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

NHTSA

DOT HS 813 564a

September 2026

NHTSA's DrIIVE Program: Phase 2 Summary

This page is intentionally left blank.

DISCLAIMER

Suggested APA Format Citation:

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

The other two parts in this series:

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

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 564a		

4. Title and Subtitle	5. Report Date
NHTSA's DrIIVE Program: Phase 2 Summary	September 2026
	6. Performing Organization Code
7. Authors Timothy Brown, Chris Schwarz, John Gaspar, John Lee, Dawn Marshall, & Omar Ahmad	8. Performing Organization Report No. DOT-VNTSC-NHTSA
9. Performing Organization Name and Address National Advanced Driving Simulator University of Iowa 2501 Oakdale Blvd Iowa City, IA 52242	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH22-12-D-00264/0001
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration Office of Vehicle Safety Research 1200 New Jersey Avenue SE Washington, DC 20590	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Digital Object Identifier: https://doi.org/10.21949/2xfn-fa85

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.

Driver impairment from drowsiness, distraction, and alcohol intoxication accounts for a substantial percentage of vehicle crashes. The goal of the NHTSA Driver Monitoring of Inattention and Impairment Using Vehicle Equipment program known as DRIIVE is the detection and mitigation of driver impairment using data available from the vehicle. This project demonstrates that impairment from distraction, drowsiness, and alcohol-intoxication can be detected and, to a large degree, differentiated. Additionally, the project has demonstrated that, for drowsiness, staged alerts can improve driver lane keeping performance at least in the short term. This report is the first of three parts. The detection track and the mitigation track are covered in associated reports.

17. Key Words	18. Distribution Statement
human factors, drowsiness, distraction, driver monitoring, impairment, alcohol intoxication, DRIIVE, detection, mitigation	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	35	

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
DVI	driver-vehicle interface
EEG	electroencephalogram; a measurement of electrical activity in the brain

[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.

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

This page is intentionally left blank.

Introduction

Executive Summary

The goal of the National Highway Traffic Safety Administration's Driver Monitoring of Inattention and Impairment Using Vehicle Equipment program known as DrIIVE was the detection and mitigation of driver impairment. Phase 1 of DrIIVE demonstrated that vehicle-based algorithms could reliably detect driver drowsiness (Lee et al., 2010). Phase 2 of DrIIVE aimed to develop and assess a system of algorithms for detecting and differentiating drowsy, distracted, and alcohol-impaired driving and evaluate potential countermeasures for drowsy driving. The project divides into two research tracks, with the detection track focused on developing a system of algorithms and collecting self-paced distraction data considered to better represent real world in-vehicle distraction than the forced-pace distractions studied in an earlier phase of the present research program (Lee et al., 2013). The mitigation track focused on developing and evaluating a drowsiness mitigation system. The interim reports for these two tracks and their associated appendices provide the complete details of the methods, algorithms, countermeasures, results, and conclusions, and are described in two associated reports.

Background and Motivation

Driver impairment from drowsiness, distraction, and alcohol intoxication accounts for a substantial percentage of motor vehicle crashes. NHTSA reports that 684 fatalities, representing about 1.6 percent of total fatalities in 2021 involved drowsy drivers (Stewart, 2023). However, drowsiness leaves no physical evidence, and so its true contribution might be underestimated. A naturalistic study, for instance, estimated that drowsy driving contributed to 22 percent of the crashes observed in Klauer et al. (2006) and a review of crash data from 2009 to 2013 estimated drowsiness was a factor in 6 percent of crashes requiring a vehicle to be towed and 21 percent of fatal crashes (Tefft, 2014). Blincoe et al. (2023) estimated distraction from a naturalistic observation study and found that distraction was involved in 29 percent of all crashes, resulting in 10,546 fatalities, 1.3 million nonfatal injuries, and $98.2 billion in economic costs in 2019. Finally, there were 13,384 alcohol-impaired driving fatalities in 2021, an increase from 2020 (Stewart, 2023). Individually, and in combination, drowsiness, distraction, and alcohol intoxication represent a threat to driving safety that may be addressed through vehicle systems. These systems intend to detect and classify an impaired driver state and invoke appropriate and effective interventions.

General impairment-detection algorithms and warning systems that use existing low-cost vehicle-based sensor suites may be one implementation option. The challenge to such an approach lies in creating algorithms that effectively detect different types of impairments and differentiate them from one another. For example, because alcohol is a depressant and most drunk driving occurs during nighttime and early morning, intoxication often co-occurs and exasperates drowsiness, presenting additional challenges in classifying impairment types.

Two main criteria for the performance of algorithms were used throughout this project. The first compares a whole family of algorithms obtained by varying one of its two inversely related outcomes, the correct detection rate or the false detection rate. The receiver operator characteristic (ROC) is a way of analyzing such a family of algorithms and has very useful tools associated with it (Brown & Davis, 2006). The ROC curve is a simple visualization of performance across the whole family. A single model is associated with an operating point on the

ROC curve and has a confusion matrix and several other measures that relate to the number of true and false categorizations the algorithm makes. The second criteria describes the timeliness of the algorithm, or its ability to issue a prediction of impairment soon after the onset of impairment that increases the likelihood of a hazardous event such as an unintended lane departure.

The DrIIVE projects have focused on vehicle-based sensor data; however, they also incorporated driver-based sensor signals and compared the results to their vehicle-based counterparts. Additionally, the benefits of taking individual differences between drivers into account in the training and operation of an algorithm were investigated. Driver-based sensors provided an added benefit to the performance and generalization of the distraction-detection algorithm, while individualizing the algorithms provided an added benefit to a drowsiness algorithm and an alcohol-intoxication algorithm.

DrIIVE phase 2 created algorithms for implementation in real-time, in a driving simulator or in a production vehicle. It also added a real-time implementation of a drowsiness-mitigation system and a simulator study aimed at testing its effects. The design of optimal warning interfaces for each type of impairment is an important problem for the detection and mitigation system. The phase 2 study examined two types of mitigation systems (discrete versus staged) and three warning modalities (audio/visual, haptic, and combined).

An earlier distraction simulator study in this research program included several instances of a reaching task, a visual-manual matching task, and a cognitive (auditory) menu information task presented to the driver in a force-paced manner that restricted their ability to delay engagement with the tasks (Lee et al., 2013). A suite of eye-based distraction algorithms characterized driver distraction and exemplified systems that use driver-based sensors. These results motivated a phase 2 simulator study that used self-paced secondary tasks to elicit more natural driver behaviors.

During phase 1, the drowsiness simulator study drives occurred during the day, early night, and late night. A variety of vehicle-based algorithms were developed, and ultimately, a random forest algorithm proved to be able to predict drowsiness 6 seconds in advance of drowsy lane departures (Brown et al., 2014). The mitigation track interim report described the video rating procedure used to identify drowsy lane departures. These results motivated the phase 2 project to improve the drowsiness algorithm, implement a real-time version along with countermeasures, and collect new data with early night and late-night mitigation conditions.

Prior to DrIIVE, the Impairment Monitoring to Promote Avoidance of Crashes Using Technology simulator study tested drivers with blood alcohol concentrations of .00, .05, and .10 g/dL (Lee et al., 2010). Algorithms were developed to distinguish drivers with BAC levels greater than .08 g/dL from those with the lower levels. However, the algorithms were not suited for real-time implementation and used inconsistently sized time segments. The results motivated the phase 2 project to revisit the IMPACT simulation dataset and create an algorithm that could be real-time and possibly differentiate between intoxication and other kinds of impairment.

Research Facility and Simulated Driving Scenario

The phase 2 research was completed in two tracks: (1) detection track, the focus was to develop a system of algorithms to detect three types of driver impairment and differentiate between them; and (2) mitigation track, the focus was to design drowsy-driving countermeasures and assess their effect on mitigating drowsy-driving lane departures.

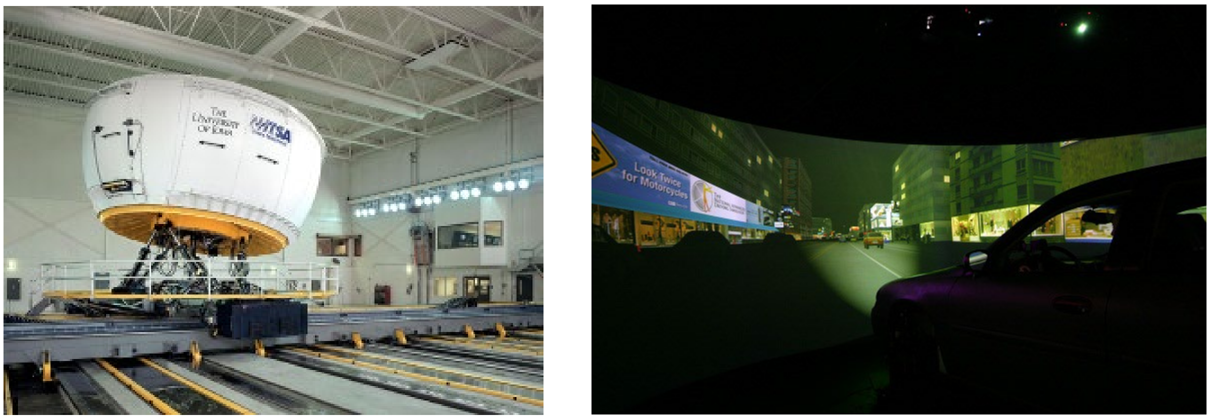

Figure 1. Images of the exterior and interior of the NADS-1 simulator

Data collection occurred in the National Advanced Driving Simulator at the University of Iowa, shown in Figure 1. The NADS-1 simulator consists of a 24-foot-diameter dome in which an entire vehicle is mounted. 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 as if driving a real vehicle. The cab is equipped with a Face Lab 5.0 (Seeing Machines) eye tracking system that is mounted on the dash in front of the drivers' seat above the steering wheel. The steering wheel and pedals have active feedback providing an appropriate amount of torque or force against the drivers' control inputs.

Drivers participated in a standard set of nighttime scenarios in the IMPACT simulation and phases 1 and 2 of the DrIIVE program including the detection track and mitigation track experiments. The scenarios start with an urban environment, shown in Figure 1, followed by an interstate highway environment with both straight and curved road portions. The drive ends in a rural environment that includes dark, lighted, straight, curved, gravel, and paved sections. In the mitigation track, the scenario included a final 10-minute drive on a section of straight rural roadway to increase the likelihood of drowsy lane departures. These scenarios provide a common set of events that every driver experienced, making it possible to compare driving performance across drivers and types of impairment.

Each drive lasted approximately 45 minutes and started with an urban segment, 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 a few slower-moving trucks. The drives concluded with another rural segment, this one 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. The drive was segmented into 19 “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. Light traffic was present in the urban and interstate sections of the drive.

Detection Track: Develop and Evaluate a System of Algorithms to Detect and Differentiate Distracted, Drowsy, and Alcohol-Impaired Driving

Objectives

The main objective of the detection track was to develop a system of algorithms that could detect and differentiate different types of driving impairment: distraction, drowsiness, and alcohol intoxication. Vehicle-based sensor signals were the preferred inputs to the algorithms. Driver-based signals were added to determine the extent they improved algorithm performance. Individual differences among drivers were also considered through an approach that standardized each participant's vehicle and driver-based signals.

The detection track also involved collecting data of drivers engaged in self-paced secondary tasks. A review of visual and manual secondary tasks in the laboratory and naturalistic driving literature was conducted to identify appropriate tasks. The detection track data collection included visual and visual-manual secondary tasks that yielded task engagement behaviors that were relatively realistic and natural because they were self-paced. This section describes the experiment and algorithms and summarizes results of the project.

Data Collection and Analysis

Method

Twenty-four participants 21 to 34 years old completed all three drives of the study in the NADS-1 in a single session—one with no distraction, one with moderately difficult versions of visual and visual-manual tasks, and one with difficult versions of those tasks. Age and gender driver characteristics matched those from the phase 1 distraction study. The order of the conditions was counterbalanced across participants.

Drivers engaged in two types of distraction tasks: a *visual-only* task and *visual-manual* task. For the *visual-only* task, drivers read aloud text about "interesting facts" from the display. Each message was no longer than 160 characters. Message complexity and readability determined the task difficulty. Moderate-difficulty messages were easily chunked by line and had no errors, while high-difficulty messages contained more words and run-on sentences, were not easily chunked, and contained abbreviations and spelling and grammatical errors.

For the *visual-manual task,* drivers scrolled through a list of musical pieces to find a target piece that the experimenter provided verbally to the driver (Lee et al., 2012). The driver selected the correct piece by touching its name on the display. Incorrect selections turned red and the task remained open. Correct selections turned blue and the display cleared. Moderate difficulty lists consisted of a single page of relatively distinct pieces. High difficulty lists extended to several pages and the target and distractor items on the list were relatively similar. The display was implemented such that it tilted toward the drivers, and they were able to perform the tasks by turning their head away from the road toward the display.

Dependent measures included standard deviation in lane position, mean and standard deviation in speed relative to the speed limit (normalized), lane departures (per minute), and the percent of time speed was 10 percent above or below the posted speed limit (percent speed high and low).

Results

Secondary task difficulty levels were verified by comparing driving performance during the events that included the tasks. This approach is more conservative than only examining performance during task execution periods because the event level data also contains periods where drivers were not engaged with the secondary task. Events consisted of distinct driving epochs that may have also included something unique, such as a yellow-light dilemma event in the urban segment. Events ranged in duration from half a minute to several minutes (Lee et al., 2010). Tasks were included within events and consisted of distinct periods of time in which the distraction tasks were available to the driver. Secondary tasks were included in all events. Task duration varied because the tasks were self-paced, but it was always shorter in duration than events. Driving performance included data for events that also included moderate and difficult distraction tasks and included corresponding event data from the baseline drive.

The results indicate that both levels of the secondary task led to degraded driving performance in at least some comparisons to the undistracted driving condition. Specifically, degraded performance between each of the baseline no distraction moderate distraction case, and the difficult distraction case was found for SDLP. Greater SDLP indicates more variability in lane position and less precision in lane keeping. Table 1 presents effect sizes for the effects of moderate and difficult distraction on driving performance relative to no distraction when considered at the timescale of each event (e.g., urban drive) for the seven events that contained secondary tasks for both the moderate and difficult conditions.

Table 1. Cohen's d effect sizes for moderate and difficult distraction conditions relative to the no distraction condition

	Urban Drive	Yellow Light Dilemma	Urban Curves	Following	Interstate Curves	Dark Rural	Gravel Rural
Moderate Distraction							
SDLP	-0.04	1.32	0.65	0.13	0.12	0.01	0.13
Speed relative to speed limit	0.10	0.20	-0.07	-0.16	-0.04	-0.19	-0.16
Lane departures/min	NA	NA	NA	0.16	0.13	0.36	NA
% speed high	-0.18	-0.14	-0.01	-0.20	-0.16	0.15	-0.42
% speed low	0.33	-0.06	0.17	0.05	-0.17	0.22	-0.23
Difficult Distraction							
SDLP	0.36	0.35	0.32	0.41	0.82	0.39	0.69
Speed relative to speed limit	-0.47	-0.43	-0.46	-0.42	-0.36	-0.30	-0.34
Lane departures/min	NA	NA	NA	0.39	0.55	0.82	NA
% speed high	-0.32	-0.51	-0.22	-0.16	-0.25	-0.26	-0.43
% speed low	0.53	0.15	0.57	0.51	0.39	0.30	0.16

Note. Signs represent the direction of the effect relative to no-distraction (e.g., negative values indicate a lower result than no distraction). 0.20 = small effect, 0.50 = medium effect, 0.80 = large effect.

Distraction-Detection Algorithm

Vehicle-based sensor signals from the study data were used to train a distraction-detection algorithm, which was then augmented using driver-based signals to compare their performance. NHTSA's visual-manual distraction guidelines state in-vehicle secondary tasks should be designed to take no longer than 2 second glances away from the road, as well as total glance time toward a secondary task longer than 12 seconds (NHTSA, 2013). Analysis of crashes and near-crashes from the SHRP 2 dataset reveals the particularly risky nature of long glances (Victor et al., 2015).

The AttenD algorithm was used to define ground truth of distracted driving. AttenD uses gaze location away from the roadway to indicate distraction. The algorithm counts single long glances as well as glances that are closely spaced in its distraction criteria. AttenD accumulates eyes-offroad time and indicates distraction when a threshold is crossed. Figure 2 shows an example of the AttenD output, whose units are in seconds. The shaded regions indicate periods of task engagement, and distraction occurs when the AttenD value exceeds a threshold of 2.0 seconds. Note that it is possible to be engaged in a task and not distracted, as in the first part of the third task on the right side of Figure 2. It is also possible to be distracted while not engaged in the tasks, as in the short periods on the left side of Figure 2.

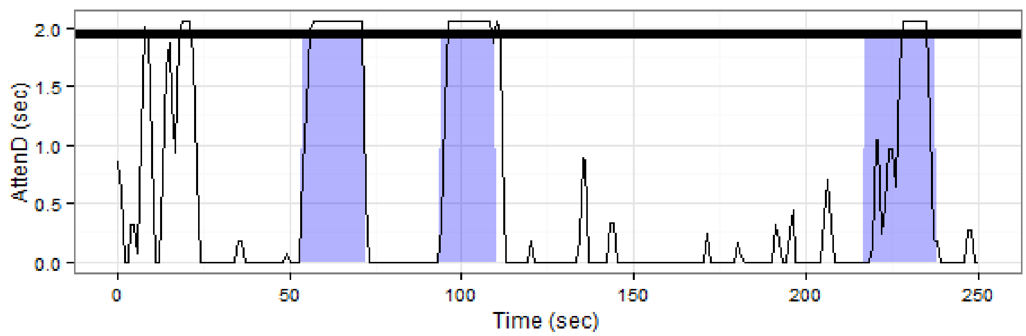

Figure 2. AttenD distraction metric defines the ground truth of distraction as values over 2.0, indicated by the bold horizontal line

A vehicle-based distraction-detection algorithm was developed using a framework like the one used for the drowsiness algorithm studied in the mitigation track. A random forest algorithm was trained to detect drowsy driving from a training dataset consisting of data from 18 drivers (75% of the sample). Algorithm performance was assessed with data from the remaining six (25% of the sample) drivers. The output of the random forest fed a hidden Markov model to issue a posterior probability indicating that the driver has been distracted. The HMM algorithm classification switched from attentive to distracted when the posterior probability from the HMM exceeded 0.5. A switch from attentive to distracted could trigger a warning to the driver; but no distraction warnings occurred in this study. The final distraction algorithm performed well, with an area under the ROC curve of 0.82 and a 95 percent confidence interval of [0.81, 0.84]. A higher AUC is better with a value of 1.0 indicating perfect detection performance. The following section provides more information on the machine learning methods.

DrIIVE System to Classify Driver Impairment

There are several ways a system of algorithms might detect, differentiate, and mitigate driver impairment. One approach is a layered framework, shown in Figure 3. 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 indicate impairment at the same time. Therefore, an arbitration layer determines the best way to mitigate the detected impairment(s), such as an alert issued in a particular modality. The driver-vehicle interface for the alert depends greatly on which type of impairment is detected, the roadway environment, and perhaps on the duration and severity of the impairment.

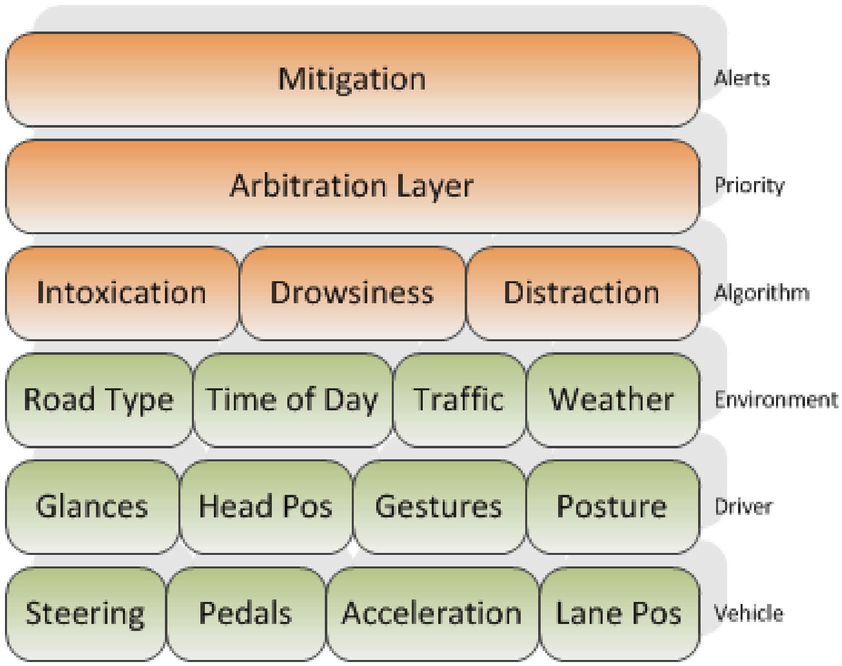

Figure 3. Layered framework for detecting, differentiating, and mitigating several impairments

The impairment detection algorithms had two components. First, a random forest algorithm 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 4).

The second component of the algorithms was an HMM that combined the random forest observations over a brief time to augment the driver state classification by assigning probabilities regarding whether the system remains in each state or transitions to a new state at each moment in time. The HMM uses state history to predict the current drivers' 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 5 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.

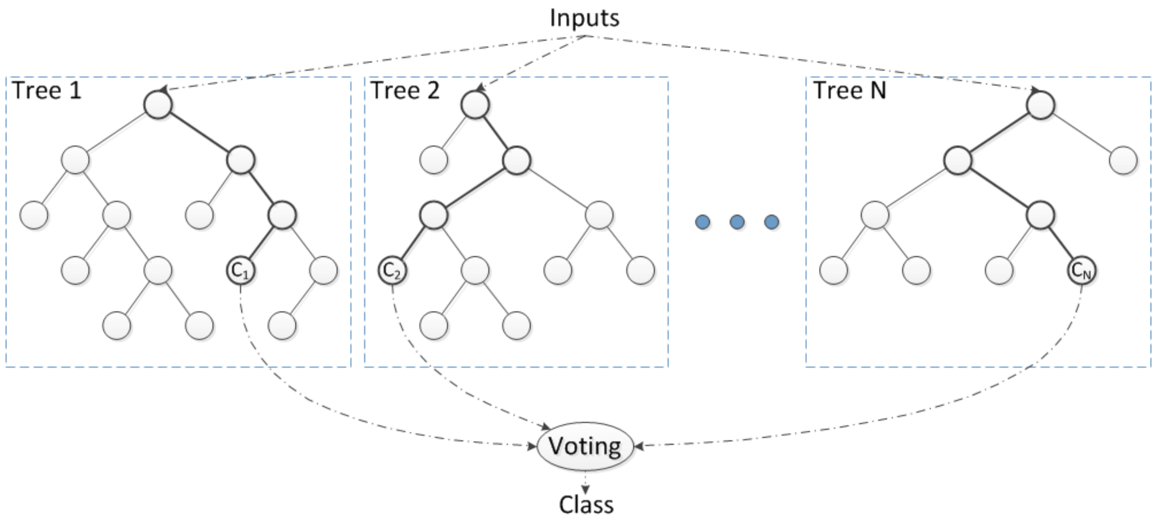

Figure 4. Random forest machine learning method uses voting to combine the output of many decision trees

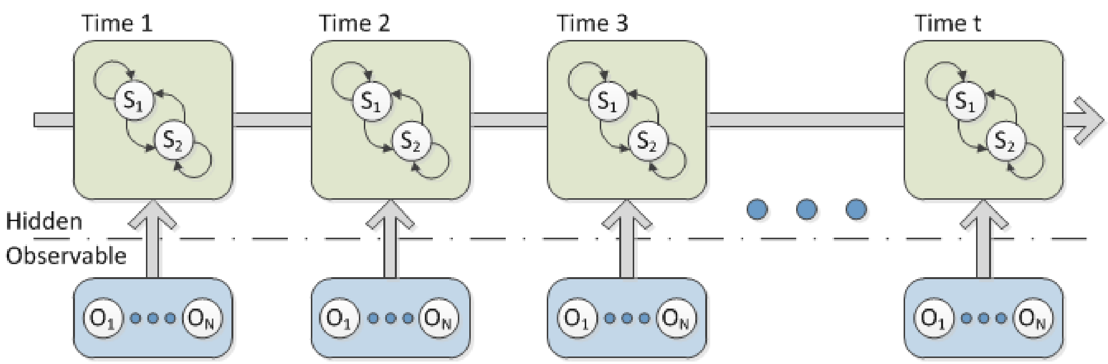

Figure 5. HMM combines the history of "observed" output of the random forest illustrating how the algorithm uses past states to estimate drivers' current state

Data from driving under the influence of each of the three driver impairments (drowsy, intoxicated, and distracted) and data from unimpaired driving was used to train separate algorithms that combined random forests with HMMs. The three impairment algorithms were then evaluated with data where the driver state matched the intended purpose of the algorithm (e.g., applying the drowsiness-detection algorithm to drowsy drivers), as well as across impairments (e.g., applying the drowsiness-detection algorithm to distracted drivers). This cross-impairment evaluation used ground truth for all the datasets. For example, instances of drowsiness in the distraction dataset were identified to cross-evaluate the drowsiness-detection algorithm with distracted driving.

Table 2 shows the ground truth for each combination of impairment algorithm and dataset. The AttenD algorithm determined the presence of distraction in each of the datasets. Manual coding of the video after the drives determined the presence of drowsy lane departures, but only in the drowsy driving dataset. To enable the evaluation of the alcohol and distraction algorithms' sensitivity to drowsiness in the other datasets it was assumed that there were no cases of drowsiness in either the distraction dataset, where this was likely the case, or in the alcohol-impairment datasets, where there likely was some drowsiness.

Although distraction and drowsiness can co-occur while driving, for the purposes of these analyses, assuming the absence of drowsiness in the distraction dataset is reasonable because these data were collected in the daytime and there were secondary tasks regularly spaced throughout the drives. However, it cannot be assumed that there is no drowsiness in alcohol-impairment because the participants drove late at night, and some were under the depressive effects of alcohol. Consequently, when the drowsiness detection algorithm indicates alcohol-impaired drivers are drowsy it may be a correct classification. The assumption of no alcohol intoxication in the other two datasets is reasonable because of the tightly controlled subject-research protocols.

Table 2. Ground truth used for each algorithm/dataset combination

Algorithm	Distraction data	Drowsiness data	Intoxication data
Distraction	AttenD	AttenD	AttenD
Drowsiness	Assume none	Video coded lane departures	Assume none
Intoxication	Screened out	Screened out	BAC level

The effect of including driver-based signals (i.e., head pose) into the algorithm, as well as the effect of removing the effects of individual differences among drivers was evaluated. The driver-based signals were head-pose metrics from a research-grade eye tracker (FaceLab v5). To evaluate the removal of individual difference effects, z-scores were calculated for each measure for each driver. A z-score is calculated by subtracting the drivers' mean score on a metric and dividing by its standard deviation. The degree to which the z-score transformation affects algorithm performance indicates the influence of individual differences. Head-pose signals and z-score adjustments improved the AUC measure of some, but not all the impairments.

Head pose improved distraction detection, improving the area under the ROC curve from 0.82 (95% CI: [0.81, 0.84]) to 0.92 (95% CI: [0.92, 0.93]). However, z-score normalization did not. The drowsiness algorithm benefited the most from considering individual differences. Drowsiness detection without z-score normalization performed relatively poorly, with an AUC of only 0.64 (95% CI: [0.56, 0.71]). However, z-score normalization improved performance to 0.83 (95% CI: [0.77, 0.89]). Including head-pose data further increased performance to 0.89 (95% CI: [0.85, 0.93]). Finally, the intoxication algorithm benefited from the z-score normalization. This normalization takes advantage of unimpaired driving data for each person to improve sensitivity. Although this improves performance, it does create technical challenges outside the laboratory environment where driver BACs are not available to normalize driving performance. The base algorithm had an AUC of only 0.50 with 95 percent CI of [0.45, 0.56]; however, with normalized inputs, the AUC increased to 0.78 (95% CI: [0.75, 0.82]). Head-pose data failed to improve the intoxication algorithm.

Applying the three algorithms to each of the three datasets tests the ability of the algorithms to differentiate the impairments, that is, it indicates their cross-impairment classification performance. For example, the drowsiness-detection algorithm should not indicate any instances of drowsiness in the distraction data if it is not present. This evaluation produced mixed results. The distraction algorithm generalized to the other impairment datasets better with a driver-based component than without it. The head-pose version of the distraction algorithm yielded an AUC of 0.8 (95% CI: [0.80, 0.82]) with the drowsy driving data, and an AUC of 0.87 (95% CI: [0.87, 0.91]) with the alcohol-impaired driving data. The vehicle-based algorithm generalized poorly to the drowsy and alcohol datasets having AUCs of 0.58 (95% CI of [0.56, 0.59]) and 0.73 (95% CI of [0.70, 0.76]) respectively.

The intoxication algorithm performed well on the distraction data, generating no false positives. AUC was not reported for the cross-impairment performance of the intoxication algorithm because the full array of ROC measures was not available in the absence of true positive cases. The intoxication algorithm, using only the random forest, also performed well on the drowsiness dataset with a specificity of 0.91 with 95 percent CI of [0.89, 0.95]. Unfortunately, including the HMM degraded performance to a specificity of 0.11 (95% CI: [0.08, 0.15]).

Finally, the drowsiness model performed poorly with the distraction data, having a specificity of only 0.38 (95% CI: [0.32, 0.46]). If drowsiness was absent from the distraction dataset, as was assumed, then the target metric for perfect performance was a sensitivity of one. Including environmental inputs improved the drowsiness algorithm. Specifically, including time of day reduced false alarms because the distraction data were collected during the day. The drowsiness detection algorithm had a similarly low specificity when applied to the intoxication dataset, but the interpretation is more difficult because episodes of drowsiness likely occurred at a higher rate in this dataset.

For the purposes of these analyses, timeliness of each algorithm was defined as how quickly it can detect impairment after it appears. Timeliness is different for the three impairment types due to the unique characteristics of the target driver states as reflected in the different time scales built into the algorithms. Timeliness for the distraction algorithm assumes that distraction can begin very quickly and abate just as quickly, and that it is difficult to predict ahead of time. Thus, the distraction algorithm was trained on sequential 1-second chunks of data. The 1-second aggregation gives the distraction algorithm good timeliness in line with the time scale of the impairment itself.

Timeliness for the drowsy driving algorithm assumes that drowsy driving consists of acute critical episodes within a chronic or persisting degraded driving state (Grace & Stewart, 2001). Accordingly, the detection track and mitigation track drowsiness algorithms operate on a rolling 1-minute chunk of data that updates every 6 seconds. Timeliness is potentially constrained by the minute during which the algorithm accumulates data before it can issue a reliable classification as well as by its update rate. The timeliness of these algorithms is consistent with the desire to predict a drowsy lane departure 6 seconds prior to its occurrence.

Humans reach their peak breath alcohol concentrations in 10 to 90 minutes following ingestion (Friel et al., 1995) suggesting a slower time course for alcohol-impaired driving effects. Consistent with this time scale, the alcohol intoxication algorithm accumulates data in rolling 5-minute chunks that update every 2.5 minutes. This means that it would take longer for the algorithm to issue its first prediction and that its initial resolution is lower; however, the resulting timeliness is consistent with the slowly varying effects of alcohol intoxication.

Mitigation Track: Assess Potential Countermeasures for Drowsy Driving Lane Departures

Objectives

The mitigation track simulator study investigated the effect of alert interface characteristics on mitigation effects. To that end, six drowsiness mitigation alerts were compared that varied along two dimensions: the number of stages in the alert: a single alert (discrete) or a multiple-stage alert (staged), and the modality of the alert (audio/visual-manual, haptic, or combined audio/visual-manual plus haptic). To test the effects of mitigation at different levels of drowsiness, drivers completed both an early night (moderately drowsy) drive and a late night (highly drowsy) drive.

Methodology

Forty-eight participants 21 to 34 years old completed two drives, during two visits, in the NADS-1. 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). It was anticipated that the early night drives would lead to moderate drowsiness, and the late-night drives would lead to severe drowsiness, and this assumption was confirmed via analysis of subjective survey responses before and after the study drives. During both drives, participants experienced one of the six drowsiness alert interfaces, as seen in Table 3. These interface components are described below.

The only difference between this study and the phase 1 study 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.

Table 3. Mitigation type and alert modality

	Audio/Visual-Manual	Haptic	Audio/Visual-Manual + Haptic
Staged	N=8	N=8	N=8
Discrete	N=8	N=8	N=8

The driving scenario was the same as described in Section 1.3, except for an additional 10-minute drive on a section of straight rural roadway with no other traffic.

Specific hypotheses regarding drowsiness mitigation included these.

- Drowsiness mitigation alerts will reduce the rate of drowsy lane departures and improve lane keeping relative to drowsy driving without mitigation
- Staged alerts will reduce lane departures more than single-stage discrete alerts
- Warning effect will vary by modality
- Drowsiness mitigation will not reduce subjective drowsiness

Mitigation System

Both the discrete and staged warning types used three warning modalities: audio/visual-manual, haptic, and combined haptic with audio/visual-manual. The drivers were instructed to clear the visual-manual alert when it appeared. The visual icon persisted until drivers pressed the button. The discrete alert was binary and triggered when the algorithm detected drowsiness. The visual component of the alert was the middle, yellow icon in Figure 6 (Schwarz et al., 2015). It displayed on a black instrument panel. 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 drivers' seat bottom.

Figure 6. The visual icons presented for audio/visual and combined alerts

The staged alert consisted of three binary alerts that 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 6. The changing colors, from white to yellow to red, indicated the increasing urgency of each stage. The audio component had three sounds, one for each icon. The first stage 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 developed under the mitigation track determined when the alerts initially occurred. The main difference between the detection and mitigation track drowsiness algorithms was that the latter used two separate random forest algorithms for steering and pedal signals. Figure 7 shows the stages of the staged alert; and the first 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.

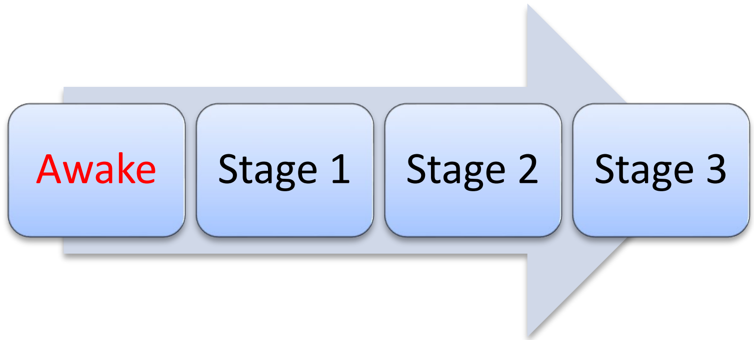

Figure 7. The staged alerts escalate from awake to stage three when the drowsy state persists

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 three, then an escalation simply re-triggered the alert for that stage.

Results

Effect of Drowsiness Mitigation

Did the presence of drowsiness mitigation reduce the frequency of drowsy lane departures, improve lane keeping, and decrease subjective drowsiness relative to the no-mitigation group from phase 1? As seen in Figure 8, the presence of mitigation resulted in a marginally significant reduction in drowsy lane departures compared to no-mitigation for both early and late-night drives. This reduction in drowsy lane departures for the mitigation groups was accompanied by significantly better lane keeping during both early and late-night drives as evidenced by a reduction in standard deviation of lateral position, shown in Figure 9. As noted earlier, larger SDLP indicates greater variability and less precision in lateral lane keeping. Subjective drowsiness, as measured by the Stanford Sleepiness Scale, showed no effect of the presence of the drowsy driving alerts.

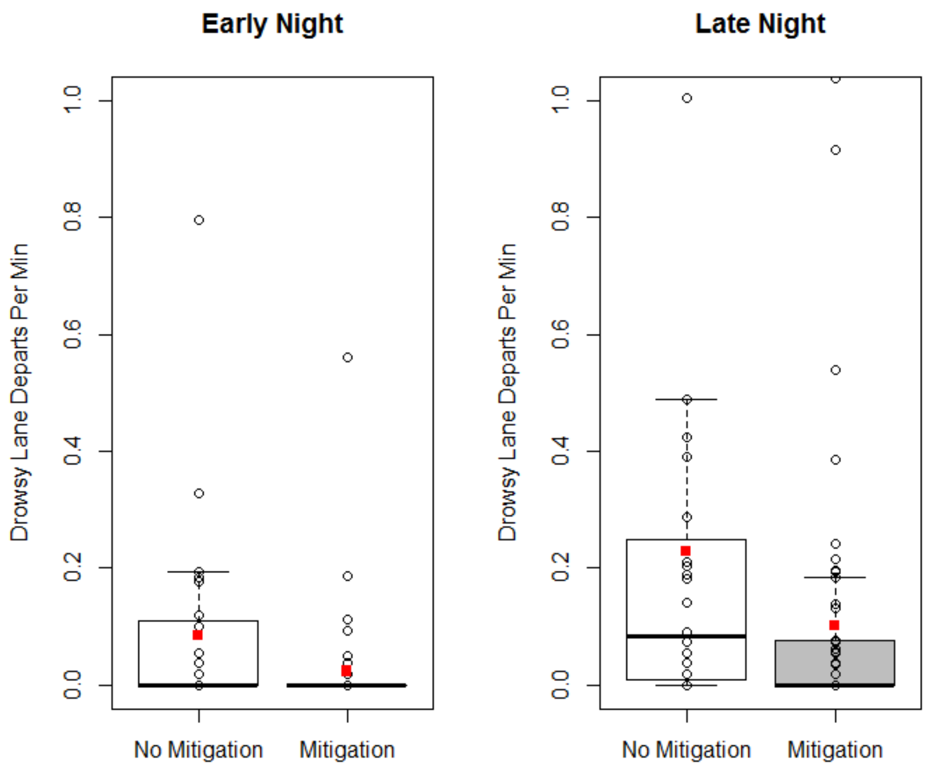

Figure 8. *Boxplots of drowsy lane departures per minute in the early and late drives.*
Note: Circles represent individual participant means and red squares indicate group means. Boxplots represent the median (dark bar), first and third quartiles (box edges), and min/max values. Points beyond the horizontal bars represent outliers.

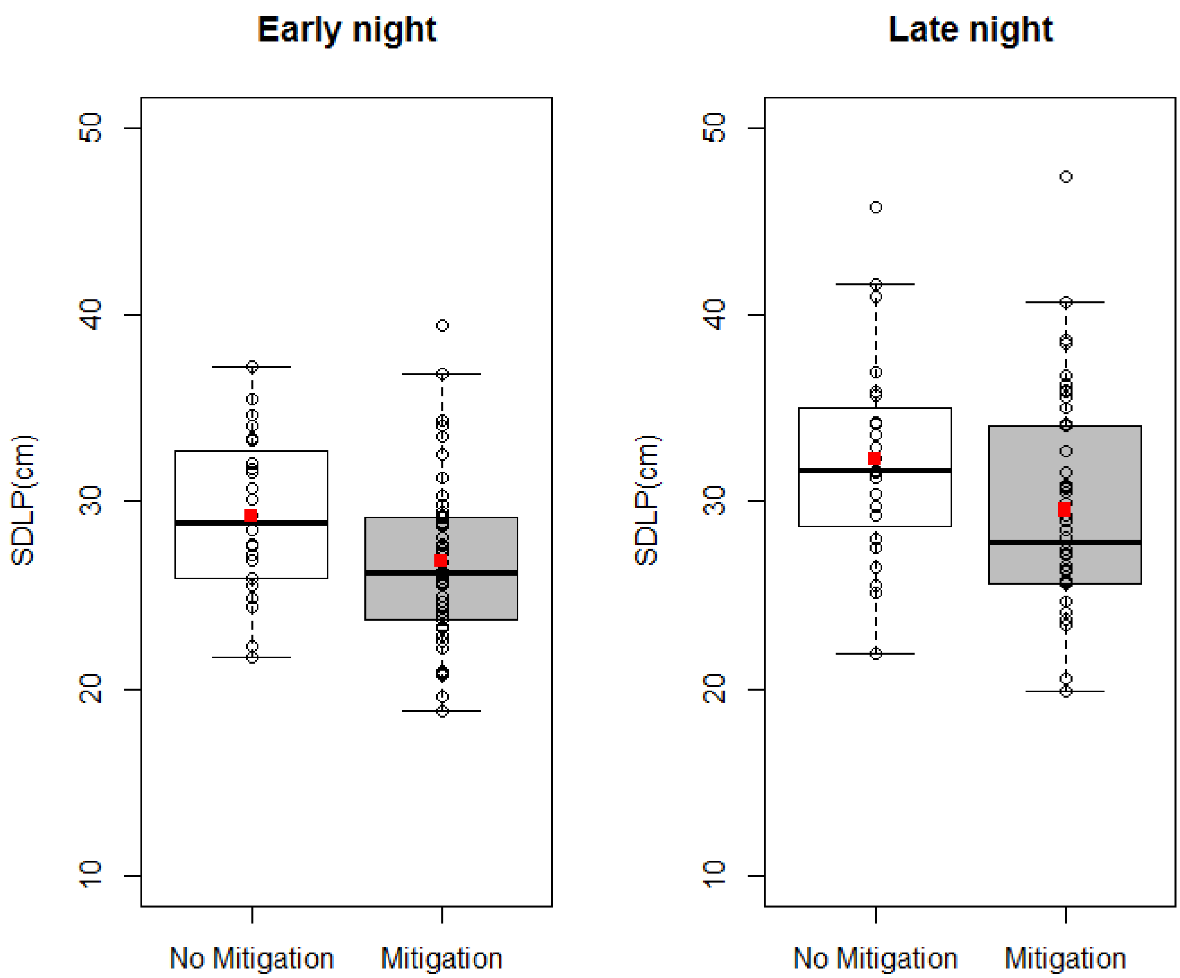

Figure 9. Boxplots of SDLP for early night and late-night drives.
Note: Circles represent individual participant means and red squares indicate group means. Boxplots represent the median (dark bar), first and third quartiles (box edges), and min/max values. Points beyond the horizontal bars represent outliers.

Influence of Mitigation Characteristics

How did interface characteristics, specifically mitigation type (discrete or staged) and alert modality (audio/visual-manual, haptic, or combined audio/visual-manual plus haptic), influence driving performance relative to that of the baseline group with no alerts?

To address this question, for each performance measure, difference scores were computed by subtracting the mean of the baseline group from each individual mitigation participant's mean for early and late-night drives separately. These difference scores 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, the difference score would be negative one, indicating an improvement relative to baseline of one fewer lane departure per minute. If difference values were significantly more negative as a function of mitigation type, alert modality, or both, this would provide evidence that one or more alert interfaces were more salient, relative to baseline, than others.

As seen in Figure 10, there was a significant interaction between drive time and mitigation type indicating that staged alerts resulted in fewer lane departures than discrete alerts relative to baseline during the late-night drives. A significant interaction also occurred for SDLP, indicating that the staged alerts reduced lane variability more than discrete alerts during late-night drives. There were no significant differences among alert modalities.

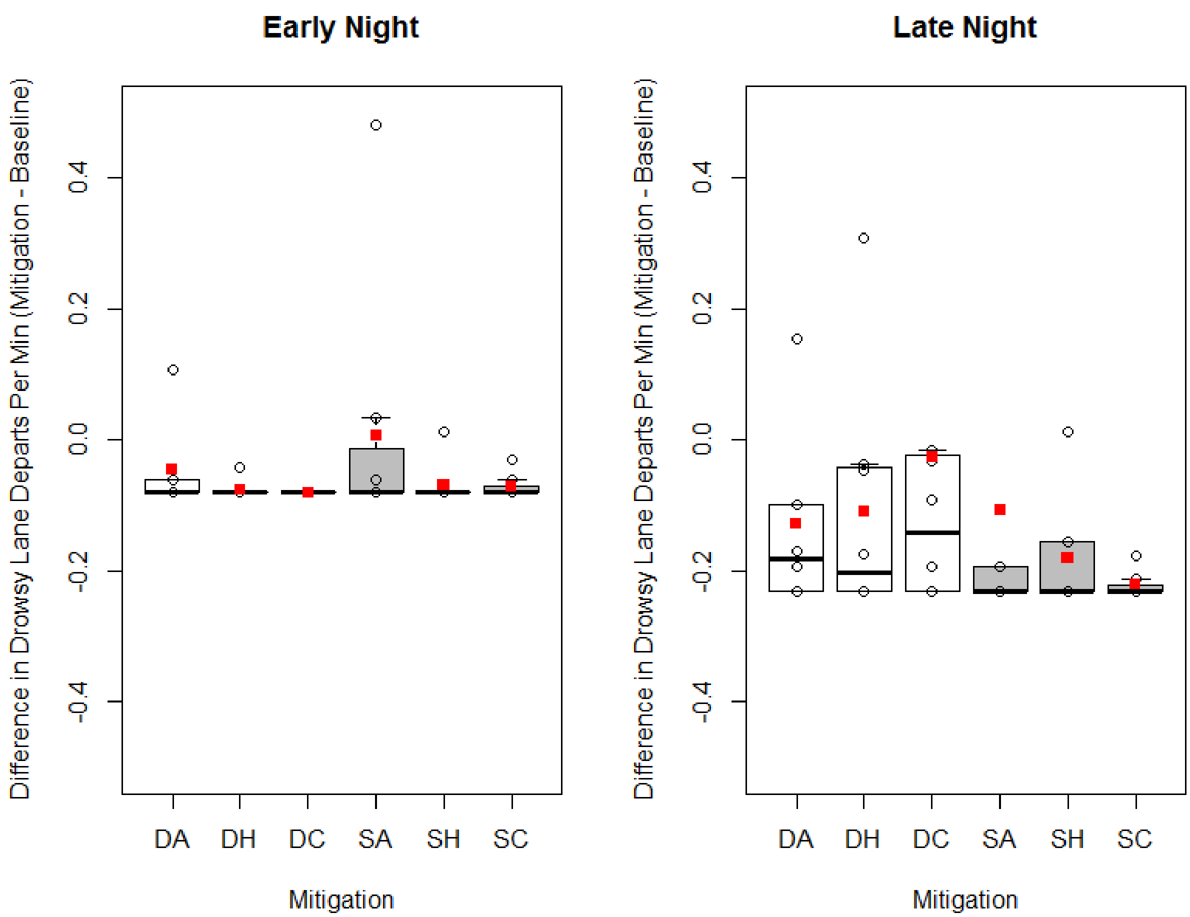

Figure 10. Boxplots of difference in drowsy lane departures per minute between each mitigation condition and the corresponding mean baseline value from Brown et al. (2014).

Note: Lower values reflect fewer drowsy lane departures 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. Boxplots represent the median (dark bar), first and third quartiles (box edges), and min/max values. Points beyond the horizontal bars represent outliers. The use of mean baseline values (Brown et al., 2014) results in a restriction in range that can be observed with all conditions within a drive time having the same minimum values.

Conclusions

Phase 2 of the DrIIVE program had two goals. The detection track aimed to develop a system of algorithms that could be useful in detecting and differentiating several impairments. To achieve this goal, data was collected and validated from the performance of drivers engaged in self-paced visual and visual-manual secondary tasks. The mitigation track addressed the ability of drowsiness alerts in reducing drowsy lane departures. It also addressed the effect of warnings for undistracted drowsy drivers who had received drowsy driving alerts.

The self-paced distraction tasks used in the detection track degraded driving performance, much like the forced-pace tasks from an earlier study in this line of research (Lee et al., 2013), but presumably with more natural task engagement patterns. Impairment-detection algorithms were trained using the distraction data from this study as well as drowsiness and alcohol-intoxication data from previous studies. Each type of algorithm exhibited some efficacy in detecting its intended target impairment. The inclusion of head pose as a driver-based sensor input improved the distraction detection algorithm. Both the drowsiness and intoxication algorithms benefitted from the use of z-score normalization to remove individual differences. One might expect that head pose data would include nods that are indicative of drowsiness. Either there were not enough of these nods to be useful to the algorithm, or there were other head movements throughout the drive that also looked like nods. Why did the use of z-scores not benefit the distraction algorithm? It may be that the short, yet intense, tasks that represent serious distraction impose a kind of structural uniformity to the drivers' performance decrements. While not looking at the road, they compensate as best they can, and the compensation strategies do not vary much from driver to driver.

The algorithms had mixed results when generalizing to other impairment datasets. The distraction algorithm generalized much better when it included head pose data. Most types of visual distraction have some degree of head-turning behavior. However, it could be that the characteristics of driving performance decrements vary widely with the specifics of a secondary task, resulting in poor generalization for the vehicle-based algorithm. Tying this to the conclusion regarding z-scores, the performance decrements observed through vehicle-based sensors are consistent with a dependence on the task but are relatively uniform across drivers. They may be caused by looking and reaching, and by using limited proxy cues that substitute for not looking at the road. On the other hand, the intoxication-detection algorithm generalized to the distraction dataset and to the drowsiness dataset. However, it would have issued some false positive classifications to the drowsy driver. Due to the presumed distinct nature of an intoxication countermeasure, the rate of false positives is an important consideration. Finally, the drowsiness algorithm generalized rather poorly to the other impairment datasets and exhibited similar results on two datasets that were assumed to have quite different levels of actual drowsiness. The reasons for this were unclear and could have been due to many factors.

Obtaining robust indicators of ground truth for cognitive distraction and drowsiness is difficult. It is difficult to identify measures that capture the physiological processes at play in the presence of impairment. Glances away from the road seem to capture visual distraction, but do not address cognitive distraction. Drowsiness is an internal state and can be difficult to differentiate from general fatigue. Moreover, it can be costly to measure ground truth. This research adopted a practical definition of drowsiness ground truth related to safety-critical events caused by drowsiness episodes.

The mitigation track focused on the use of the drowsiness detection algorithm to trigger drowsy driving alerts. A combination of random forest algorithms and HMMs produced an algorithm that significantly improved drowsy driving classification performance compared to the random forest model from phase 1. Drowsiness alerts from a developed mitigation system produced marginal reductions in drowsy lane departures per minute, and significantly reduced SDLPs compared to no-mitigation, in both the early and late drives. These findings suggest that the drowsiness mitigation affected behavior over the course of a short drive where the drivers' only choice was to continue driving.

The results also suggest that relative to results from a no-alert baseline condition, staged alerts reduced lane departure more than discrete alerts during late-night drives, where drivers reported greater sleepiness. Staged alerts showed significantly more reduction than discrete alerts in the rate of drowsy lane departures as well as SDLP relative to no mitigation during the late-night drives. This provides some evidence that staged alerts may be better able reduced lane departures than discrete alerts over the course of relatively short drives, particularly when drivers are very drowsy. However, the staged alerts had escalated salience relative to the discrete alerts, which could account for these differences. There was no effect of warning modality on driver behavior. Alert type may be more effective than alert modality for producing short-term improvement of drowsy driving performance.

Limitations

Several limitations are worth noting in the present research. As the data indicate, there was significant variability among people in their susceptibility to the effects of both distraction and drowsiness on driving performance. Additional understanding of the nature of individual differences and their impact on impairment will be important for future refinement of state detection algorithms. Additionally, the similarity of intoxication and drowsiness remains a significant challenge. It has been difficult to tease out the unique signature of each type of impairment for algorithm development.

Ground truth accuracy and availability are important considerations for this line of research. Types of ground truth that use manually intensive coding procedures are costly and time-consuming to obtain. The use of video-coded drowsy lane departures as ground truth for drowsiness had this difficulty and left large parts of the dataset uncategorized as either drowsy or awake. For that reason, an algorithm that employs a different ground truth definition for drowsy driving might complement the present drowsy-detection algorithm, and large coded datasets such as SHRP2 could prove useful in future research (Victor et al., 2015). Likewise, the ground truth for distracted driving was not complete, lacking a cognitive component. One may speculate that cognitive distraction and its effects on driving may vary from person to person even more than visual distraction (Lee et al., 2013).

Finally, the present research, with its new sample of self-paced distracted driving represents an important step in establishing a database of impaired driving data based on a common set of scenarios. The DrIIVE dataset may ultimately facilitate more efforts to understand and mitigate the effects of driver impairment and aid in the development of methods to assess driver state and enhance safety. Other driving-safety researchers may benefit from the availability of carefully controlled simulator data that complements broad naturalistic datasets that lack experimentally controlled conditions.

References

Blincoe, L., Miller, T., Wang, J. -S., Swedler, D., Coughlin, T., Lawrence, B., Guo, F., Klauer, S., & Dingus, T. (2023, February). *The economic and societal impact of motor vehicle crashes, 2019 (Revised)* (Report No. DOT HS 813 403). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813403

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. http://doi.org/10.1016/j.chemolab.2005.05.004

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. (2005, November 30-December 2). *Adjusted minimum time-to-collision (TTC): A robust approach to evaluating crash scenarios*. The Driving Simulation Conference North America, Orlando, FL.

Friel, P. N., Baer, J. S., & Logan, B. K. (1995). Variability of ethanol absorption and breath concentrations during a large-scale alcohol administration study. *Alcoholism: Clinical and Experimental Research*, *19*(4), 1055–1060. https://doi.org/10.1111/j.1530-0277.1995.tb00988.x

Grace, R., & Stewart, S. (2001, August 14–17). *The co-pilot: A low-cost drowsy driver and driver inattention monitor*. The International Driving Symposiums on Human Factors in Driver Assessment, Aspen, CO.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Report No. DOT HS 810 594). National Highway Traffic Safety Administration. https://rosap.https://rosap.ntl.bts.gov/view/dot/62931

Lee, J. D., Roberts, S. C., Hoffman, J. D., & Angell, L. S. (2012). Scrolling and driving how an MP3 player and its aftermarket controller affect driving performance and visual behavior. *Human Factors*, *54*(2), 250–263.

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 Admnistration.

Lee, J. D., Moeckli, J., Brown, T. L., Roberts, S. C., Schwarz, C., Yekhshatyan, L., Nadler, E., Liang, Y., Victor, T., & Marshall, D. (2013, May). *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

National Highway Traffic Safety Administration. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*, [Docket submission, Document ID NHTSA-2010-0053, in Regulations.gov]. www.nhtsa.gov/sites/nhtsa.gov/files/docketnhtsa-2010-0053.pdf

Romano, E., & Pollini, R. A. (2013). Patterns of drug use in fatal crashes. *Addiction*, *108*(8), 1428–1438. doi:10.1111/add.12180

Schwarz, C., Brown, T. L., Gaspar, J., Marshall, D., Lee, J., Kitazaki, S., & Kang, J. (2015, June 8–11). *Mitigating drowsiness: Linking detection to mitigation* [Paper presented]. 24th International Technical Conference on the Enhanced Safety of Vehicles, Gothenburg, Sweden.

Stewart, T. (2023, April). *Overview of motor vehicle traffic crashes in 2021* (Report No. DOT HS 813 435). National Highway Traffic Safety Administration.

Tefft, B. C. (2014). *Prevalence of motor vehicle crashes involving drowsy drivers, United States, 2009 – 2013*. AAA Foundation for Traffic Safety. https://aaafoundation.org/wp-content/uploads/2017/12/PrevalenceofMVCDrowsyDriversReport.pdf

Victor, T., Dozza, M., Bärgman, J., Boda, C. -N., Engström, J., Flannagan, C., Lee, J. D, & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk.* Transportation Research Board.

DOT HS 813 564a
September 2026

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

NHTSA

16175a-090226-v3