed limit, on average. Figure 20 presents speed relative to the speed limit during early and late drives. Somewhat unexpectedly, the main effect of drive time was not significant, $F(1,68) = 1.08$, $p = .30$, nor was the main effect of mitigation, $F(1,68) = 0.11$, $p = .74$, or the interaction between drive time and mitigation, $F(1,68) = 1.93$, $p = .17$. Standard deviation in speed relative to the speed limit was calculated to estimate variability in speed and presented in Figure 21. The drowsiness mitigation significantly reduced the standard deviation of relative speed, $F(1,68) = 10.53$, $p = .002$. Drive time resulted in a marginal increase in standard deviation speed relative to the speed limit, $F(1,68) = 1.46$, $p = .08$. The drive time by mitigation interaction was not significant, $F(1,68) = 0.04$, $p = .76$.

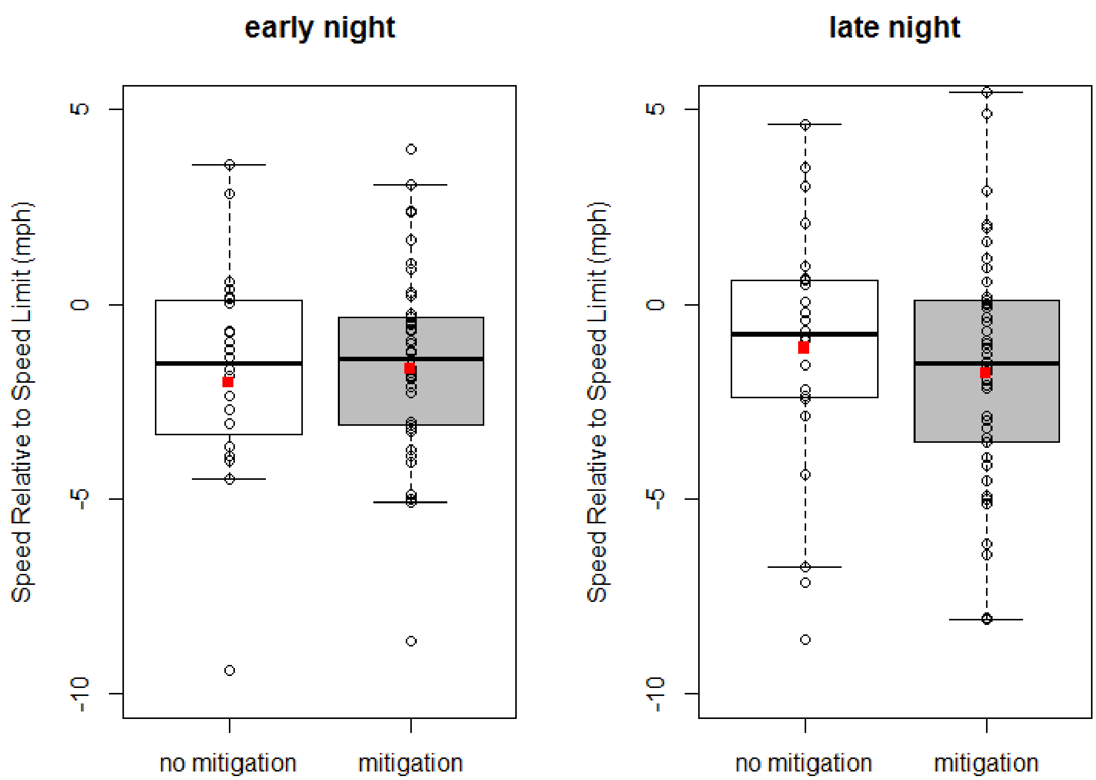



*Figure 20. Boxplots of speed relative to the speed limit during early and late drives. Circles represent individual participant means. Red squares represent group means.*

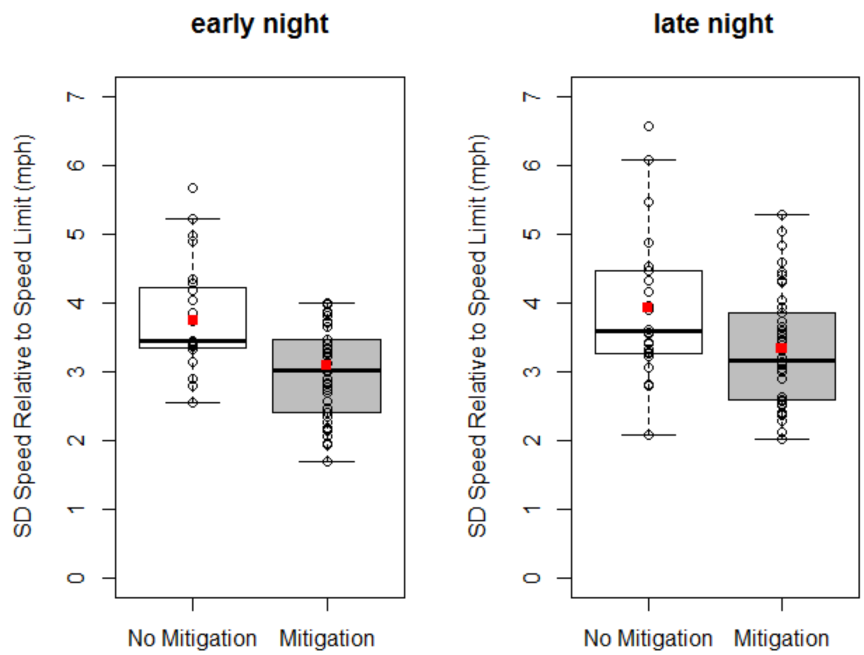



*Figure 21. Boxplots of speed relative to the speed limit during early and late drives. Circles represent individual participant means. Red squares represent group means.*

## Comparison of Mitigation Type and Alert Modality

### *Overview*

The above analyses demonstrate that the drowsiness mitigation reduced the rate of drowsy lane departures and SDLP relative to the phase 1 no-mitigation baseline. The goal of the next set of analyses was to determine the influence of mitigation type (discrete versus staged) and alert modality (audio/visual-manual, haptic, and combined audio/visual-manual and haptic).

For each performance measure, difference scores were computed by subtracting the mean of the baseline (no-mitigation) group from each individual mitigation participant's mean for early and late-night drives separately. These difference scores thus provide an index of how much performance differed from baseline, and in what direction, for each of the mitigation groups. For instance, if there were an average of three lane departures per minute in the no-mitigation group and the discrete audio/visual mitigation resulted in two lane departures per minute for a given participant (on average), the difference score for that participant would be -1, showing an improvement relative to baseline of one fewer lane departure per minute. Next, the average of these difference scores for each of the mitigation conditions was calculated.

These difference scores were analyzed using mixed-factor ANOVAs with mitigation type (discrete and staged) and alert modality (A, H, C) as between-subjects factors and drive time (early and late) as a within-subjects factor. Pairwise comparisons with Tukey correction were performed where appropriate. Difference values significantly more negative as a function of mitigation type, alert modality, or both, would provide evidence that one or more mitigation interfaces were associated with a larger reduction in drowsy lane departures than other interfaces.

### ***Drowsy Lane Departures per Minute***

Neither the main effect of mitigation type, $F(1,41) = 0.62$, $p = .44$, nor the main effect of alert modality, $F(2,41) = 0.41$, $p = .66$, were significant, nor was the interaction between mitigation type and modality, $F(1,41) = 0.32$, $p = .57$. Most importantly, however, there was an interaction between drive time and mitigation type showing that during the late-night drives, the staged mitigation reduced lane departures per minute relative to baseline more than the discrete mitigation, as shown in Figure 22.

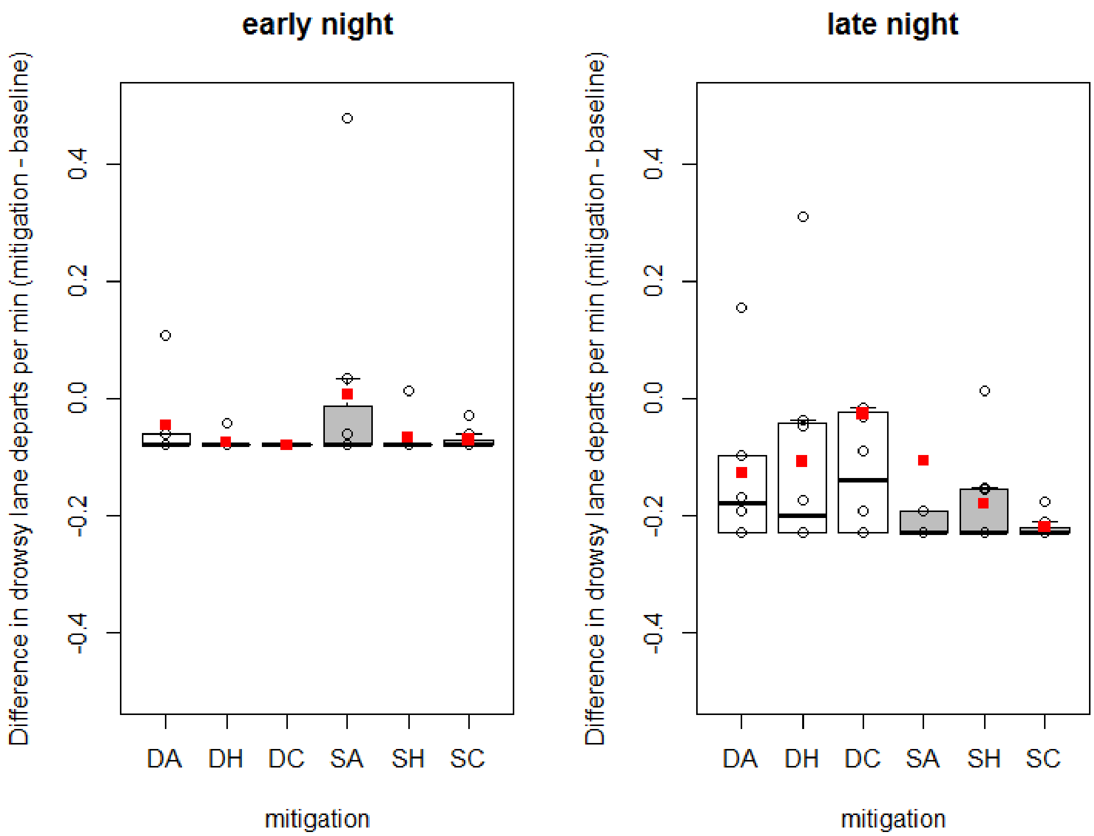



*Figure 22. Boxplots of difference in drowsy lane departures per minute between each mitigation condition and baseline. Lower values reflect a lower rate of drowsy lane departures relative to the baseline no-mitigation group. DA = discrete audio/visual-manual, DH = discrete haptic, DC = discrete combined, SA = staged audio/visual-manual, SH = staged haptic, SC = staged combined. Circles represent individual participant means. Red squares represent group means.*

### *Standard Deviation in Lane Position*

For the difference in SDLP from baseline, neither the main effect of mitigation type, $F(1,41) = 2.35$, $p = .13$, nor the main effect of alert modality, $F(2,41) = 1.81$, $p = .18$, were significant. There was a significant interaction between mitigation type and drive time showing that during the late night drives the staged mitigation reduced SDLP relative to baseline more than did the discrete mitigation, as seen in Figure 23, $F(1,41) = 5.14$, $p = .03$.

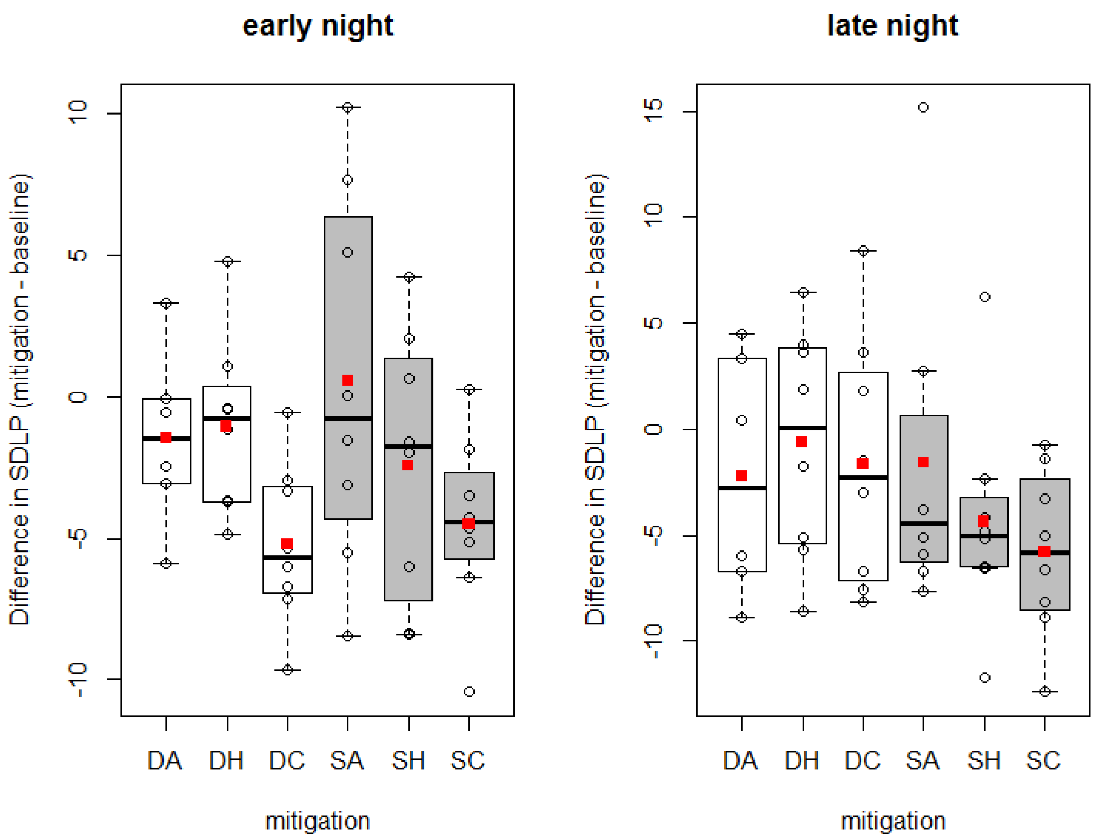



*Figure 23.* *Boxplots of difference in SDLP between each mitigation condition and baseline. Lower values reflect better lane keeping relative to the baseline no-mitigation group. DA = discrete audio/visual, DH = discrete haptic, DC = discrete combined, SA = staged audio/visual, SH = staged haptic, SC = staged combined. Circles represent individual participant means. Red squares represent group means.*

### ***Speed Relative to Speed Limit***

For differences from baseline in standard deviation in speed relative to the speed limit, neither the main effect of mitigation type, $F(1,41) = 0.36$, $p = .55$, nor alert modality, $F(2,41) = 0.38$, $p = .68$, were significant. Furthermore, the interaction between mitigation type and time of day was not significant, $F(1,41) = 0.01$, $p = .99$, nor was the interaction between time of day and alert modality, $F(1,41) = 1.38$, $p = .26$, as seen in Figure 24.

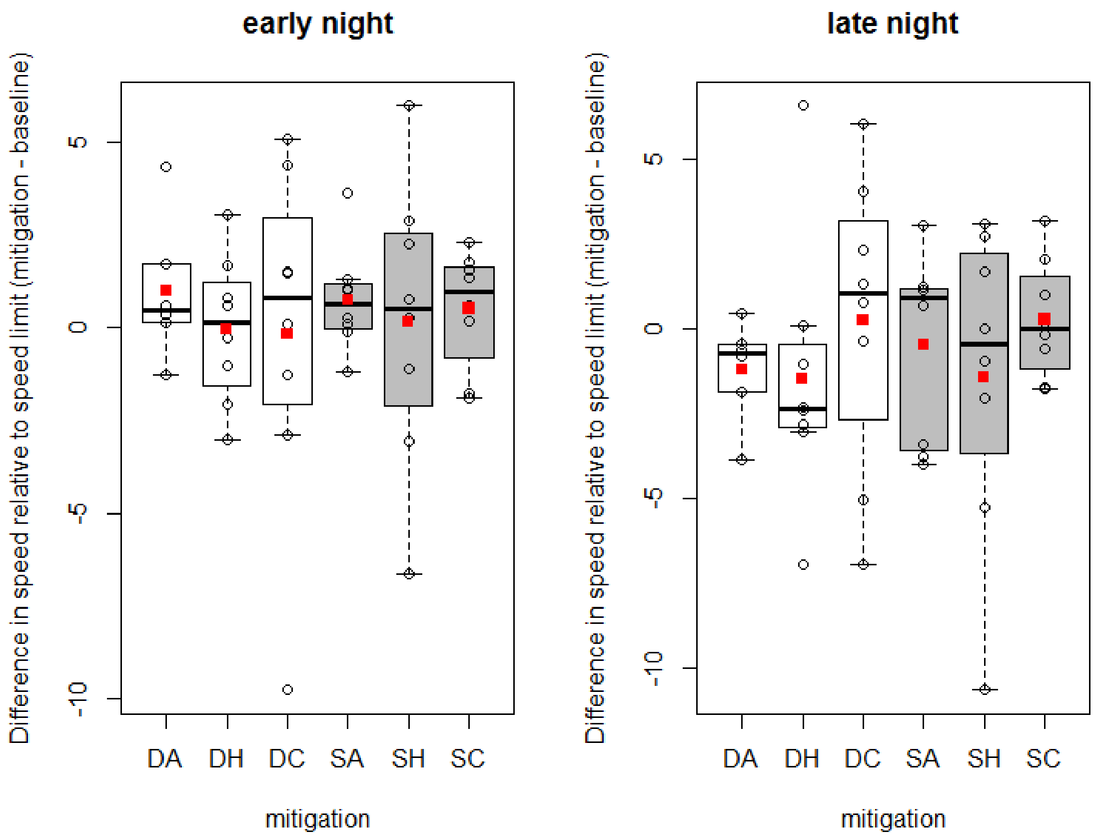



*Figure 24.* *Boxplots of speed relative to the speed limit between each mitigation condition and baseline. Lower values show that the mitigation reduced speed compared to the baseline no-mitigation group. DA = discrete audio/visual, DH = discrete haptic, DC = discrete combined, SA = staged audio/visual, SH = staged haptic, SC = staged combined. Circles represent individual participant means. Red squares represent group means.*

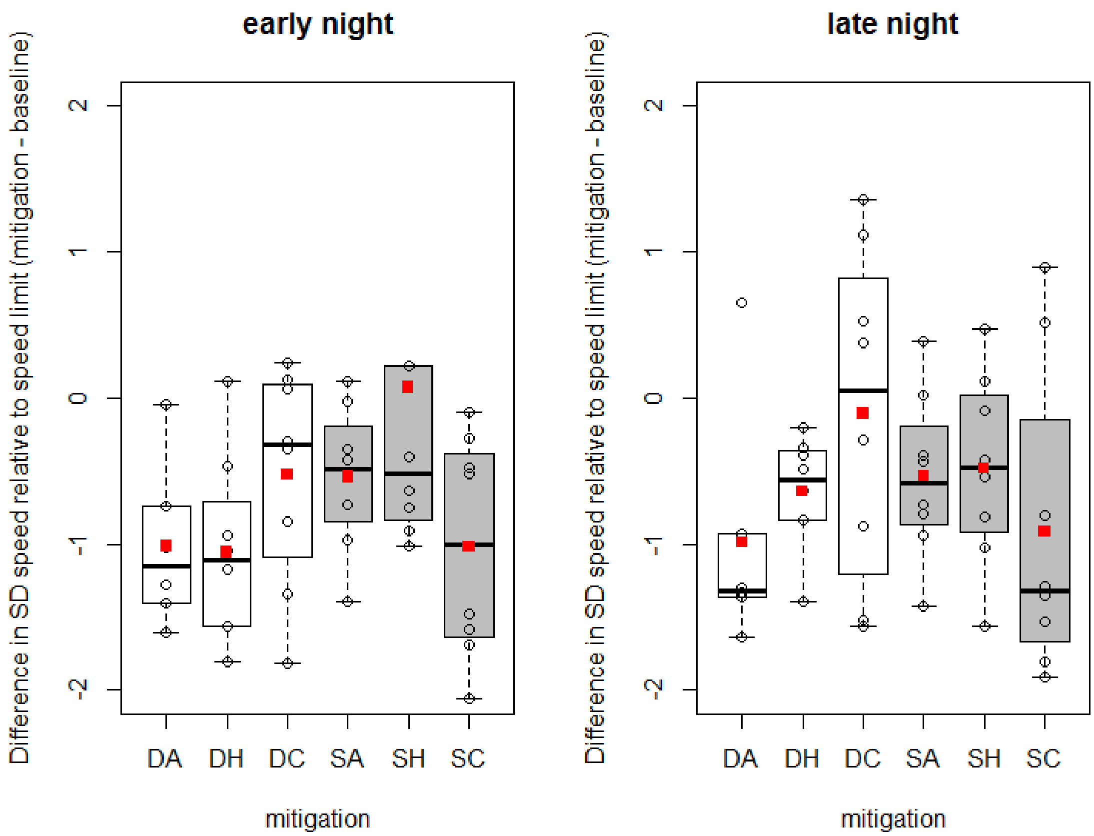



*Figure 25.* *Boxplots of difference in standard deviation in speed relative to the speed limit between each mitigation condition and baseline. Lower values reflect reduced variability with the mitigation compared to the baseline no-mitigation group. DA = discrete audio/visual, DH = discrete haptic, DC = discrete combined, SA = staged audio/visual, SH = staged haptic, SC = staged combined. Circles represent individual participant means. Red squares represent group means.*

### ***Standard Deviation in Speed Relative to Speed Limit***

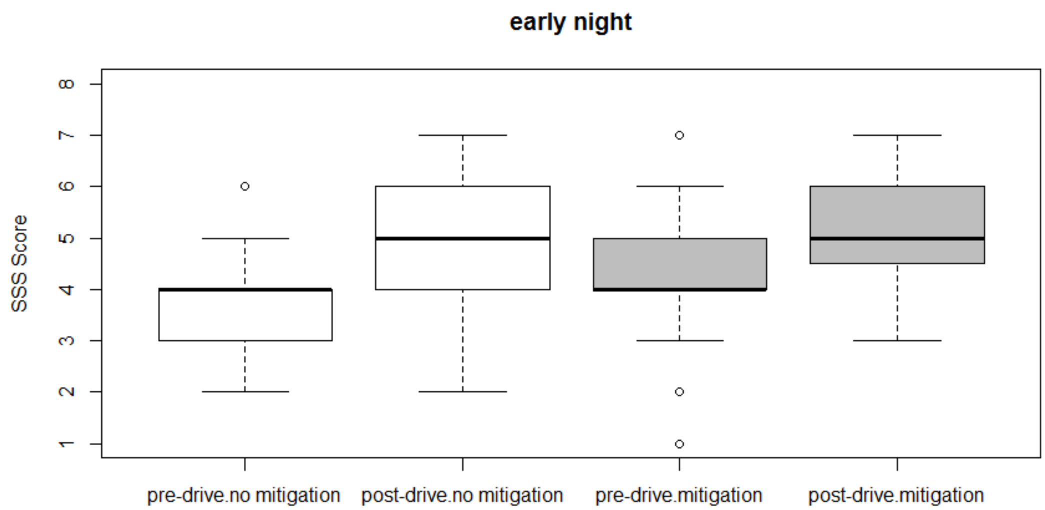
For standard deviation in speed relative to the speed limit, shown in Figure 25, neither the main effect of mitigation type, $F(1,41) = 0.36$, $p = .55$, nor alert modality, $F(2,41) = 0.38$, $p = .68$, were significant. Furthermore, the interaction between mitigation type and modality was not significant, $p > .20$. The main effect of drive time was not significant, $F(1,41) = 0.37$, $p = .55$, and no interactions involving drive time, mitigation type, and alert modality were significant, $p > .19$.

## Subjective Drowsiness

The final goal of the analyses was to determine whether drowsiness mitigation reduced subjective drowsiness relative to the no-mitigation baseline. Subjective drowsiness was obtained from scores on the SSS, which was administered immediately before and after each drive. Higher scores on the SSS show greater levels of subjective drowsiness.

SSS scores were compared with a mixed-factor ANOVA with mitigation condition (mitigation, no-mitigation) as a between-subjects factor and drive order (pre-drive, post-drive) and drive time (early, late) as within-subjects factors. One subject from the mitigation group was excluded because of technical failures that prevented recording a post-drive SSS during the early drive.

Figure 26 presents pre- and post-drive SSS scores for the early and late-night drives. As expected, SSS scores increased significantly in the late drives compared to the early drives, $F(1,133) = 55.02$, $p < .001$, and also increased significantly from pre-drive to post-drive, $F(1,133) = 71.42$, $p < .001$. However, there was no effect of mitigation condition on SSS scores, $F(1,133) = 2.79$, $p = .10$, or significant interactions involving mitigation, drive time, and drive order , $p > .15$.



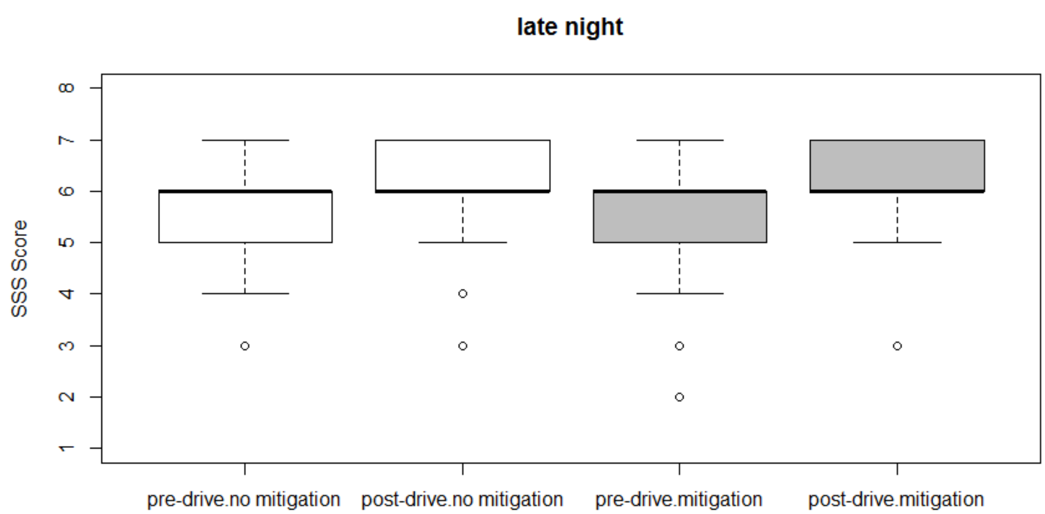




*Figure 26. Boxplots of scores on the SSS for early and late drives in the mitigation and no-mitigation groups. Higher scores show greater levels of sleepiness. DA = discrete audio/visual, DH = discrete haptic, DC = discrete combined, SA = staged audio/visual, SH = staged haptic, SC = staged combined.*

## Conclusions

These three sets of analyses addressed the potential for three drowsiness mitigation warnings to reduce the risk of drowsy lane departures and improving SDLP. Specific conclusions are presented below.

**Does drowsiness mitigation reduce drowsy lane departures?**

Results showed a marginally significant reduction in lane departures per minute for the mitigation compared to no-mitigation in both the early and late drives. The drivers' only choice was to continue, even if they would have pulled over to rest in actual driving. Importantly, there were no significant differences between the mitigation and no-mitigation groups in the much higher rate of awake lane departures. This demonstrates that the mitigation had a targeted effect on drowsy lane departures.

**Does drowsiness mitigation improve consistency in lane keeping and speed?**

In addition to the benefit of mitigation for lane departures, drowsiness alerts also resulted in reduced SDLP compared to SDLP for drowsy driving without-mitigation, suggesting that the mitigation improved lateral control. Furthermore, while the mitigation did not appear to affect speed relative to the speed limit, it reduced variability in speed relative to the speed limit compared to the no-mitigation baseline, showing that the mitigation resulted in more consistent speed control.

**Do multi-staged mitigations have a greater impact than discrete single-stage mitigations?**

Staged alerts result in larger reductions in drowsy lane departures and lane position variability relative to baseline compared to discrete alerts during the late-night drives when drivers were drowsier. Thus, staged alerts may have a greater impact than discrete alerts for extremely drowsy drivers over the course of a short (i.e., 1 hour) drive.

**Does alert modality affect mitigation impacts?**

We found no significant differences among the evaluated drowsiness mitigation modalities with respect to any of the driving performance metrics: drowsy lane departures, SDLP, and average and SD of speed relative to the speed limit.

**Does drowsiness mitigation influence subjective drowsiness?**

As expected, late night drives resulted in significantly higher subjective sleepiness scores than early night drives. This shows that the drive time manipulation increased drowsiness. However, drowsiness alerts did not alter subjective drowsiness in either the early or late drive times. This is not, however, altogether surprising. Unlike distraction, which is under drivers' voluntary control, drivers may be unable to simply stop feeling drowsy even if they receive feedback via the mitigation. However, another possibility is that the SSS is less sensitive than SDLP or other driving performance measures.

### Limitations and Gaps

Several limitations exist in the present research. First, the baseline no-mitigation data were collected from a different sample of drivers during a different data collection. It is possible that small, unintentional differences between the two studies occurred (due to different research staff, upgraded projectors in the simulator, etc.), however this possibility is not supported by data. As evidence, subjective drowsiness levels did not differ significantly in the early or late-night drives across studies, nor did the number of awake lane departures. This suggests that the critical difference was the intended factor of drowsiness mitigation. Furthermore, sample size in the present study was relatively small, with a total of eight subjects in each mitigation condition. Although the drive time manipulation used to induce drowsiness was consistent across participants and studies, substantial inter-individual variability exists in terms of drowsiness.

While these results suggest that the mitigation warnings improved lane keeping relative to no mitigation, it is unclear whether the drowsiness detection algorithm contributed to this mitigation. The lack of a control group that received random mitigation warnings makes it impossible to determine for certain that the changes are the result of the mitigation approaches tested.

## References


Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences, 12*(3), 241–253.

Åkerstedt, T., Ingre, M., Kecklund, G., Anund, A., Sandberg, D., Wahde, M., Philip, P., & Kronberg, P. (2010). Reaction of sleepiness indicators to partial sleep deprivation, time of day and time on task in a driving simulator–the DROWSI project. *Journal of Sleep Research, 19*(2), 298–309.

Arnedt, J. T., Wilde, G. J., Munt, P. W., & MacLean, A. W. (2001). How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accident Analysis & Prevention, 33*(3), 337–344.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. http://doi.org/10.1023/A:1010933404324

Brown, C. D., & Davis, H. T. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems, 80*(1), 24–38.

Brown, T., Lee, J., Schwarz, C., Fiorentino, D., & McDonald, A. (2014, February). *Assessing the feasibility of vehicle-based sensors to detect drowsy driving* (Report No. DOT HS 811 886). National Highway Traffic Safety Administration. www.nhtsa.gov/sites/nhtsa.gov/files/811886-assess_veh-based_sensors_4_drowsy-driving_detection.pdf

Brown, T. L., Schwarz, C., Lee, J. D., Gaspar, J., Marshall, D., & Ahmad, O. (2024, May). *Driver monitoring of inattention and impairment using vehicle equipment — detection track: Develop and evaluate a system of algorithms to identify signatures of alcohol-impaired, drowsy, and distracted driving* (Report No. DOT HS 813 564c). National Highway Traffic Safety Administration.

De Valck, E., & Cluydts, R. (2001). Slow-release caffeine as a countermeasure to driver sleepiness induced by partial sleep deprivation. *Journal of Sleep Research, 10*(3), 203–209.

Dinges, D. F., & Grace, R. (1998). *PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance* (Report No. FHWAMCRT-98-006). Federal Motor Carrier Safety Administration. https://rosap.ntl.bts.gov/view/dot/113

Eskandarian, A., Sayed, R., Delaigue, P., Blum, J., & Mortazavi, A. (2007). *Advanced driver fatigue research* (Report No. FMCSA-RRR-07-001). Federal Motor Carrier Safety Administration. https://ntlrepository.blob.core.windows.net/lib/51000/51200/51269/Advanced-Driver-Fatigue-Research-Final-Report-April2007.pdf

Gabel, R. A., & Roberts, R. A. (1980). *Signals and linear systems* (2nd ed.): John Wiley & Sons.

Ji, Q., Lan, P., & Looney, C. (2006). A probabilistic framework for modeling and real-time monitoring human fatigue. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 36*(5), 862–875. https://sites.ecse.rpi.edu/~qji/Fatigue/smc_fatigue_proof.pdf

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26.

LaViola, J. J. (2003). *Double exponential smoothing: An alternative to Kalman filter-based predictive tracking.* Workshop on Virtual Environments 2003, Association for Computing Machinery, Zurich, Switzerland.

Lee, J. D., Fiorentino, D., Reyes, M. L., Brown, T., Ahmad, O., Fell, J., Ward, N. & Dufour, R. (2010). *Assessing the feasibility of vehicle-based sensors to detect alcohol impairment* (Report No. DOT HS 811 358). National Highway Traffic Safety Administration. www.nhtsa.gov/sites/nhtsa.gov/files/811358_0.pdf

Lee, J. D., Moeckli, J., Brown, T. L., Roberts, S. C., Schwarz, C., Yekhshatyan, L., Nadler, E., Liang, Y., Victor, T., Marshall, D. & Davis, C. (2013). *Distraction detection and mitigation through driver feedback* (Report No. DOT HS 811 547A). National Highway Traffic Safety Administration. www.nhtsa.gov/sites/nhtsa.gov/files/811547a.pdf

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems, 8*(2), 340–350.

Lin, C. -T., Wu, R. -C., Liang, S. -F., Chao, W. -H., Chen, Y. -J., & Jung, T. -P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I Regular Papers, 52*(12), 2726–2738.

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research, 40*(4), 239–245.

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers, 36*(2), 339–346.

Marshall, D., & Schmitt, R. (2015). *Participant orientation PowerPoint presentation: DrIIVE track B* (Report No. N2015-005). University of Iowa Driving Safety Research Institute. www.nads-sc.uiowa.edu/publications.php?specificPub=N2015-005

McDonald, A. D., Lee, J. D., Schwarz, C., & Brown, T. L. (2014). Steering in a random forest ensemble learning for detecting drowsiness-related lane departures. *Human Factors, 56*(5), 986–998.

Oron-Gilad, T., & Ronen, A. (2007). Road characteristics and driver fatigue: A simulator study. *Traffic Injury Prevention, 8*(3), 281–289.

Otmani, S., Pebayle, T., Roge, J., & Muzet, A. (2005). Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & Behavior, 84*(5), 715–724.

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Charles, A., & Bioulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep, 28*(12), 1511–1516.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286. www.cs.cmu.edu/~cga/behavior/rabiner1.pdf

Schmitt, R., & Brown, T. (2015). *Driver monitoring of inattention and impairment using vehicle equipment (DrIIVE) – Track B: Assess potential countermeasures for drowsy driving lane departures* (Report No. N2015-013). University of Iowa Driving Safety Research Institute. http://www.nads-sc.uiowa.edu/publications.php?specificPub=N2015-013

Schwarz, C., Brown, T. L., Gaspar, J., Marshall, D., Lee, J., Kitazaki, S., & Kang, J. (2015, June). *Mitigating drowsiness: Linking detection to mitigation.* 24th International Technical Conference on the Enhanced Safety of Vehicles, Gothenburg, Sweden.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. *Accident Analysis & Prevention, 35*(3), 381–391.

Wallis, S. (2013). Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics, 20*(3), 178–208.

Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention, 26*(5), 571–581.

Wilkinson, R. T., & Houghton, D. (1982). Field test of arousal: A portable reaction timer with data storage. *Human Factors, 24*(4), 487–493. doi:10.1177/001872088202400409

Yang, G., Lin, Y., & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences, 180*(10), 1942–1954.

Yang, J. H., Mao, Z. -H., Tijerina, L., Pilutti, T., Coughlin, J. F., & Feron, E. (2009). Detection of driver fatigue caused by sleep deprivation. *IEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 39*(4), 694–705. https://web.mit.edu/coughlin/Public/Publications/YangMaoTijerinaPiluttiCoughlinFeron2009.pdf

***This page is intentionally left blank.***

# Appendix A: Means and Standard Deviations From Driving Performance Measures

*Table A-1. Driving performance measures*

| Measure | No-Mitigation | Discrete Audio/Visual | Discrete Haptic | Discrete Combined | Staged Audio/Visual | Staged Haptic | Staged Combined |
|---|---|---|---|---|---|---|---|
| Drowsy lane departs per min | 0.16 (0.30) | 0.07 (0.12) | 0.06 (0.14) | 0.10 (0.26) | 0.11 (0.26) | 0.03 (0.07) | 0.01 (0.02) |
| Alert lane departs per min | 0.68 (0.27) | 0.62 (0.27) | 0.61 (0.18) | 0.68 (0.43) | 0.99 (0.54) | 0.60 (0.19) | 0.58 (0.24) |
| SDLP (cm) | 30.76 (5.11) | 28.94 (4.53) | 29.92 (4.63) | 27.34 (5.64) | 30.24 (6.84) | 27.37 (4.78) | 25.61 (3.58) |
| Speed in relation to speed limit (mph) | -1.56 (3.24) | -1.66 (1.77) | -2.31 (2.97) | -1.51 (4.43) | -1.42 (2.14) | -2.19 (4.14) | -1.42 (2.14) |
| SD speed in relation to speed limit (mph) | 3.83 (0.93) | 2.84 (0.69) | 2.99 (0.59) | 3.52 (0.98) | 3.30 (0.52) | 3.64 (1.20) | 2.87 (0.90) |
| SSS pre-drive | 4.48 (1.62) | 4.58 (1.51) | 5.19 (1.80) | 4.63 (1.71) | 4.19 (1.94) | 5.31 (1.25) | 5.13 (0.89) |
| SSS post-drive | 5.38 (1.52) | 5.67 (1.23) | 5.50 (1.63) | 5.69 (1.25) | 4.94 (2.17) | 5.69 (1.40) | 5.50 (0.82) |

*Note*. Values are presented as mean (SD).

*Table A-2. Difference scores between mitigation conditions and baseline*

| Measure | Discrete Audio/Visual | Discrete Haptic | Discrete Combined | Staged Audio/Visual | Staged Haptic | Staged Combined |
|---|---|---|---|---|---|---|
| Drowsy lane departs per min | -0.09 (0.12) | -0.09 (0.13) | -0.05 (0.24) | -0.05 (0.26) | -0.12 (0.09) | -0.15 (0.08) |
| Alert lane departs per min | -0.06 (0.27) | 0.00 (0.43) | -0.08 (0.18) | 0.31 (0.54) | -0.08 (0.19) | -0.10 (0.23) |
| SDLP (cm) | -1.83 (4.39) | -0.85 (4.28) | -3.43 (4.86) | -0.52 (6.91) | -3.40 (4.85) | -5.16 (3.54) |
| Speed in relation to speed limit (mph) | -0.10 (2.00) | -0.75 (3.05) | 0.05 (4.39) | 0.14 (2.22) | -0.63 (4.20) | 0.40 (1.61) |
| SD speed in relation to speed limit (mph) | -1.00 (0.68) | -0.85 (0.55) | -0.32 (0.95) | -0.54 (0.51) | -0.20 (1.21) | -0.97 (0.89) |

*Note*. All values represent (mitigation condition – baseline) and are presented as mean (SD).

DOT HS 813 564b
September 2026

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA

16175b-090226-v2