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

DOT HS 813 564c

September 2026

NHTSA's DrIIVE Program: Alcohol-Impaired, Drowsy, And Distracted Driving Detection Algorithms

This page is intentionally left blank.

DISCLAIMER

Suggested APA Format Citation:

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

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., 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

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 564c		

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

15. Supplementary Notes

Digital Object Identifier: https://doi.org/10.21949/2b5z-0398

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.

The main goal of the Driver Monitoring of Inattention and Impairment Using Vehicle Equipment program detection track was to develop and evaluate a system of algorithms to identify signatures of impairment caused by alcohol-impaired, drowsy, and distracted driving (the second report in this series explores potential mitigations to those impairments). This research collected data from drivers performing self-paced distracting tasks (secondary tasks). Algorithm development began with the definition of "ground truth" which showed when each type of impairment occurred in the data. The ground truth was the target of the detection algorithms. It continued with data processing which extracted the mathematical features of each type of impaired driving. These were used to train the three detection algorithms. Each of the three algorithms exhibited some effectiveness in detecting its target; but mixed results were obtained when used to distinguish among the three targets by applying them to cross-impairment datasets.

17. Key Words	18. Distribution Statement
alcohol-impaired driving, distraction, drowsy driving, impairment detection, driving under the influence, driving while intoxicated, driving simulator, DrIIVE	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	72	

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

100-Car Study	naturalistic data collection from 100 drivers
accuracy	ratio of true positives plus true negatives to all positives plus all negatives
alert	stimulus from a driver vehicle interface intended to warn the driver; see also *warning*
algorithm	process to be followed for solving a problem; a computerized process to detect driving impairment
ANOVA	analysis of variance
AttenD	eye-based distraction algorithm used as ground truth here
AUC	area under an ROC Curve
BAC	blood alcohol concentration, expressed in grams per deciliter (g/dL)
C++	general purpose programming language
caret[1]	Classification And REgression Training; a statistical toolbox in programming language R used to train and test machine learning algorithms, always lowercased
chunking	grouping individual task or subtasks into larger units
CI	confidence interval
confusion matrix	simple way to represent an algorithm's predictions as true positives, true negatives, false positives and false negatives
CWIM	Crash Warning Interface Metrics
decision tree	decision-making algorithm that consists of a number of sequential choices leading to a binary classification; a component of a random forest
deep learning	new approach to neural networks for machine learning
DrIIVE	Driver Monitoring of Inattention and Impairment Using Vehicle Equipment
driver-based	Using sensors that monitor the driver, such as eye and head trackers, EEG or other physiological sensors
driver distraction	diversion of attention away from activities critical for safe driving toward a competing activity
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.

dual-task cost	cost of one or both tasks of performing two tasks concurrently versus independently
DVI	driver vehicle interface
EEG	electroencephalogram, measures electrical activity in the brain
ensemble techniques	class of machine learning methods that have many constituent parts that contribute to the final classification
event level	data averaged over the duration of an entire event within the context of the drive
EWMA	exponentially weighted moving average filter
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)
forced-pace	task the participant must complete within a preselected timeframe
GLM	generalized linear models
Google Scholar	online scientific article search engine
ground truth	gold standard classification of the presence or absence of impairment against which algorithm performance may be judged
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
Latin square	experimental design used to counterbalance conditions to control order effects
MATLAB	computing and statistical software
model	representation of a physical or algorithmic system; can be graphical, computer code, or equations
N-back task	auditory working memory task where participants must recall a number heard N spaces ago in a stream of numbers
NADS	National Advanced Driving Simulator

naturalistic data	driver behavior observed during real, on-road driving through cameras and monitors in the vehicle
neural networks	type of machine learning algorithm that is loosely analogous to the function of the human brain
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
R	open source statistical software programming language
random forest	machine learning algorithm that uses a group of decision trees to vote on a binary classification problem, abbreviated RF
roadside observation	observing individual drivers as they pass a certain point to assess the frequency of secondary task interactions
ROC	receiver operator characteristic; a way to represent binary classifier performance for a family of algorithms that vary on a single parameter. Also called ROC curve. Implemented using pROC package in R.
SDLP	standard deviation of lane position
secondary task	task that is not necessary for driving that may result in distraction
self-paced	task in which the participant controls when to engage in the task
sensitivity	ratio of true positives to true positives plus false negatives
SHRP2	Second Strategic Highway Research Program naturalistic data collection
Simulink	graphical programming environment for modeling
SMS	short message service
specificity	ratio of true negatives to true negatives plus false positives
SVM	support vector machines
TA	truly awake
TD	truly drowsy
TLC	time to lane crossing
timeliness	ability of an algorithm to predict impairment soon after it appears
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)
vehicle-based	using sensors that monitor the vehicle, including inputs (e.g., steering, braking) and lane position

warning	stimulus from a DVI intended to warn the driver; see also *alert*
Web of Knowledge	servicemarked tradename of an online article database that changed its name to Web of Science in 1992
z-score	transform used to standardize data

Executive Summary

Three impairment-detection algorithms covering impairment from alcohol intoxication, drowsiness, and distraction successfully detected their targets but had mixed results when applied to cross-impairment datasets. The alcohol intoxication algorithm adapted well to the distracted and drowsy datasets. The distraction algorithm also worked moderately well when applied cross-impairment, although it worked better with head pose incorporated as a driver-based sensor signal. Finally, the drowsiness algorithm struggled in the other two datasets. One challenge of this exercise was creating ground truth signals for all impairment types in all three datasets. Some assumptions were made to fill in the ground truth gaps. Differentiating intoxication from drowsiness in data where both are present remains a challenge.

The simulator data collected in the DrIIVE program (Driver Monitoring of Inattention and Impairment Using Vehicle Equipment) formed a database of over 250 drivers in controlled scenarios at different times of day and under different types of impairment. As part of this research, driving simulator data from a new distraction study was collected using self-paced secondary tasks. Distraction is an important topic in driving safety. Driver distraction and inattention contribute to crash risk and are likely to have an increasing influence on driving safety in the future as vehicle technology and portable devices used in vehicles become increasingly common and complicated.

Three parts of an impairment mitigation systems are: (1) define the impairment states to be mitigated, (2) specify sensors needed to derive impairment-related features used by the detection algorithm, and (3) match algorithms and associated impairment states to alerts, warnings or other mitigation strategies. This phase of the DrIIVE project deals with detection and mitigation of impairment, but the focus of the detection track (the current report) was only on the impairment detection.

Algorithm development and evaluation needed detailed data from distracted drivers in controlled yet realistic situations. As part of algorithm development, a high-fidelity, full-motion driving simulator was used to collect self-paced distraction data. Data collection involved 24 drivers 21 to 34 years old driving through representative situations on three types of roadways (urban, freeway, and rural) at three levels of distraction (no distraction, moderate distraction, and high distraction). A verification analysis confirmed that assigned secondary tasks negatively affected driving performance, thus providing fertile ground for algorithm development.

An iterative process was used to create and improve the impairment-detection algorithms. This process began with defining ground truth for each type of impairment and continued with extracting features from the simulator data that might have been sensitive to the presence of impairment. Three detection algorithms (distraction, drowsiness, and alcohol intoxication) were trained using these features. This report documents the performance of each algorithm. The report also describes the ability of each algorithm to detect a specific impairment in the presence of other impairments (i.e., in a different dataset). While an overall mitigation system was not implemented under this scope of work while briefly discussing three algorithm outputs that might be arbitrated and prioritized for an overall mitigation system.

This page is intentionally left blank.

Introduction

This report covers work for NHTSA's DrIIVE program. The main goal of the detection track was to develop and evaluate a system of algorithms to identify signatures of alcohol, drowsiness, and distraction impairment. To meet the project objectives, the collection and validation of data from drivers performing self-paced secondary tasks (tasks apart from driving), as well as use of previously collected impairment datasets. The Summary of Datasets section outlines the datasets mined for algorithm development. The System of Algorithms and System Evaluation sections describe algorithm development and evaluation. Finally, the Discussion section summarizes the overall findings and conclusions from the project.

Background

One of the most prevalent sources of driver distraction is the use of mobile phones. Voice calls on mobile phones are associated with a four-fold increase in the odds of a crash (McEvoy et al., 2007). A naturalistic driving study found that 78 percent of crashes involved drivers looking away from the road (Klauer et al., 2006). The magnitude of the distraction problem may be underestimated due to the challenges of identifying distraction as the cause of a crash and driver reluctance to admit to having been distracted. The extent to which the estimate of crashes due to distracted driving in this study generalizes to the overall population of crashes remains a topic of debate; regardless there remains a concern even if the crash risk is smaller than that observed in this study. 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.

The DrIIVE program envisioned a single integrated system capable of detecting and mitigating several major categories of impaired driving: drowsiness, alcohol-intoxication, and distraction. For convenience, this report will refer to all three as "impaired driving." Work that preceded the current DrIIVE program consisted of the development and evaluation of algorithms to detect alcohol-impaired driving in the Impairment Monitoring to Promote Avoidance of Crashes Using Technology program, and the development and evaluation of algorithms to detect drowsy driving (DrIIVE Phase 1) (Brown et al., 2014). The current research program evaluated detection algorithms that used both vehicle-based signals and driver-based signals (head position and rotation).

Study Objectives

The initial task was to collect data for different levels of distracted driving. These data represent driving performance with the addition of "self-paced" visual and visual-manual secondary tasks. That is, tasks that afforded drivers the ability to determine when to begin and how much time to spend on each task. Self-paced tasks more realistically represent actual distracted driving than forced-pace secondary tasks. DrIIVE Phase 2 employed these data in conjunction with the alcohol-impaired driving data collected in the IMPACT program and the drowsy-driving data collected in DrIIVE Phase 1 for algorithm development and evaluation.

Two main hypotheses were tested to verify that the experimental manipulations were successful.

Hypothesis 1. Driver performance without distraction differs from driver performance when the driver is engaged in moderate and high levels of distraction.

Hypothesis 2. Driver performance differs between moderate and high levels of distraction.

The goal of the algorithm task was to develop a system of algorithms that could detect and differentiate different types of impairment. This goal was divided into the following objectives.

1. Standardize data organization and data processing scripts used in the distracted driving, drowsy driving, and alcohol-impaired driving datasets.
2. Create detection algorithms for distracted driving, drowsy driving, and alcohol-impaired driving.
3. Evaluate the performance of each algorithm in detecting its target impairment as well as in distinguishing its target impairment from the other two types of impaired driving.
4. Evaluate the potential benefit of including a driver-based signal (head position).
5. Evaluate the potential benefit of developing individualized algorithms that adjust for an individual driver's manner of driving.
6. Address general issues of real-time implementation, algorithm arbitration, and impairment mitigation.

Summary of Datasets

The data used for the development, validation and refinement of the impairment-detection algorithms was drawn from a set of studies using the same general driving environments and scenario events with a few modifications to support collection of relevant data.

Each drive was composed of three nighttime driving segments. The drives started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph with signal-controlled and uncontrolled intersections (see Figure 1 and Figure 2). An interstate segment followed that consisted of a four-lane divided expressway with a 70-mph posted speed limit. After a period in which drivers followed the vehicle ahead, they encountered infrequent lane changes associated with the need to pass several slower-moving trucks (see Figure 3). The drives concluded with a rural segment composed of a two-lane undivided road with curves (see Figure 4 and Figure 5). A portion of the rural segment was gravel. These three segments mimicked a drive home from an urban downtown location to a rural home via an interstate. Events in each of the three segments combined to provide a representative trip home in which drivers encountered situations that might be encountered in a real drive. For the drives used in studying drowsiness, a 10-minute-long segment of straight road was included at the end of the drive following the gravel roadway. For the drives used for studying distraction, secondary tasks were added throughout the drive.

Figure 1. Approach to curve in urban drive

Figure 2. Straight roadway segment in urban drive

Figure 3. Passing truck on interstate


Figure 4. Approach to rural curve


Figure 5. Rural vertical curve

Alcohol Data

This dataset included data from 108 participants equally split by gender across three age groups (21 to 34, 38 to 51, and 55 to 68). Each participant in the simulator drove at three levels of blood alcohol concentration (.00, .05, and .10 g/dL) approximately 1 week apart. All data was collected as the BAC declined to minimize extraneous variation associated with the effect of rising and falling BAC levels. Participants possessed valid U.S. driver licenses, were licensed drivers for 2 or more years, drove at least 10,000 miles per year, had no restrictions on the driver licenses except for vision, were not taking drugs that interacted with alcohol, did not use any special equipment to drive, and had been a moderate to heavy drinker, but not a chronic alcohol abuser. Lee et al. (2010) provides a full summary of this study.

Drowsiness Data

This dataset included data from 72 participants equally split by gender across three age groups (21 to 34, 38 to 51, and 55 to 68). Each participant drove at three levels of drowsiness based on time of day (9 a.m. to 1 p.m., 10 p.m. to 2 a.m., and 2 a.m. to 4 a.m.). Data was collected on two visits separated by at least 3 days. For each visit, participants were asked to log their sleep and food intake, get at least 6 hours of sleep the preceding night and avoid caffeine and alcohol leading up to the visit. For the daytime-alert visit, participants arrived at the simulator approximately 1 hour before their scheduled drive. For the nighttime-drowsy visit, participants were picked up at their homes after eating dinner and transported to the simulation facility around 7 p.m. The Stanford Sleepiness Scales were collected during the visit to document self-reported drowsiness. To be eligible, participants had to possess valid U.S. driver licenses, to have been licensed for 2 or more years, drove at least 10,000 miles per year, had no restrictions on the driver licenses except for vision, and did not use any special equipment to drive. Brown et al. (2014) provides a full summary of this study.

Distraction Data

Two sources of distraction data were used as part of this work. One set of data was collected using forced-paced tasks (Lee et al., 2013), and the other set using self-paced tasks collected as part of this project. The following sections summarize the data from those two studies.

Forced-Paced Distraction

This dataset included data from 32 participants equally split by gender between 25- and 50-year-olds. Each participant drove twice with and without engaging in secondary tasks. Secondary tasks started at fixed points during the drive and timed out after a specified time regardless of whether the task was completed. An incentive system (score) was used to encourage the participants to engage in the distracting tasks. The incentive was a function of overall task performance, including the time to initiate the distraction task, continuous attention to the task, and response accuracy. To be eligible, participants had to possess valid U.S. driver licenses, to have been licensed for at least 1 year, drove at least 3,000 miles per year, had no restrictions on driver's license except for vision, did not use any special equipment to drive, and had experience engaging in secondary tasks while driving. Lee et al. (2013) provides a full summary of this study.

Self-Paced Distraction

This dataset included data from 24 participants equally split by gender between 21- and 34-year-olds. Each participant drove three times without engaging in secondary tasks, engaging in moderately difficult secondary tasks, and engaging in difficult secondary tasks. Secondary tasks were located at specific events during the drive, but the driver could engage at any time after the task became available and take as long as needed to complete the tasks, thus the description 'self-paced.' No incentive system was used. To be eligible, participants must possess valid U.S. driver licenses, to have been licensed for at least 1 year, drove at least 10,000 miles per year, had no license restrictions other than for vision and good general health. Appendix A provides a full summary of this study.

System of Algorithms

Introduction

General impairment-detection algorithms and warning systems that use existing vehicle-based sensor suites would be attractive as they have the potential to be adopted quickly. The challenge to such an approach lies in the creation of algorithms able to detect different types of impairments and differentiating them from one another. We consider several aspects of algorithm design, including cross-impairment performance, individual differences, driver versus vehicle-based sensors, and issues of real-time implementation. While the mitigation track (Brown et al., in press) was a deep dive into the specific impairment of drowsiness, the detection track was a broad investigation of multi-impairment detection algorithms. This section describes a system of impairment-detection algorithms; and the System Evaluation section describes its evaluation on three impairment datasets.

Phase 1 Review

The phase 2 scope of work built on phase 1 by creating algorithms that could be implemented in a real-time environment, be it a driving simulator or a production vehicle. The phase 1 distraction simulator study included several instances each of a reaching task, a visual-manual task and a cognitive task presented to the driver in a force-paced manner that restricted their ability to delay engagement with the tasks. A suite of vision-based distraction algorithms characterized driver distraction and exemplified systems that use driver-based sensors (Lee et al., 2013). These results motivated a phase 2 simulator study that used self-paced secondary tasks that elicited more natural behaviors from the drivers.

The phase 1 drowsiness simulator study drives occurred during the day, early night, and late night. We explored several notions of drowsiness ground truth and developed a variety of vehicle-based algorithms. 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 on the drowsiness algorithm, implement a real-time version along with countermeasures, and collect new data with early night and late-night mitigation conditions.

Drowsiness ground truth within a drive was obtained by video-coding lane departures to show whether the driver was or was not drowsy (Wierwille & 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. The final evaluation dataset consisted of 578 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 truly drowsy and truly awake cases were unallocated to either state, and might reasonably have been either one.

The IMPACT simulator study tested drivers with .00, .05, and .10 g/dL BACs (Lee et al., 2010). Algorithms were developed to classify drivers with BAC levels greater than .08 (the legal limit for DUI) from those with lower levels; however, the algorithms were not designed for real-time implementation and used inconsistently sized time segments. The results motivated the phase 2 project to revisit the IMPACT dataset and create an algorithm that could be real-time and possibly differentiate between intoxication and other kinds of impairment.

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 6. The bottom three layers relate to the sensors that collect data from the vehicle, driver, and environment to be used as 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 determines the best way to mitigate the detected impairments, such as an alert issued in a particular modality. The DVI 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.

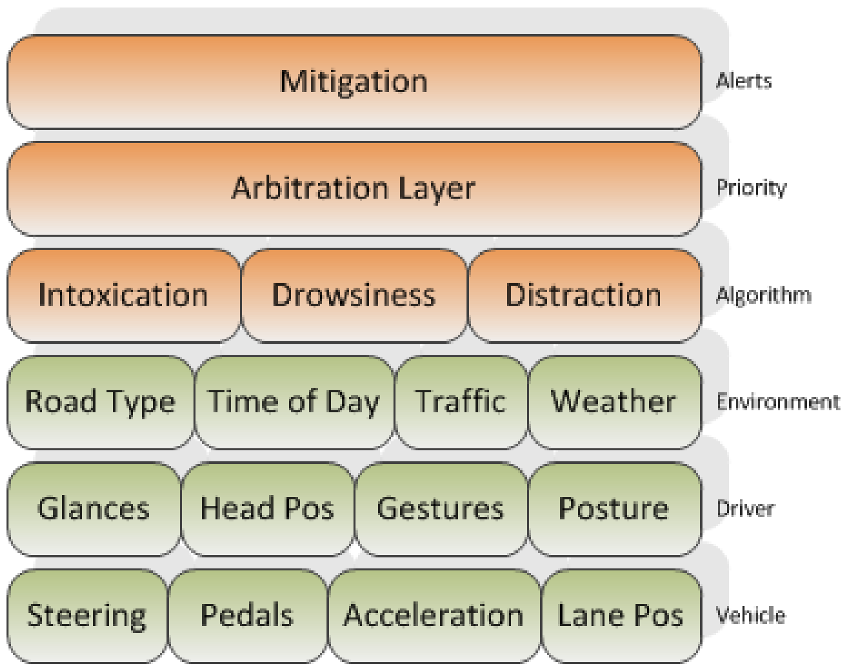

Figure 6. Layered framework for detecting, differentiating, and mitigating impairments

The first challenge in developing a coherent system of impairment algorithms is to adopt a standard methodology of processing data. Previous efforts were highly unique and used very different approaches to cleaning, processing, reducing, and extracting measures from data. In a commercialized system, this challenge would relate to designing a system of sensors, electronics and in-vehicle networks that process the various signals and present them to the algorithms. In the context of driving simulator data, it had to do with processing the data from three separate studies using a common set of scripts with common sampling methods and feature extraction. The following steps were excluded on the IMPACT alcohol dataset, the phase 2 distraction dataset and the phase 1 drowsiness dataset.

1. Convert NADS data acquisition files into MATLAB format.
2. Calculate common measures from simulator raw data.
3. Calculate measures unique to impairment (e.g., task engagement).
4. Save selected raw data and measures into comma-separated text files.

5. Compute aggregated measures on each sequential second of data.
6. Aggregate measures at resolution appropriate to each impairment.

Some measures and steps are still unique to the specific impairment dataset. The alcohol dataset includes BACs; and the drowsiness dataset includes time-of-day condition. Moreover, the drowsiness algorithm operates on inputs that are further aggregated from 1-second windows into 60-second windows. The key though is that these unique steps are added to a common processing framework.

The researchers adopted a data-centric, rather than model-centric, approach to developing impairment algorithms. A model-centric approach might develop a model of the physical and mental processes of the driver in their normal and impaired state and use data to identify a set of parameters for the model. In contrast, the data-centric approach examines the data from many drivers and learns from patterns and differences that can be observed between normal and impaired drivers. Machine learning techniques that are often applied to "big data" are data centric.

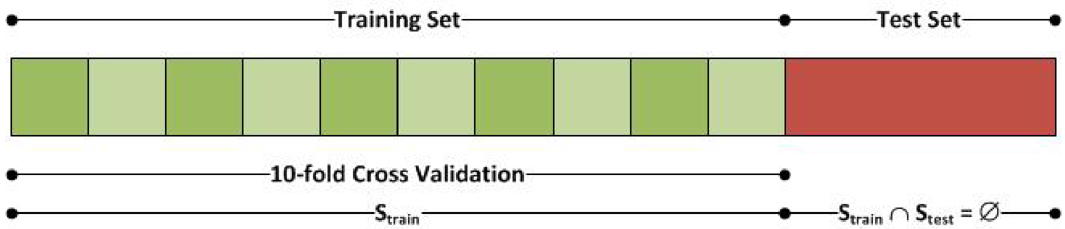

Figure 7. Algorithm training and testing approach

The standard machine learning procedure is to use a portion of data to train an algorithm and reserve another portion to test the performance of the algorithm fairly. A good training/testing regime prevents common pitfalls of the data-centric approach like overfitting to the training data. The method followed is represented graphically in Figure 7. About 75 percent of the data was used to train the model using 10-fold cross validation. Then the remaining 25 percent was used to run the model and test its performance on data it has not seen before. Given that there are many ways to split the training and test sets, a conservative approach was used in creating distinct sets of participants. Each participant was assigned to either the training set or the test set. The set allocation was random; however, an allocation could be rejected if it did not preserve the ratio between impaired cases and normal cases. If rejected, the allocation was randomized again until preservation of the impairment ratio was sufficiently achieved.

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 and, different subsets of features, thresholds, and different branching conditions. Each decision tree predicts the driver state and these predictions decide according to a majority vote. The output of the RF is the state predicted by the greatest number of decision trees. Figure 8 shows how predictions from many decision trees are combined through voting to show driver state.

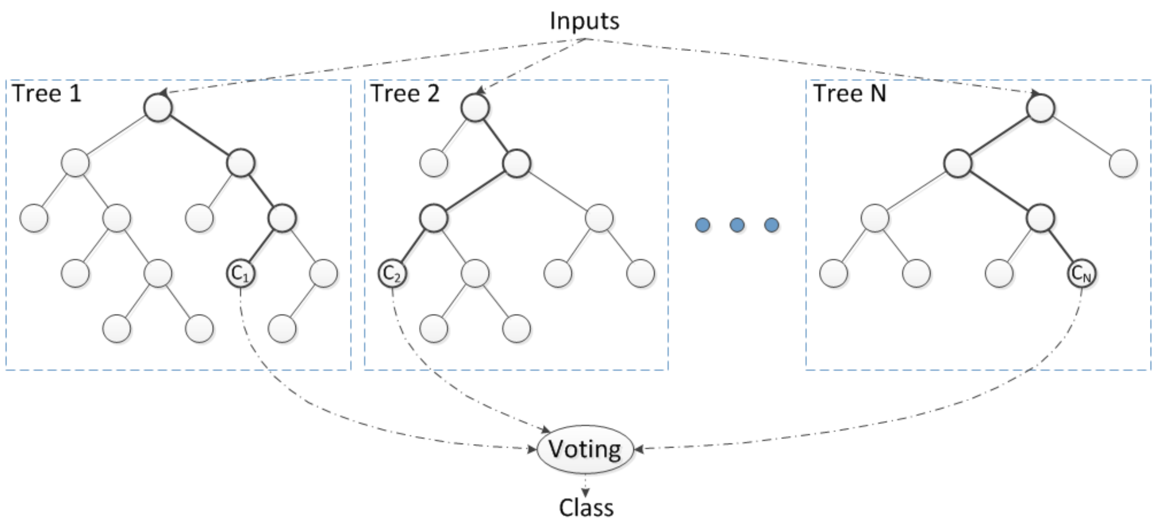

Figure 8. Random forest machine learning model

The second component of the algorithms was an HMM that combined the RF observations over a brief time to augment the driver state classification. An HMM assigns probabilities regarding whether the system remains in a given state or transitions to a new state at each moment in time. It uses the history of the state to predict the driver's current state. This history is particularly important in predicting driver state assuming that a driver who was recently drowsy, for example, is likely to remain drowsy. Distraction depends on the shortest history and alcohol depends on the longest, corresponding with the natural timescales of those impairments. Figure 9 shows how the HMM combines a series of inputs, (the classifications of the RF) to better predict driver state in a way that reflects how the current state depends on past states.

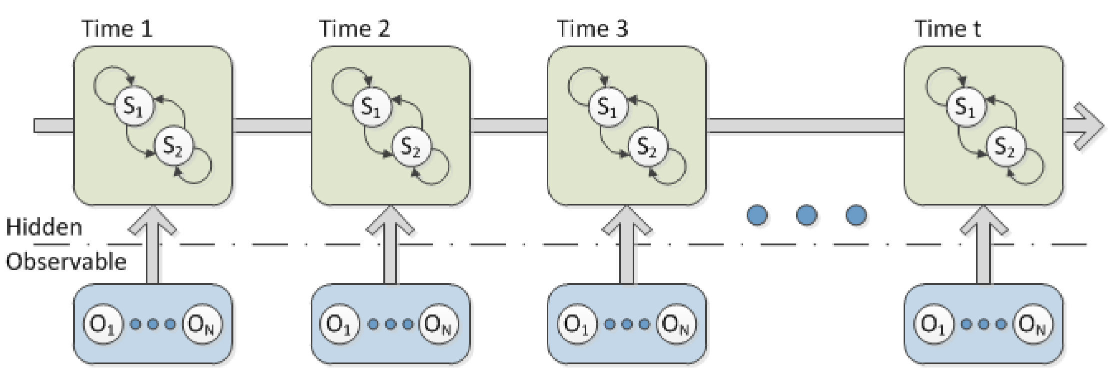

Figure 9. Hidden Markov model

Distraction Algorithm

A distraction-detection algorithm was a specific deliverable under the system-of-algorithms task. Vehicle-based sensor signals were used to train a distraction-detection algorithm from the study data. Then an algorithm that used driver-based signals was compared to the vehicle-based algorithm to compare performance. This section describes the approach to algorithm design. Results are presented in the System Evaluation section.

The definition of ground truth for distracted driving is a challenge. Does task engagement equal distraction? If not, when does distraction begin and what constitutes the threshold that separates distraction from alertness? NHTSA's distraction guidelines penalize glances longer than 2 seconds, as well as total glance time greater than 12 seconds to complete a particular secondary task (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).

A slightly modified AttenD (Kircher & Ahlstrom, 2009) algorithm was used to define ground truth of distracted driving. AttenD uses gaze location and requires an eye tracker to collect data. It defines a 90° horizontal field of view in which gaze is interpreted to be on the road. The allowed vertical field of view is 22.5° downward from center. Glances outside of the front field of view are tested to see whether they are aimed at any of the mirrors, which are also allowed. Failing to meet the mirror exception, AttenD accumulates eyes-off-road time and shows distraction when a threshold is crossed. When the driver returns their gaze to the front, the output begins to drop, after a short delay, at the same rate until it reaches zero. Distraction shows when the output exceeds some threshold. A threshold of 2.0 was used. Finally, the output is limited at a maximum value, set to 2.05 in this research.

AttenD detects single long glances. Due to its accumulation mechanism, it also penalizes densely spaced glances separated by short glances to the front. The published algorithm initializes the output at two seconds and then drops it towards zero when distracted, while the algorithm used in this study starts at zero and raises the output towards two. Figure 10 shows an example of the AttenD output, whose units are in seconds. The shaded regions show 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 10. It is also possible to be distracted while not engaged in the tasks, as in the short periods on the left side of Figure 10.

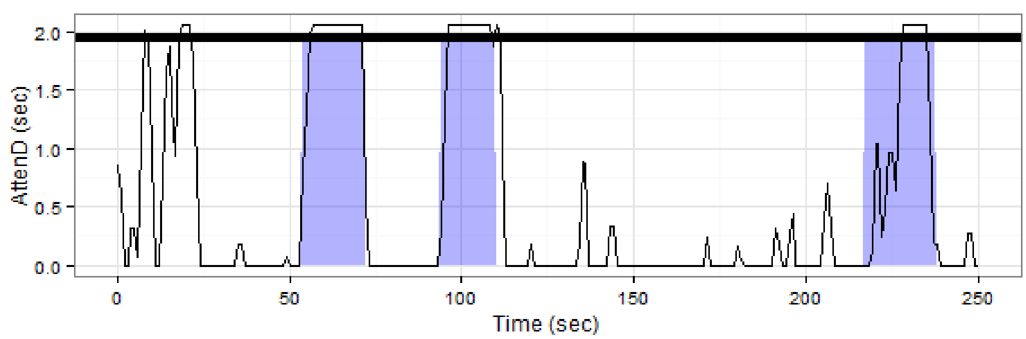

Figure 10. AttenD distraction metric defines the ground truth of distraction as values over 2.0, shown by the bold horizontal line

Driver-Based Signals

Driver-based sensor signals may provide the most direct measure of impairment and could be expected to benefit an impairment-detection algorithm. Visual distraction would especially benefit from sensors that monitor the drivers gaze direction; however, there are obstacles to the approach. Generally, robust eye-tracking is difficult to achieve in all realistic lighting conditions, and drivers may not accept this type of monitoring technology. Specifically, an eye-tracking algorithm was used to establish ground truth for distraction so an eye-based detection algorithm would have an unfair advantage.

A compromise is to consider head pose as a driver-based signal since it is easier to detect than gaze location. Examples of head pose signals are shown in Figure 11. Notice the clear pattern of horizontal head rotation that is associated with task engagement. We note that the study tasks were in one fixed location; but in reality, there are numerous ways for a driver to look away from the road.

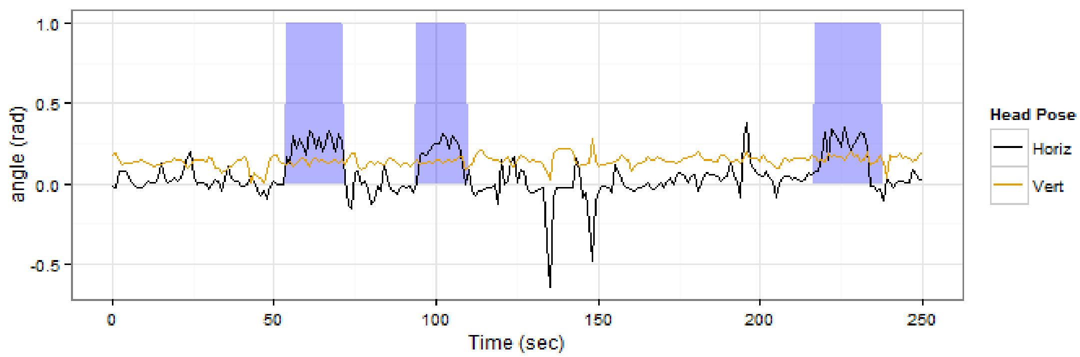

Figure 11. Mean vertical and horizontal head pose signals with task engagement shown by shaded regions

Individual Differences

A typical analysis averages measures across drivers to examine characteristics of a larger population, and this research has followed this approach too. Furthermore, each driver exhibits individual differences from other drivers and has a "driving signature" that might be used to customize algorithm performance.

We studied the potential effects of individual differences on algorithms through a counterintuitive approach of smoothing away differences among drivers during algorithm training and testing. Z-scores were used to standardize sensor inputs from each driver with the following logic: if applying the z-score transformation to the algorithm inputs during training and testing improves its performance, then there must have been individual differences present that contributed to a performance decrement. Improved performance under z-score transformation therefore serves as an estimate of the presence and size of an effect, but not necessarily the nature of the individual differences.

The method involves calculating the z-score for each signal that feeds into the algorithm under normal, unimpaired driving conditions. If a signal, s, has a mean, μ, and a standard deviation, σ, then its z-score is given by the equation:

$$z = \frac{s - \mu}{\sigma}$$

A normally distributed signal would have, after transformation, a standard normal distribution with mean of zero and standard deviation of one; however, the method is still applicable to signals that are not normally distributed. A baseline set of means and standard deviations was computed for each signal for each driver during unimpaired driving. These statistics were applied to their signals for each driver in normal and impaired driving conditions; and the transformed signals were used to train and test alternate algorithms. Results are presented in the System Evaluation section.

Real-Time Implementation

The ultimate deployment scenario for a system of impaired driving algorithms involves integration into a vehicle's hardware and software, and an implementation that can process sensor inputs and make predictions in real-time.

In the present program of research, the framework of RF 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 drowsiness data collection under the mitigation track. We discuss some general issues regarding real-time implementation of this framework.

Running the entire RF at each time step was too computationally intensive to consider, so the prediction was only updated every 6 seconds. To meet this need, a Simulink model called a C++ implementation of the algorithm. In fact, the update rate could have been reduced to around once every half a second on a desktop workstation.

The HMM implementation was quite simple. 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.

Fortunately, just the value of the most recent state is of interest in this application. Thus, a more suitable approach is to apply smoothing or filtering to the HMM to estimate the posterior probability of it being in the impaired state at time T, both of which are reasonable for real-time implementation. As a matter of practical consideration, it is not feasible to consider time intervals that extend back to infinity (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 RF. The time horizon was fixed at 5 minutes for the HMM algorithm.

Z-score transformations could also be included in real-time implementation. Such an approach would begin to collect baseline means and standard deviations for each signal and update them continuously. After minimum time has passed, the baseline would be considered useable and be applied to the algorithm signals. The baseline updates would continue except during such times that the algorithm predicts impaired driving. An exponentially weighted moving average could be used to gradually forget about old z-scores.

Arbitration and Mitigation Layers

If several impairment-detection algorithms run in parallel, then it may happen that several algorithms report impairment at the same time, and arbitration would be used to decide what is passed on to the mitigation layer.

Then the mitigation layer is responsible for acting on the arbitrated algorithm output to issue a warning of some type to the driver. Some issues relating to drowsiness mitigation were studied in the mitigation track, and we discuss arbitration and mitigation only superficially here.

There is a natural temporal ordering of driving impairment from distraction to drowsiness to alcohol intoxication based on the inherent timescale of the impairments. In the presence of intoxication, drowsiness might reasonably wax and wane; and during a drowsy episode, distraction could come and go. For this reason, distraction mitigation (i.e., warning) may benefit from being prioritized higher than drowsiness mitigation. This implies that if both distraction and drowsiness algorithms show impairment, then distraction would be passed to the mitigation layer because it is considered in this model to be the more time-sensitive of the two. Note that this time sensitivity does not necessarily correlate with the seriousness of the impairment, nor the ability to recover from that impairment state.

A second consideration is the immediate interest in and expected use of, issuing a warning for each type of impairment. Here again, a distraction alert is likely associated with a long glance away from the road and is hypothesized to benefit from an immediate warning to prevent a crash or near crash situation. In contrast, drowsiness develops over a longer period and there is presumably more time in which an alert could be issued in advance of a drowsy lane departure.

System Evaluation

A system of impairment algorithms was trained and tested. The System of Algorithms section describes the algorithm framework and presents results summarizing its performance. Discussion regarding the algorithm results is given in the next section. All the measures that were used as algorithm inputs, both in raw and aggregated form, are listed in Appendix E (see Table E-19 and Table E-20).

Distraction Detection Algorithm

The distraction algorithm was trained and tested using the phase 2 distraction dataset, which had two different types of visual distractions. Both tasks involved looking at a screen near the center console. The tasks were self-paced in that the drivers had a choice of when to begin and how long to spend on each task. Overwhelmingly, drivers did not choose to put off task engagement; however, there was large variation in how long different drivers took to complete the tasks.

The AttenD measure was calculated as a ground truth signal. About 75 percent of participants were randomly selected into the training set, with the remaining 25 percent allocated to the test set. No further aggregation of the data was done beyond the initial aggregation into 1-second segments. Thus, the distraction algorithm operated on 1-second windows of data and could provide prediction updates every second.

A RF was trained on the 1-second data using several vehicle-based signals and some environmental-based signals as inputs. Then the output of the RF was used as an input to an HMM. The performance summary of just the RF stage as well as the total algorithm (HMM) is given in Table 1.

Each set of parentheses in Table 1, and throughout the section, presents the estimated value of the measure (center value) as well as the 95 percent confidence interval (left and right values). The CI for the area under the ROC curve (AUC) was estimated using the Delong method in the pROC package (Robin et al., 2011) using the R statistics software (R Development Core Team, 2011). The CIs for all other measures were estimated using the Wilson score interval (Wallis, 2013).

Table 1. Distraction detection algorithm statistics

Measure	RF	HMM
AUC	(0.78,0.80,0.81)	(0.81,0.82,0.84)
Sensitivity	(0.37,0.38,0.39)	(0.55,0.56,0.57)
Specificity	(0.94,0.94,0.94)	(0.90,0.90,0.91)
Accuracy	(0.90,0.90,0.91)	(0.88,0.88,0.89)
Kappa	(0.28,0.29,0.29)	(0.31,0.32,0.33)
Positive Predictive Value	(0.30,0.31,0.31)	(0.28,0.29,0.29)

An alternate version of the distraction detection algorithm was trained and tested with driver-based signals for head pose added to the input data. The performance statistics are summarized in Table 2.

Table 2. *Distraction detection algorithm statistics with head pose*

Measure	RF	HMM
AUC	(0.90,0.91,0.92)	(0.92,0.92,0.93)
Sensitivity	(0.66,0.67,0.68)	(0.69,0.70,0.70)
Specificity	(0.93,0.94,0.94)	(0.95,0.95,0.95)
Accuracy	(0.92,0.92,0.92)	(0.93,0.93,0.94)
Kappa	(0.47,0.48,0.48)	(0.53,0.54,0.54)
Positive Predictive Value	(0.41,0.42,0.43)	(0.48,0.48,0.49)

The raw outputs of the RF stage and HMM may be compared and contrasted by examining the signal traces shown in Figure 12 and Figure 13. The vertically distributed plots in the figure show individual drivers in the test set. Periods of task engagement are displayed as shaded regions in the plots. The ground truth from AttenD is shown as a black line and has units of seconds. The colored dots show the output of the RF and should be interpreted as a ratio of distracted votes to total votes having values between zero and one. One color shows a normal classification, while the other shows distraction. The threshold between the two is at 0.5, which corresponds to the point at which half the trees vote for distraction. Finally, the colored line is the posterior probability of the HMM being in the distracted state obtained by smoothing and can vary between zero and one. Observe how much less noisy the RF output is using driver-based head signals.

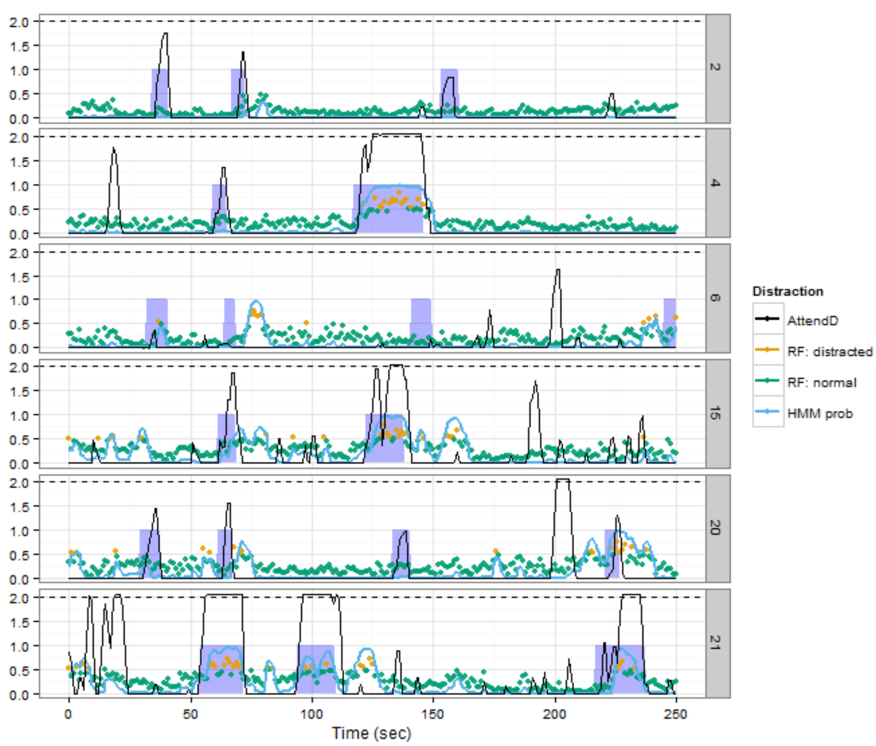

Figure 12. Distraction detection algorithm performance

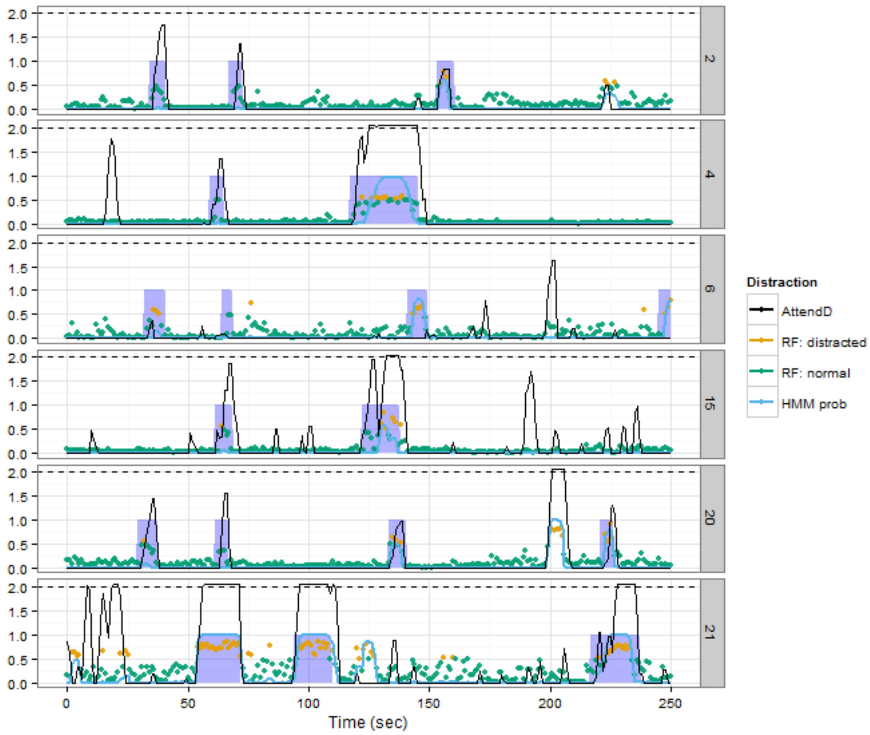

Figure 13. Distraction detection algorithm performance with head pose

The RF training process records the relative importance of each of its input variables after training. Examination of variable importance can yield clues about which signals are most important to accurately estimate impairment and which could be removed from the algorithm with little to no impact on its performance. Throughout this section, we have adopted the approach of including many variables and have not taken the logical next step of pruning the less important ones.

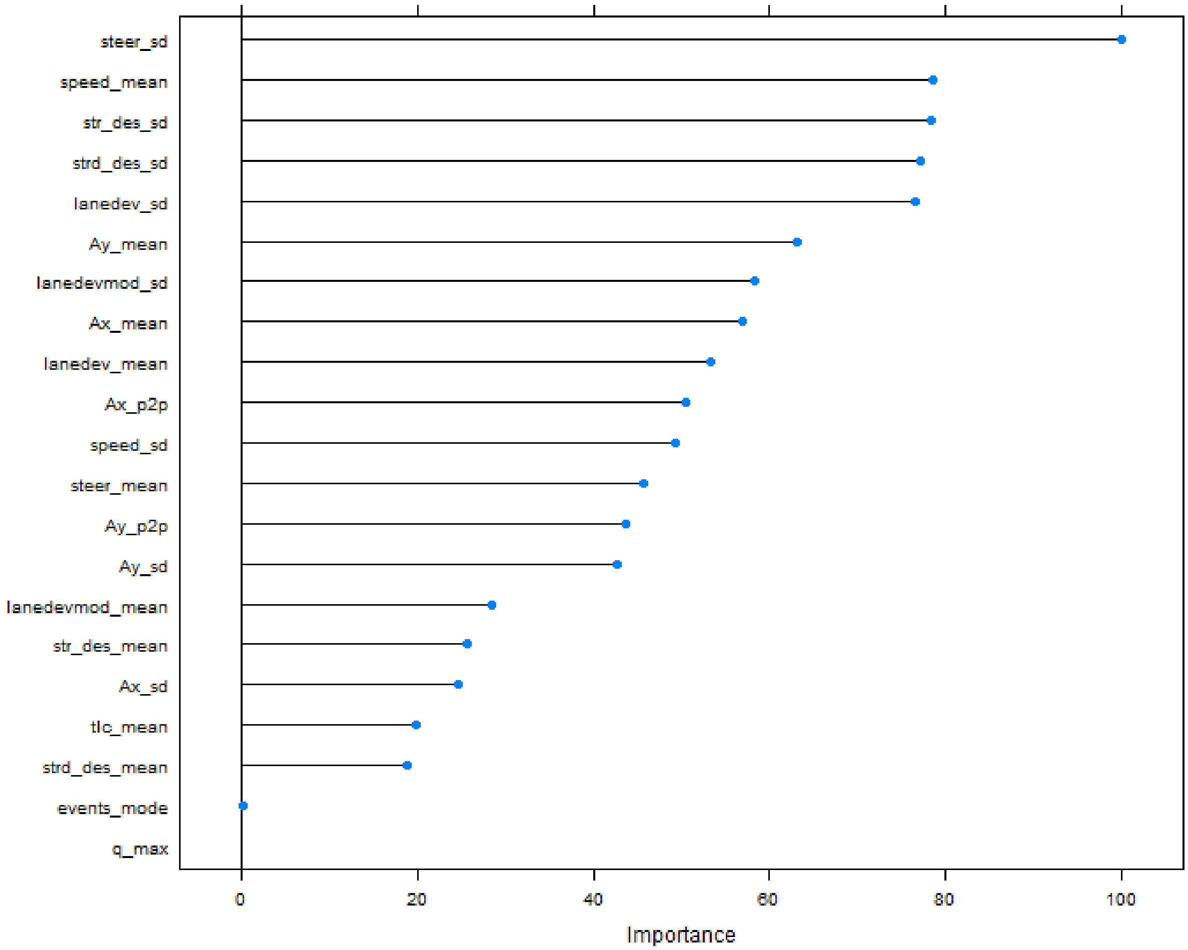

Figure 14. Distraction detection algorithm variable importance

A variable importance plot is a graphical representation of the relative variable importance in a RF model in which each variable's importance is shown on a horizontal stem plot. It can be automatically generated by the caret package in R (Kuhn, 2008). The variable importance plot for the vehicle-based RF algorithm stage is shown in Figure 14 while the driver and vehicle-based RF stage is shown in Figure 15. Observe the high importance attributed to all the head pose variables in the driver-based algorithm.

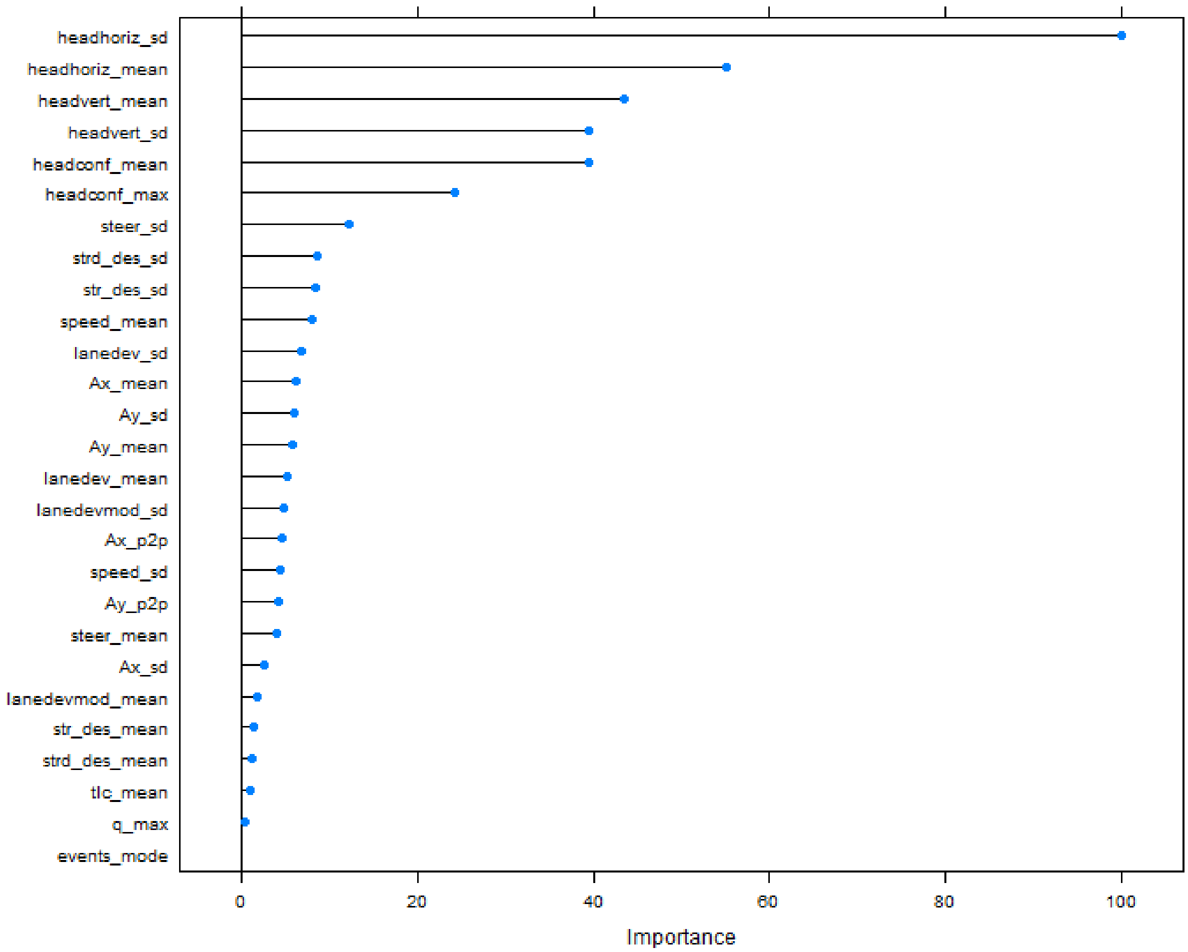

Figure 15. Distraction detection algorithm variable importance with head pose

A System of Impairment Algorithms

We applied a common methodology for processing impairment data not only to the distraction dataset, but also to the drowsiness and intoxication datasets for the purpose of developing a system of algorithms for detecting and differentiating the three types of impairment. Consequently, a new drowsiness algorithm was trained, distinct from the one developed in the mitigation track. Drowsiness and intoxication algorithm results are presented in the next two sections, followed by an evaluation of their cross-impairment performance.

Drowsiness Detection Algorithm

The structure of the new algorithm is like the one in track B except that this one used one RF as input to an HMM, while the track B model used two of them. The mitigation track used a more carefully curated set of ground truth from a subset of drivers deems truly drowsy and truly awake, while this model was allowed to draw drowsy lane departures from any drive in which they occurred.

The performance of the vehicle-based drowsiness model is shown in Table 3. The base algorithm, shown in the first column, has poor performance in sensitivity to drowsy lane departures. However, when *z*-score normalization is used to smooth away individual differences between drivers, the performance improves significantly. The sensitivity jumps to 0.92 (95% CI: [.89, .96]); but the specificity drops to 0.68 (95% CI: [.63, .75]). The net gain in performance is reflected in the increased AUC of 0.83 with 95 percent CI of [.77, .89].

Table 3. Drowsiness detection algorithm with z-score normalization

Measure	HMM	HMM Z
AUC	(0.56,0.64,0.71)	(0.77,0.83,0.89)
Sensitivity	(0.13,0.17,0.23)	(0.89,0.92,0.96)
Specificity	(0.91,0.93,0.97)	(0.63,0.68,0.75)
Accuracy	(0.36,0.41,0.48)	(0.80,0.84,0.90)
Kappa	(0.05,0.07,0.11)	(0.57,0.62,0.69)
Positive Predictive Value	(0.81,0.84,0.90)	(0.82,0.86,0.91)

Table 4. Drowsiness detection algorithm with z-score normalization and head pose

Measure	HMM	HMM Z
AUC	(0.60,0.67,0.74)	(0.85,0.89,0.93)
Sensitivity	(0.44,0.49,0.57)	(0.86,0.89,0.94)
Specificity	(0.66,0.71,0.78)	(0.63,0.68,0.75)
Accuracy	(0.51,0.56,0.63)	(0.78,0.82,0.88)
Kappa	(0.13,0.17,0.22)	(0.53,0.58,0.66)
Positive Predictive Value	(0.74,0.79,0.85)	(0.82,0.86,0.91)

Performance is further improved using driver-based signals for head pose information as shown by a slightly larger AUC of 0.89 (95% CI: [.85, .93]) in Table 4. However, the picture is less clear when examining the other performance statistics. There are tradeoffs between sensitivity and specificity that are at play. Note that AUC is a measure of an ROC curve and represents a family of algorithms varying over some parameter, while the other statistics are obtained by selecting an operating point on the ROC curve to produce one algorithm, and one confusion matrix, from the family. The default operating points for all RF algorithms is 0.5 meaning that a simple majority of trees in the forest must vote for impairment. The operating points for all HMMs is also 0.5, meaning that the posterior probability of impairment from the HMM must be greater than one half. Thus, it could happen that AUC increases while other statistics look worse because a better operating point might be available from the ROC curve.

The variable importance plots for the RF segments of the algorithms with and without z-score normalization are shown in Figure 16 and Figure 17 respectively. What can account for the increase in performance evident in Table 3? Consider the variables that move towards the top of the importance plot with z-score normalization. In addition to small steering reversals, large and very large steering reversals are also present. The standard deviation of brake force increased in importance, as did the road demand metric, *q_max*. Road demand metric is included as an environmental input and serves as a proxy for a group of road measures that might be collected using a forward-facing camera and GPS/Geographic Information System data.

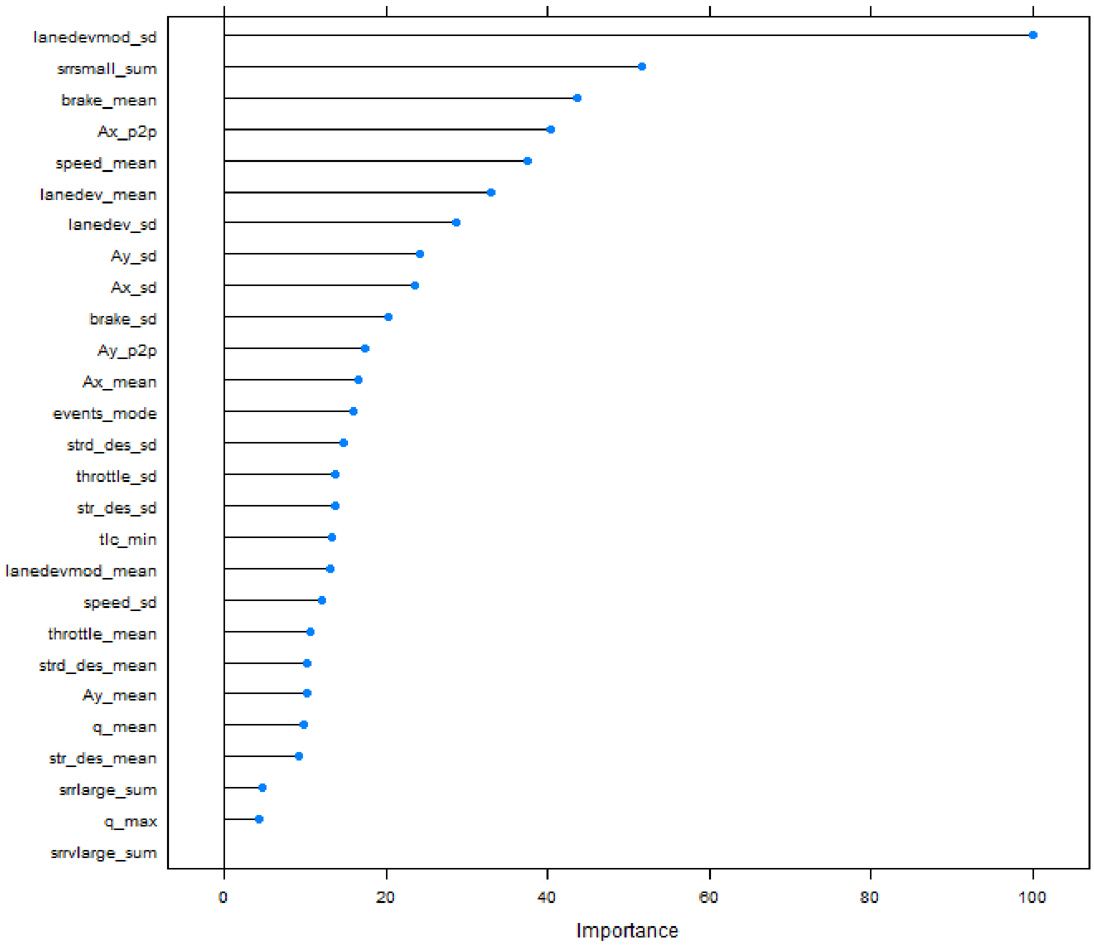

Figure 16. Variable importance in drowsiness RF

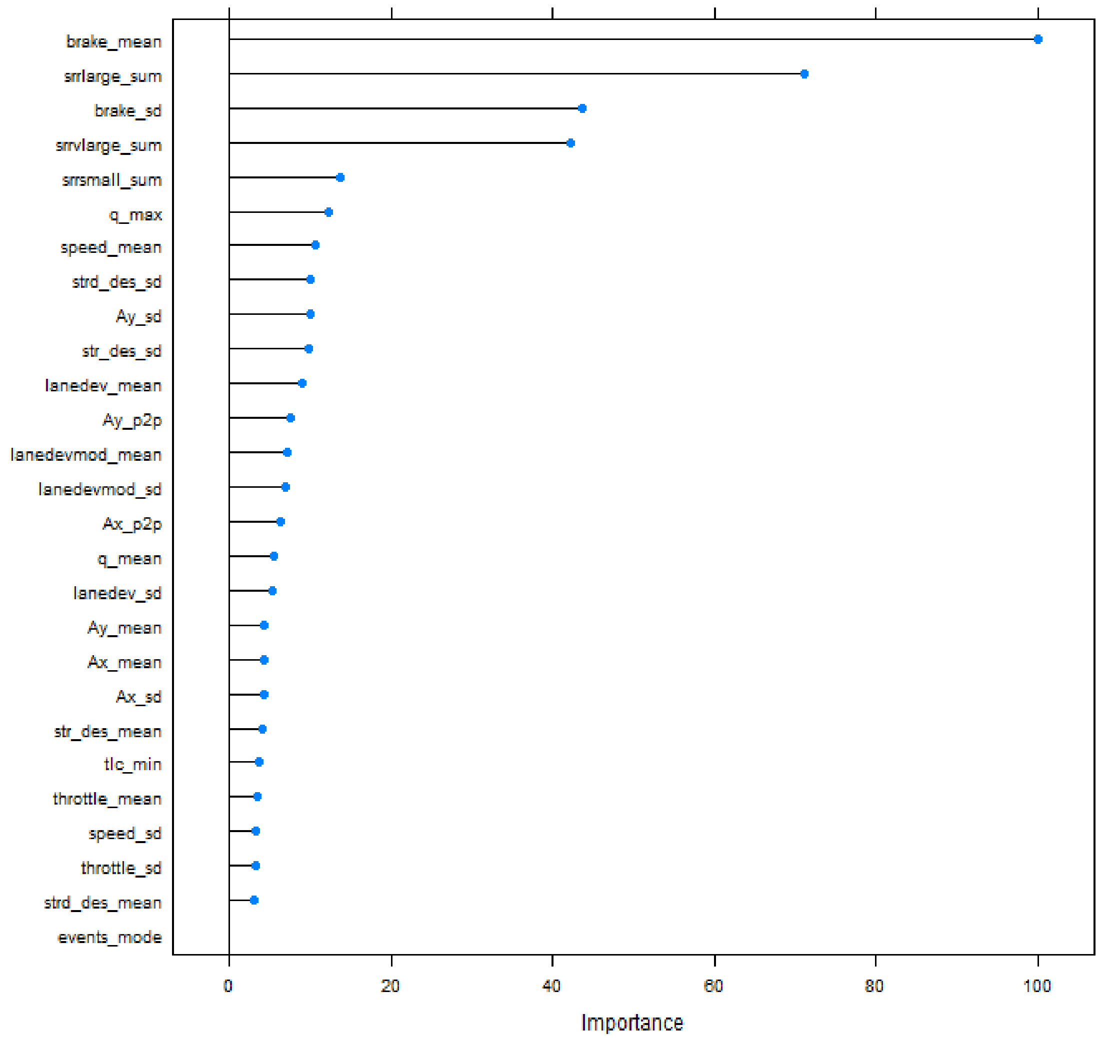

Figure 17. Variable importance in drowsiness RF with z-score normalization

Intoxication Detection Algorithm

An intoxication detection algorithm was trained using the same RF/HMM framework and a common set of scripts and functions used with distraction and drowsiness. As with both of the other impairments, a single RF generates a classification probability and is fed in as input to an HMM, from which a posterior probability of impairment is estimated with a smoothing procedure. The algorithm operates on 5-minute chunks of data, which are aggregated from the 1-second segments. A 50-percent overlap is allowed, resulting in an update period of 2.5 minutes for new predictions.

The performance of an initial algorithm, as well as that of a version with z-score normalization, is summarized in Table 5. Observe that without z-score normalization, the algorithm performs no better than pure chance in classifying intoxication, but with it, the performance is reasonable. Head pose signals were added to the algorithm, but they did not result in improved performance in the z-score normalized algorithm, as seen in Table 6, and only slightly improved the non-normalized algorithm.

Table 5. Intoxication detection algorithm with z-score normalization

Measure	HMM	HMM Z
AUC	(0.45,0.50,0.56)	(0.75,0.78,0.82)
Sensitivity	(0.02,0.03,0.05)	(0.87,0.89,0.92)
Specificity	(0.98,0.99,1.0)	(0.55,0.59,0.63)
Accuracy	(0.64,0.68,0.72)	(0.65,0.68,0.73)
Kappa	(0.02,0.02,0.04)	(0.36,0.40,0.44)
Positive Predictive Value	(0.51,0.55,0.59)	(0.47,0.51,0.55)

Table 6. Intoxication detection algorithm with z-score normalization and head pose

Measure	HMM	HMM Z
AUC	(0.53,0.58,0.63)	(0.74,0.77,0.81)
Sensitivity	(0.01,0.02,0.03)	(0.93,0.95,0.97)
Specificity	(1.0,1.0,1.0)	(0.53,0.56,0.61)
Accuracy	(0.65,0.68,0.72)	(0.65,0.69,0.73)
Kappa	(0.01,0.02,0.04)	(0.38,0.42,0.46)
Positive Predictive Value	(1.0,1.0,1.0)	(0.47,0.51,0.56)

The few most important variables in the RF were small steering reversals, standard deviation of lane position, average brake, standard deviation of throttle, and average steering (de-trended). The variable importance plot using z-score normalization is shown in Figure 18. It is interesting that many variables appear to have played an important role in classifying intoxication.

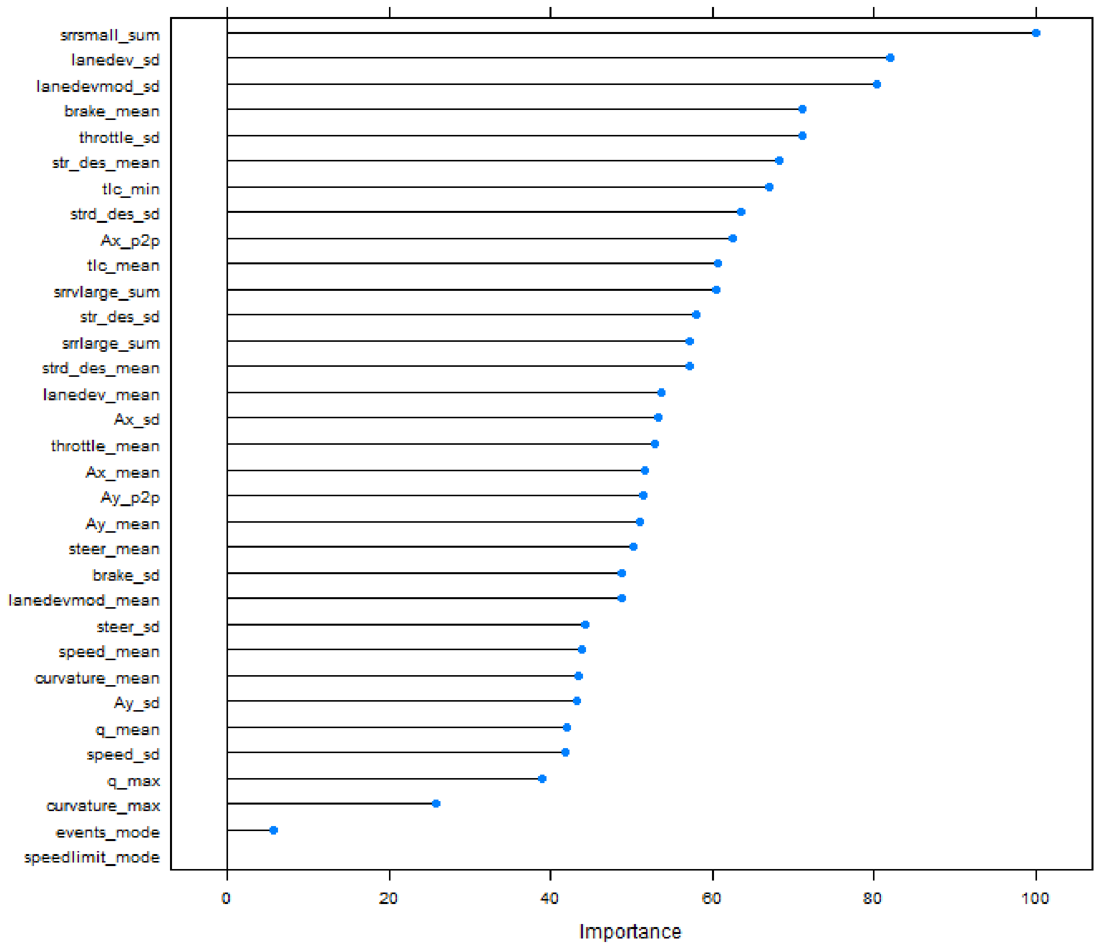

Figure 18. Variable importance in intoxication RF with z-score normalization

Cross-Impairment Performance

Each of the three driver impairments (drowsy, intoxicated, and distracted) supplied datasets that trained algorithms using a combination of RF and HMMs. The three impairment algorithms were applied to 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 required that ground truth be available across all the datasets. For example, any instances of drowsiness in the distraction dataset had to be coded, otherwise the drowsiness ground truth would not be present and predictions from the drowsiness-detection algorithm could not be properly evaluated. Table 7 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. We were unable to evaluate the alcohol and distraction algorithms' sensitivity to drowsiness in the other datasets without this show of ground truth, so it is 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.

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, we cannot assume 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 shows 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 7. Ground truth for each algorithm/dataset combination

	Distraction data	**Drowsiness data**	**Intoxication data**
Distraction alg.	AttenD	AttenD	AttenD
Drowsiness alg.	Assume none	Video coded lane departures	Assume none
Intoxication alg.	Screened out	Screened out	BAC level

Applying the three algorithms to each of the three datasets tests the ability of the algorithms to differentiate the impairments—cross impairment classification performance. For example, the drowsiness-detection algorithm should not show any instances of drowsiness in the distraction data if it is not present. Applying the algorithms to cross-impairment datasets produced mixed results.

The distraction algorithm was applied to the drowsiness and intoxication datasets. All datasets had the AttenD measure for distraction calculated for use as ground truth. Table 8 shows the performance of the distraction-detection algorithm for detecting distraction in the drowsiness dataset. Observe that the algorithm is largely ineffective using vehicle-based signals but still works quite well with head pose signals included.

Table 8. Distraction detection algorithm applied to drowsy dataset

Measure	**HMM**	**HMM Head**
AUC	(0.56,0.58,0.59)	(0.80,0.81,0.82)
Sensitivity	(0.16,0.16,0.16)	(0.40,0.40,0.40)
Specificity	(0.94,0.94,0.94)	(0.98,0.98,0.98)
Accuracy	(0.90,0.91,0.91)	(0.95,0.95,0.95)
Kappa	(0.08,0.09,0.09)	(0.41,0.41,0.41)
Positive Predictive Value	(0.11,0.11,0.12)	(0.47,0.47,0.48)

The performance of the distraction algorithm applied to the alcohol dataset is summarized in Table 9. Here, the vehicle-based algorithm preserves most of its performance but is still outperformed by the version with head pose included.

Table 9. Distraction detection algorithm applied to alcohol dataset

Measure	HMM	HMM Head
AUC	(0.70,0.73,0.76)	(0.87,0.89,0.91)
Sensitivity	(0.45,0.45,0.46)	(0.36,0.36,0.37)
Specificity	(0.86,0.86,0.86)	(0.98,0.98,0.98)
Accuracy	(0.86,0.86,0.86)	(0.98,0.98,0.98)
Kappa	(0.01,0.01,0.15)	(0.10,0.11,0.11)
Positive Predictive Value	(0.01,0.01,0.01)	(0.06,0.07,0.07)

It may be unfair to report AUC in a cross-impairment evaluation since a pre-trained algorithm is being applied and the freedom to pick a new operating point on a different dataset is not realistic. Note that the AUC measure, in this context, shows how well an optimal algorithm could perform given the opportunity to tune its parameters.

The drowsiness detection algorithm was then applied to the detection and intoxication datasets. Video-coded drowsy lane departures for ground truth in the other two datasets were unable to be generated, so an assumption of no drowsiness was made for evaluation purposes. Given this assumption, several of the statistics take on values of zero, including sensitivity. The key measure to consider is specificity, which is the ratio of true negatives to total negatives. In the absence of positive cases, the value of specificity should be one. The performance of the drowsiness algorithm cross-impairment was poor but was better for the RF stage than it was for the HMM. The sensitivity of the vehicle-based RF stage applied to the distraction data is 0.37 with a 95 percent CI of [.31, .45], while that of the driver-based RF stage using head pose is 0.38 with 95 percent CI of [.32, .46]. Similar results were obtained on the intoxication dataset. The sensitivity of the vehicle-based RF stage was 0.21 with 95 percent CI of [.19, .24]; and that of the driver-based RF stage was 0.34 with 95 percent CI of [.31, .37].

The intoxication detection algorithm was also applied to the drowsiness and distraction datasets. Here we are reasonably certain that true BAC levels were zero in both cases, so the assumption of no intoxication is the most reasonable one to make. In the drowsiness dataset, both RF and HMM specificity is reported due to an interesting anomaly in the results. The specificity of the RF stage is 0.91 with 95 percent CI of [.89, .95], while the corresponding HMM output had a specificity of only 0.11 (95 percent CI: [.08, .15]). The driver-based algorithms with head pose signals had an RF sensitivity of 0.99 (95% CI: [.99, 1.0]) and an HMM sensitivity of 1.0 (95% CI: [1.0, 1.0]). Meanwhile, the sensitivity in the distraction dataset for both the vehicle-based and driver-based algorithms was 1.0 with 95 percent CI of [1.0, 1.0]. The gap between the RF and HMM specificity of the vehicle-based algorithm might be closed by simply tuning the HMM parameters further. The result is interesting because such fine-tuning was not needed in any of the other HMMs.

Discussion

A distraction algorithm was successfully developed using the AttenD distraction metric as a ground truth signal. This choice is appropriate for visual and visual-manual types of distraction and matches well with the current wisdom on the risk of glances away from the road. However, it does not consider the total engagement time that is part of NHTSA's distraction guidelines.

Not surprisingly, the use of head pose data improved the performance of the distraction-detection algorithm. Head pose did not provide a benefit to the drowsiness or intoxication-detection algorithms. One might expect that head pose data would include nods that show 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.

Z-score normalization for smoothing away individual differences between drivers had a beneficial effect on the drowsiness and intoxication detection algorithms. Further investigation could reveal clever ways that the individual driver performance could be identified and used for customization of the algorithms. Why did the use of z-scores not benefit the distraction algorithm? It may be that the short, yet intense, tasks that typify serious distraction impose a kind of structural uniformity to drivers' performance decrements.

The algorithms had mixed results when generalizing to other impairment datasets. The distraction algorithm generalized much better when it included head pose data. Many 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, it is possible that the performance decrements observed through vehicle-based sensors are heavily dependent on the task, but relatively uniform across drivers. They are 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 very well to the drowsiness dataset. However, it would have issued some false positive classifications to the drowsy driver.

The drowsiness algorithm did not generalize well to the distraction dataset, and it was difficult to characterize its performance on the intoxication dataset without having the same ground truth available. We attempted to differentiate drowsy lane departures from non-drowsy lane departures in training the algorithm; however, it may be that lane departures caused by distraction and intoxication are like drowsy lane departures and have similar driving performance characteristics as their precursors.

The timeliness of each algorithm was designed to match an estimated time constant present in each type of impairment that characterizes how quickly it might lead to changes in driving behavior. The distraction timeliness is around 1 second, while the drowsiness timeliness is between 6 and 60 seconds. The timeliness of the intoxication algorithm is greater than 2.5 minutes and could be as long as 5 minutes. We assume that running the three algorithms on different timescales helps them be less sensitive to each other's impairments.

There are no identified barriers to implementing any of the proposed algorithms in the simulator environment. Such an implementation was completed in the mitigation track for a drowsiness algorithm, and the same core implementation could be used for the system of algorithms described here. The fastest update rate is 1 second, which is the most computationally demanding of the three; however, a C++ implementation of RF inference was able to run in less time.

The algorithms made use of some environmental variables, such as road demand metric and road curvature. These variables served as proxies for signals that might normally be obtained from a front-facing camera/sensor and GPS/GIS information. The road demand metric was a relatively important variable in the drowsiness algorithm with z-score normalization. The inclusion of time-of-day would dramatically improve the performance of the drowsiness algorithm as most drowsy lane departures occurred in the late-night condition. A commercialized algorithm could seek to take advantage of environmental variables as much as possible, as those variables may help in separating different regimes of normal and impaired driving performance.

A mitigation system was implemented in the mitigation track that studied different modalities and types of warnings. In a system of algorithms, an arbitration layer would select and prioritize among the types of impairments being reported by the underlying algorithms. Distraction warnings may benefit from being prioritized highest due to their highly temporary nature and the ability for a crash to happen very quickly. It has been shown that the drowsiness algorithm can detect drowsy driving in advance of an adverse event, so there is some leeway to delay a drowsy warning until a distraction warning is completed. It is unknown what benefit may be gained from an intoxication mitigation system, or what driver vehicle interface system should be used.

Limitations

Several limitations are worth noting in the present research. As shown in the data, there was significant variability among individuals 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 may be important for future refinement of state detection algorithms. The similarity of intoxication and drowsiness also 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.

Finally, the present research represents an important step in establishing a database of impaired driving data based on a common set of scenarios. The DrIIVE dataset may 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

Anttila, V., & Luoma, J. (2005). Surrogate in-vehicle information systems and driver behaviour in an urban environment: A field study on the effects of visual and cognitive load. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 121–133.

Ashton, W. D. (1972). *Logit transformation: With special reference to its uses in bioassay*. Macmillan Publishing Company.

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, T., Lee, J., Schwarz, C., Fiorentino, D., & McDonald, A. (2014). *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., Gasper, J., Schwarz, C., Lee, J., Schmitt, R., Marshall, D., & Ahmad, O. (2026). *Driver monitoring of inattention and impairment using vehicle equipment — Mitigation track: Assess potential countermeasures for drowsy driving lane departures* (Second of three parts. Report No. DOT HS 813 564b). National Highway Traffic Safety Administration.

Dingus, T. A., Hulse, M. C., Antin, J. F., & Wierwille, W. W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research Part A: General*, *23*(4), 301–315.

Heitbrink, D., Brown, T. L., & Ahmad, O. (2013). *DrIIVE phase track A (distraction) pilot scenario specification* (Report No. N2013-007). University of Iowa, Driving Safety Institute.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel* (Report No. 1-1-1975). University of Central Florida, Institute for Simulation and Training. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

Kircher, K., & Ahlström, C. (2009). *Issues related to the driver distraction detection algorithm AttenD*. Swedish National Road and Transport Research Institute. www.diva-portal.org/smash/get/diva2:817352/FULLTEXT01.pdf

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.ntl.bts.gov/view/dot/62931

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

Lee, J. D., Fiorentino, D., Reyes, M. L., Brown, T., Ahmad, O., Fell, J., Ward, N., & Dufour, R. (2010, August). *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. https://www.nhtsa.gov/sites/nhtsa.gov/files/811547a.pdf

Marshall, D. C., & Schmitt, R. A. (2015). *Participant orientation PowerPoint presentation: DrIIVE track A* (Report No. N2015-005). University of Iowa. www.nads-sc.uiowa.edu/publicationStorage/20151432761486.N2015-005_Participant%20Orientat.pdf

Mattsson, K. (2007). *In-vehicle prediction of truck driver sleepiness*. (Master's thesis). Luleå University of Technology, Luleå, Sweden. https://ltu.diva-portal.org/smash/get/diva2:1023537/FULLTEXT01.pdf

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2007). The prevalence of, and factors associated with, serious crashes involving a distracting activity. *Accident Analysis & Prevention*, *39*(3), 475–482.

National Highway Traffic Safety Administration. (2013). *Visual-manual NHTSA driver distraction guidelines: In-vehicle electronic devices* [Docket submission, Document ID NHTSA-2010-0053, in Regulations.gov]. www.regulations.gov/document/NHTSA-2010-0053-0135

Östlund, J., Peters, B., Thorslund, B., Engström, J., Markkula, G., Keinath, A., Horst, D., Juch, S., Mattes, S., & Foehl, U. (2005). *Driving performance assessment – Methods and metrics* (AIDE Deliverable 2.2.5). AIDE Consortium.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77.

Schmitt, R., & Brown, T. (2015). *Driver monitoring of inattention and impairment using vehicle equipment (DrIIVE) – Track A: Develop and evaluate a system of algorithms to identify signatures of alcohol-impaired, drowsy, and distracted driving: Study documents* (Report No. N2015-012). The University of Iowa. www.nads-sc.uiowa.edu/publications.php?specificPub=N2015-013

Victor, T., Dozza, M., Bärgman, J., Boda, C. -N., Engström, J., Flanagan, C., Lee, J. D., & Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk*. The National Academies Press. https://doi.org/10.17226/22297

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.

Wang, J. J., Xu, W., & Gong, Y. H. (2010). Real-time driving danger-level prediction. *Engineering Applications of Artificial Intelligence*, *23*(8), 1247–1254. http://doi.org/10.1016/j.engappai.2010.01.001

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

Young, M. S., Mahfoud, J. M., Walker, G. H., Jenkins, D. P., & Stanton, N. A. (2007). Crash dieting: The effects of eating and drinking on driving performance. *Accident Analysis & Prevention*, *40*, 142–148.

This page is intentionally left blank.

.

Appendix A: Documentation of New Distracted Driving Dataset

As part of this research effort, additional driving distraction data was collected to supplement existing datasets that had been previously used in the development of algorithms to detect distracted driving using vehicle-based sensors. The purpose of this appendix is to document the decisions made in designing the scenarios, the methodology used for the collection of the data, and the validation of the data showing effects of distraction on driving performance.

Review of the Literature

The purpose of the following literature review was: (1) to identify secondary tasks for the distracted driving data collection, and (2) to identify vehicle-based distracted driving detection algorithms that have been used to detect driver impairment. The literature review concluded with a recommended system of algorithms to detect and distinguish among the three types of impaired driving using vehicle-based and driver-based algorithms to assess the benefit of adding driver-based algorithms.

Based on this literature review and the requirements of the algorithm, two tasks were selected:

Visual-Only Task: Read aloud text from display located to require a head turn. For each task, an auditory prompt was provided to alert the driver that the message was ready to be read. Messages did not exceed 160 characters. Messages contained “interesting facts” and were written for a Flesch-Kincaid grade level between 6.5 and 8. The subject read the message aloud. In previous implementations, this task was available only for a limited time, requiring a single glance that commenced within one second of the prompt ending to complete successfully. For this experiment, the time limit was eliminated to make the task self-paced. The researcher cleared the message from the display when the subject completed reading the text.

Visual/manual Task: Scrolling and selecting songs in list. This task was used by Lee et al. (2012) and described as such:

For each search task, a synthetic voice instructed participants to engage in the task at specified points during the drive. The music search tasks required drivers to scroll through a short (20 songs), medium (75 songs), or long (580 songs) playlists sorted alphabetically by song title to find a target song. Drivers had to scroll through 7 to 12 songs in the short playlist, 21 to 30 songs in the medium playlist, and 100 to 130 songs in the long playlist to find the target song. These lengths were selected to ensure that the short playlist task required little to no scrolling, the medium playlist task required one full scroll, and the long playlist task required multiple scrolls. Drivers responded verbally with the song number in the playlist (which was displayed) to indicate task completion. For each search task, drivers were required to find a different target song; that is, drivers did not search for the same song twice.

Methodology

This section provides an overview of the methodology used for collecting the self-paced distraction data used in the algorithm development. Schmitt and Brown (2015) provides all the study documents used in this data collection.

Experimental Design

Independent variables

A 3 x 3 x 2 mixed design exposed male and female participants to three distraction levels in three different orders. Between-subjects independent variables were gender and order of the distraction drives. Gender was balanced across participants and order was balanced using a single Latin square. The within-subject independent variable was distraction: no distraction, moderate distraction, and high distraction. Due to sample size limitations and a desire to minimize variability across that sample, all participants were age 21 to 34 (the youngest age group from phase 1) to capture a group of subjects that were more likely to multitask while driving and thus were more likely to have developed strategies for managing this engagement that were important to capture when considering self-paced engagement with tasks. This choice allowed for comparison with prior data.

Distraction tasks and difficulty

The purpose of the secondary tasks was to provide a distracting task of a known duration, for which the beginning and end of engagement were clearly recorded. This allowed clear demarcation in the driving performance data of when the driver was or was not engaged in the task. In addition, the task was self-paced to allow natural chunking and interruptions of the task. Secondary tasks were selected based on the literature review described above. The locations of the visual display and manual interaction were such that engaging in the tasks required a head turn by the driver.

Visual-Only Distraction Task

This visual-only task required drivers to read aloud text from a display. Each message was roughly 160 characters. Messages contained "interesting facts" and were written for a Flesch-Kincaid (Kincaid et al., 1975) grade level between 6.5 and 8. For each task, an auditory prompt, "Read Message," alerted the driver that the message was ready to be read. The subject read the message aloud. Task engagement began when the participant spoke the first word of the message and ended with the last word of the message. A voice key was used to mark these beginning and end points in the data stream. A researcher recorded task performance by noting correct and incorrect words while the participant read the message. The message was cleared from the screen remotely by the researcher when completed.

Visual-Manual Distraction Task

The visual-manual task required drivers to scroll through a list of songs and find a target song from the list as described by Lee et al. (2012). The moderate and high levels of distraction were varied using the length of the list of songs, three pages and five pages respectively. The number of songs and position of the target song ensured that the moderate level of the task required one full scroll and the high level of the task required multiple scrolls. The target song never appeared on the first page and always required at least one scroll action. Each target song was presented to drivers only once and was not repeated.

Drivers received an audio prompt, "Find [song title]," instructing them to begin the task. The beginning of task engagement was recorded by the driver's first touch to the display screen to initiate a scroll. Drivers responded by manually scrolling to the song number of the target song marking the end of task engagement.

Order

To accommodate learning effects, the order of the drives for the different levels of distraction were counterbalanced using a Latin square. Driving sessions in the simulator alternated between participants, to reduce carryover effects from one distraction level to another and to reduce simulator-induced fatigue.

Potential intervening variables

A few intervening variables could influence dependent variables or mediate the influence of distraction on the dependent variables. Table A-1 summarizes these variables and the approach to minimize any confounding they might have produced. The potential confound of fatigue caused by long uninterrupted sessions in the driving simulator was minimized by scheduling two participants for concurrent data collection visits and allowing rest periods between each of the data collection drives.

Table A-1. Intervening variables that might moderate the effect of distraction

Variable	Rationale	Mitigation
Unintended Distraction	Inadvertently distracting participants during the baseline drive would limit the usefulness of the data	Efforts were made to reduce exogenous stimuli other than the distraction tasks, thereby ensuring that distraction effects were the result of the tasks. Participants were also allowed a practice drive to become familiar with the simulator.
Physical Fatigue	Excessive physical activity may confound the effects of distraction, particularly if present during the second or third simulator drive	Participants had recovery periods between simulator drives.
Ambient temperature	Can impact simulator sickness	Temperature was set at 72° F.
Simulator sickness score	Higher levels of simulator sickness can result in potential changes in driver behavior to counteract its effects	Participants had a practice drive during their screening visit. Participants with high scores (Nausea $\geq$ 21, Oculomotor $\geq$ 32, Disorientation $\geq$ 15, or Total $\geq$ 32) related to the screening drive were excluded from the study.

Dependent Variables

The dependent variables varied across the 19 distinct events that comprised the 3 segments of the drive. The primary measures related to lane position, speed, and steering. The scenario specification describes the dependent variables for each event (Heitbrink, Brown, & Ahmad, 2013). Table A-2 shows the list of driver performance measures used in the validation analysis. These measures were tested to determine their sensitivity to detect distracted tasks.

Table A-2. Summary of simulator operational measures for distraction

Measure	Description
NumLaneDeparts	Number of lane departures. Includes variations for left and right departures
SDLP	Standard deviation of lane position (cm)
MeanSpeedNorm	Average vehicle speed relative to speed limit (mph)
PctSpeedHigh	Percent time vehicle speed is 10 percent greater than speed limit (%)
PctSpeedLow	Percent time vehicle speed is 10 percent less than speed limit (%)

Participants

Twenty-four participants completed all 3 drives of the study in a single session – one with no distraction, one with moderate distraction, and one with high distraction. They were healthy men and women aged 21 to 34 with a valid driver's license. Efforts were made to recruit a racially and ethnically diverse sample. Participants were paid $150 for completing the single study session. Pro-rated compensation was provided for participants who did not complete the study. Key inclusion criteria included having been licensed for at least 1 year, driving at least 10,000 miles per year, no license restrictions other than for vision, and good general health. The study required 41 participants (see Table A-3) to enroll to achieve 24 completed participants. Of the 17 who enrolled but did not complete, 5 did not pass the screening visit, 8 were unable to meet the study schedule or were screened as backups, and 4 either cancelled their visit or did not show up at their scheduled time. Of the five who did not pass the screening visit, four showed a propensity for simulator sickness and one did not meet the breath alcohol requirements. Demographic information for the completed subjects is provided in Table A-4.

Table A-3. Number of participants reporting to visits for main study

	Young Male	Young Female	Total
Enrolled	21	20	41
Complete	12	12	24
Screen Failures	2	3	5
Passed Screening, Unable to Schedule, or Alternate	5	3	8
Passed Screening, Cancelled Visit, or No Show	2	2	4

Table A-4. Demographic information for completed subjects

Gender	Age					
	Completed	**Mean**	**SD**	**Min**	**Median**	**Max**
Male	12	26.33	4.35	22	25	34
Female	12	28.17	3.44	23	29.5	34
Overall	24	27.25	4.02	22	26.5	34

Recruitment

Participants were recruited from the NADS participant database and through ads and referrals. An initial telephone interview determined eligibility for the study. Potential participants were screened for health history and current health status. Pregnancy, disease, sleep disorders, or evidence of substance abuse resulted in exclusion from the study. Potential participants taking prescription medications that cause or prevent drowsiness also were excluded from the study. Individuals were scheduled for a screening visit if all telephone-screening criteria were met. If all eligibility requirements were met during the screening visit, participants were scheduled for the data collection visit.

Screening Procedures

During the screening visit, each participant first gave informed consent to participate in the study and received a copy of the signed informed consent form. They also filled out a video release form and a payment voucher. During this time, a researcher confirmed the participant held a valid driver's license. Next, a breath alcohol measurement was taken to confirm that the participant was not under the influence of alcohol.

Eligible participants viewed an orientation and training presentation (Marshall & Schmitt, 2015) that provided an overview of the simulator cab and the distraction tasks they were asked to complete while driving. Participants practiced the distraction tasks until they accurately completed the task three successive times. Participants then completed an approximately eight-minute practice drive that included making a left-hand turn, driving on two- and four-lane roads, and practicing each of the distraction tasks while driving. They received recorded audio navigational instructions to guide them through the route. After the drive, participants completed a wellness survey that asked questions about how they felt. Following NADS standard operating procedure, this survey was completed following completion of the practice drive to determine level of simulator sickness. If the survey showed a propensity for simulator sickness based on a total score greater than 32, a nausea score greater than 21, an oculomotor score greater than 32, or a disorientation score greater than 15, the participant was ineligible to continue. If still eligible, the participant was scheduled for a data collection visit.

Apparatus

Simulator and Sensor Suite

NADS is located at the University of Iowa's Oakdale Campus. It consists of a 24-foot dome in which an entire car 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 she/he were driving a real vehicle. Each of the three front projectors has a resolution of 1600 x 1200; the five rear projectors each have a resolution of 1024 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 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 were visible, a fixated gaze may be measured with an RMS 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 A-1.

Figure A-1. FaceLab cameras mounted in the Malibu cab with a separate head tracking system mounted between them

Driving Scenarios

The scenarios were the same as those in phase 1 of this study, except for the gravel extension and a 10-minute straight segment at the end of the drive.[2] Each drive was composed of three nighttime driving segments. The drives started with an urban segment composed of a two-lane roadway through a city with posted speed limits of 25 to 45 mph with signal-controlled and

[2] Equivalent to scenario used for the IMPACT study (Lee et al., 2010).

uncontrolled intersections. An interstate segment followed that consisted of a four-lane divided expressway with a posted speed limit of 70 mph. 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 location via an interstate. Events in each of the three segments combined to provide a representative trip home in which drivers encountered situations they might experience in a real drive. Scenario events are summarized in Table A-5.

Table A-5. Summary of events for each of the three segments of the drive. Events containing secondary tasks are italicized and underlined

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
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 distraction 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 are faded and the road surface is grayish. The Hairpin Curve of the IMPACT scenarios is not included.	270
	Dark Rural Hairpin Curve (304HP)	A hairpin turn and a vertical curve located within the Dark Rural (304). Data for Dark Rural and Hairpin Curve are mutually exclusive.	30
	Transition to Straight Segment (305)	Transition to gravel rural	30
	Gravel Rural (306)	Dark, gravel rural road	90

Throughout the urban section, a series of potential hazards required drivers to scan the roadside. These hazards included pedestrians, motorbikes, and cars entering and exiting the roadway. These hazards had paths that did not cross the driver's path.

Each participant drove the simulator three times: once in the baseline distraction condition, once in a moderate distraction condition, and once in a high distraction condition. Three scenarios with varied event orders were used to minimize the learning effects from one drive to the next. For each of the three scenarios, there were the same number of curves and turns, but their order varied. For example, the position of the left turn in the urban section varied so that it was located at a different position for each drive. The order of the left and right rural curves also varied between drives. These were the same drives used for phase 1 of this project and the IMPACT study, but with adaptations at the end to accommodate transition to the added long straight rural road.

Experimental Procedures

At the main data collection visit, each participant completed three data collection sessions all occurring during one study visit. Two participants completed the study concurrently to allow time between drives with different distraction levels, as shown in Table A-6.

Table A-6. Illustration of concurrent participant appointments

Participant A	
Report & Task Review	**Participant B**
First Simulator Drive	Report & Task Review
Recovery Period	First Simulator Drive
Second Simulator Drive	Recovery Period
Recovery Period	Second Simulator Drive
Third Simulator Drive	Recovery Period
Debrief	Third Simulator Drive
	Debrief

Participants drove themselves to the facility. Upon arrival, participants reviewed the distraction task instructions and were allowed to practice to the same criterion if the visit was on a separate day from the screening appointment. The participants then were escorted to the simulator and eye-tracking calibrations were completed. The participant completed three data collection drives with waiting or recovery periods between each drive. The wellness survey was administered immediately following each drive to ensure continued absence of simulator sickness per NADS protocol. The waiting period between each drive minimized the potential for carryover effects of the distraction from one drive to another and alleviated the potential for fatigue induced by very long sessions in the driving simulator. The debriefing followed the third simulator drive for each participant. Participants who completed were compensated $150 for their time. Participants who did not complete the study were provided with pro-rated compensation.

Validation Results

The purpose of the validation analysis was to verify that differences were present in the driving performance data between the different levels of distraction tasks to verify that their use for the algorithm development work. A conservative approach was taken in this analysis by looking for differences in performance at the event level rather than just during the performance of the task. This approach is conservative because it also includes periods during which drivers were not directly engaged in the task.

Statistical Approach

Reduced data were examined for events in which distraction tasks were included for both distraction drives. Corresponding event data from baseline drive was also included for these events. The events included in the analysis are documented in Table A-7. For these events, density plots by event and subject were used to verify uniformity of the data for analysis purposes. The SAS general linear models procedure was used to conduct an analysis of variance for each dependent measure. The SAS PLM was used to perform post-fitting statistical analyses on the interactions. A t-test with a Bonferroni adjustment was used to identify differences between levels within an independent variable. For the presented box plots, the whiskers represent the minimum and the maximum excluding outliers.

Table A-7. Events included in validation analysis

Urban Events	Interstate Events	Rural Events
Urban Drive (102)	Following (203)	Dark Rural (304)
Yellow Light Dilemma (104)	Interstate Curves (205)	Gravel Rural (306)
Urban Curves (106)		

Independent Variables

Three independent variables were considered in this analysis: event, distraction task type, and subject. As detailed in Table A-7, three urban events, two interstate events, and two rural events were included in the analysis. The three levels of distraction tasks were none, moderate distraction tasks, and difficult distraction tasks. These variables were nested under subject to account for within-subject variability.

Dependent Measures

Dependent measures included standard deviation of lane position (SDLP), lane departures per minute (which normalized lane departures across events of different length), speed in relation to the speed limit (mean speed to posted speed limit), and the percentage of time speed was 10 percent above (high) or 10 percent below (low) the posted speed limit. Note that a logit transform was performed on percent speed high and low.

Results

Standard Deviation of Lane Position

There were significant effects associated with the task difficulty ($p = .0009$), event ($p < .0001$), and their interaction ($p = 0.0001$). The simple effects based on task difficulty show significant difference by task difficulty for all events except the urban drive and urban curves ($p < .05$). As can be seen in Figure A-2, for the three higher speed events there was a trend toward increasing SDLP with increasing task difficulty. For the urban yellow light dilemma event, the moderate distraction task had the greatest SDLP. The overall trend across all of the events (see Figure A-3) is for increasing SDLP with increasing task difficulty.

When considering the effects of the distraction on SDLP, two general conclusions can be drawn. First, there were differences between subjects in their lane keeping variability, with some individuals showing greater variability than others, even without the distracting task. Second, and most importantly, variability in lane keeping was greater in the presence of distracting tasks with increases in SDLP of 1 to 3 cm.

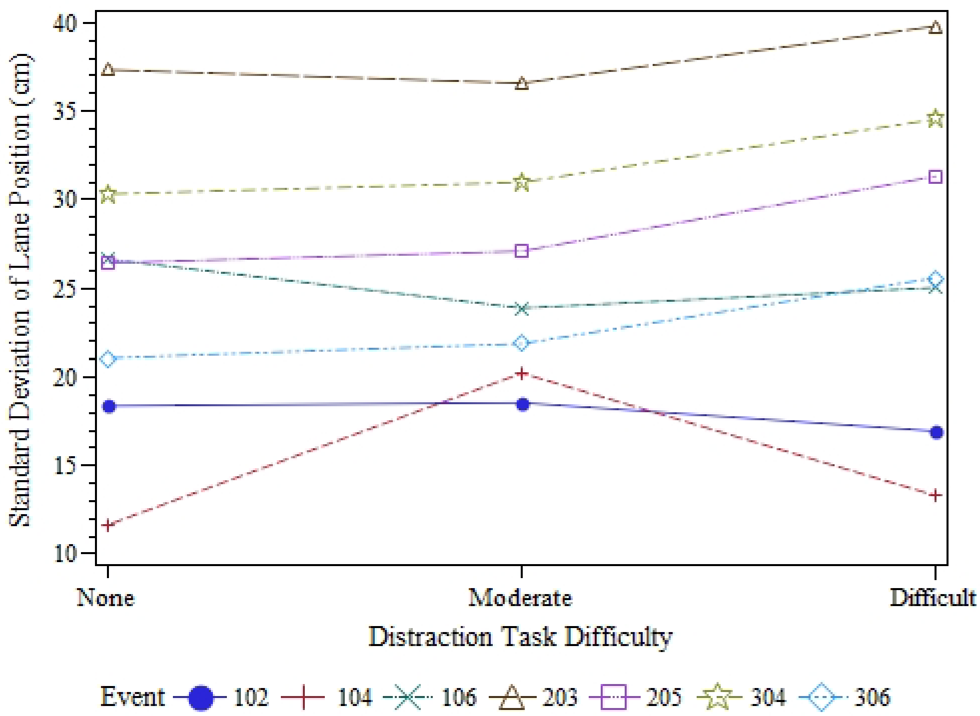

Figure A-2. Interaction effect for SDLP with all subjects for urban drive (102), yellow light (104), urban curves (106), following (203), interstate curves (205), dark rural (304), and gravel rural (306). For three higher speed events with a significant effect, there was a trend towards increasing SDLP with increasing task difficulty.

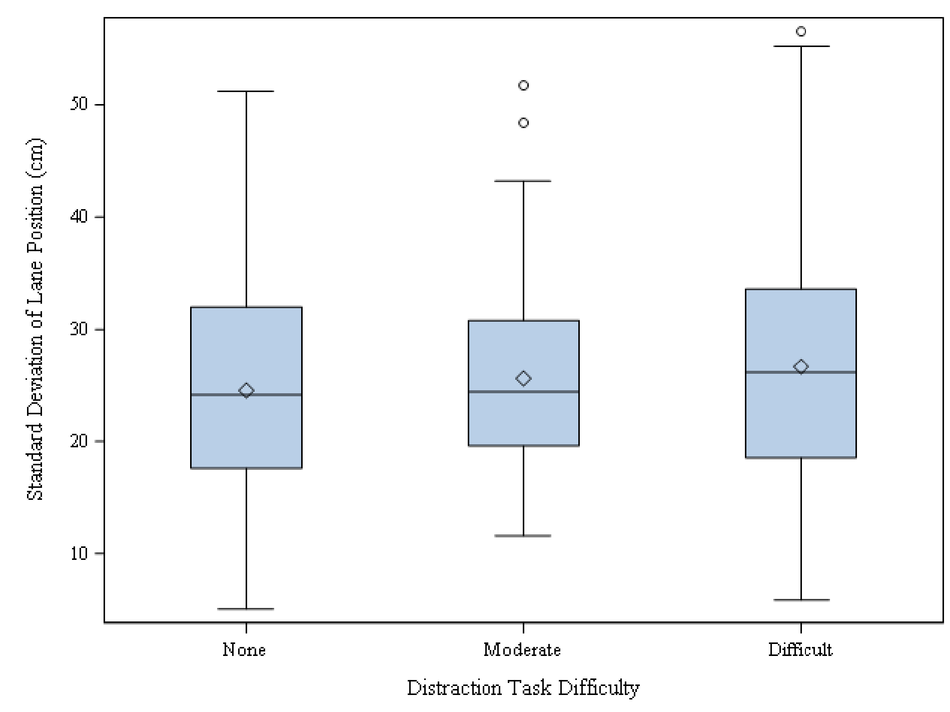

Figure A-3. Mean SDLP increases with increased task difficulty

Lane Departures per Minute

The review of the distribution of the data revealed a single outlying subject that had significantly more lane departures even during the baseline drive that needed to be removed from this analysis. There were significant effects associated with the task difficulty (p = .0008), event (p < .0001), and their interaction (p < .0001). The simple effects based on task difficulty show significant difference by task difficulty, as with SDLP, for all events except the urban drive and urban curves (p < .05). As can be seen in Figure A-4, for the three higher speed events with a significant effect, there was a trend towards increasing lane departures with increasing task difficulty. There was no effect of distraction on lane departures for any of the urban events, and lane departures were not possible on the gravel. The overall trend across all the events (see Figure A-4) was for an increased rate of lane departures when performing the difficult distraction task compared to the moderate or no distraction cases (p < .05).

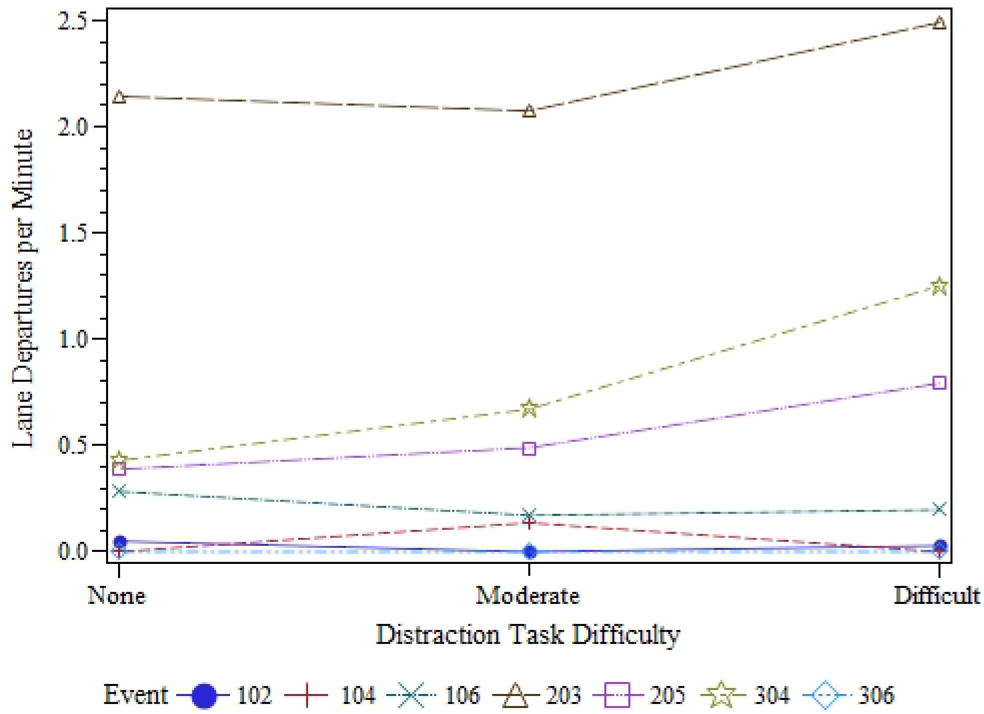

Figure A-4a. Interaction effect for lane departures per minute for urban drive (102), yellow light (104), urban curves (106), following (203), interstate curves (205), dark rural (304) and gravel rural (306). For three higher speed events with a significant effect, there was a trend towards increasing SDLP with increasing task difficulty.

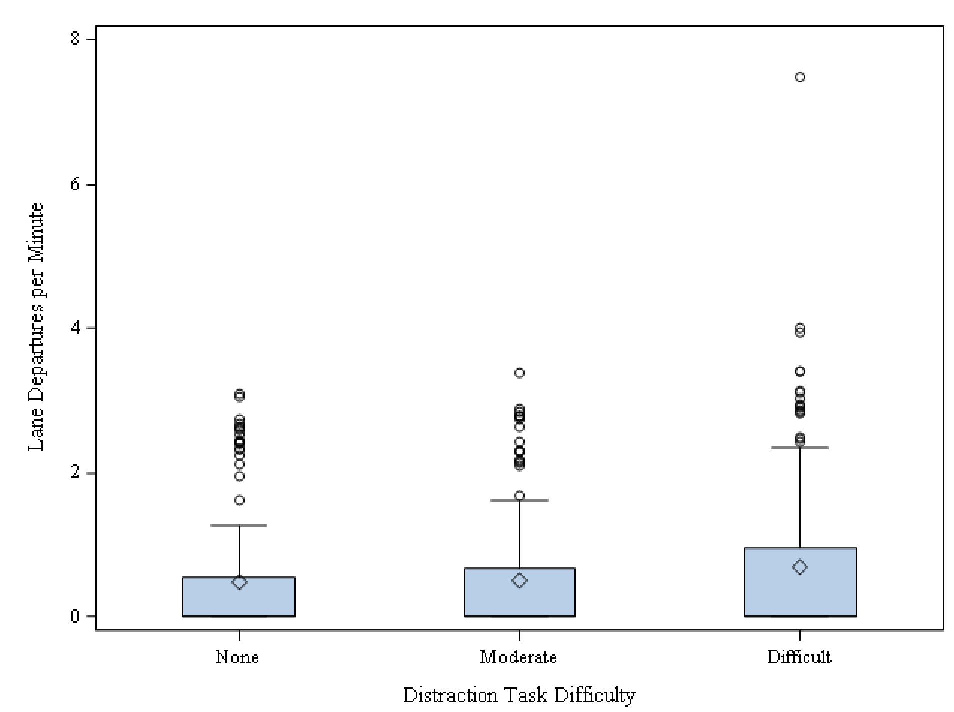

Figure A-4b. Lane departures per minute increase with increased task difficulty

Mean Speed Relative to the Speed Limit

There were significant effects associated with the task difficulty (*p* = .0076) and event (*p* < .0001). The overall effect across all the events (see Figure A-5) was for decreased average speed when performing the difficult distraction task relative to the moderate or no distraction cases (*p* < .05). There were significant differences between events (*p* < .05) that divided the events into three distinct groups. The two lowest speed events (urban drive and yellow light dilemma) produced mean speed greater than the speed limit of 0.6 to 1.8 mph. The dark rural drive (with curves) produced the greatest relative speed below the speed limit at -4.2 mph. The remaining four events resulted in mean speeds that were 1 to 2 mph below the speed limit.

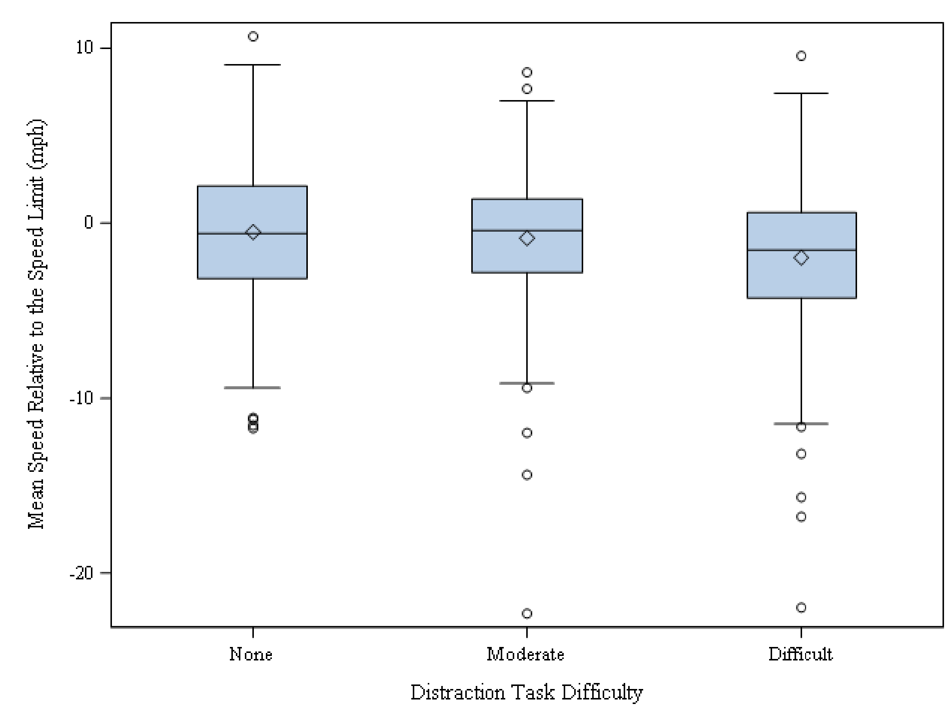

Figure A-5. Effect of task difficulty on mean speed relative to the speed limit

Percent Speed Low

Due to restriction in range for this variable, data were analyzed using a logit transformation (Ashton, 1972). There were significant effects associated with task difficulty (*p* = .0266) and event (*p* < .0001). For ease of understanding of the effects of distraction, the transformed data are presented in Figure A-6 (upper) and the raw untransformed data are presented in Figure A-6 (lower). The overall effect across all the events was for an increased percentage of time more than 10 percent below the speed limit when performing the difficult distraction task (*p* < .05) relative to the moderate or no distraction cases. There were significant differences between events (*p* < .05) that divided the events into four groups. The curviest roads (urban curves and dark rural) had the greatest percentage of time driving slowly, while the urban drive had the least percentage of time with slow driving.

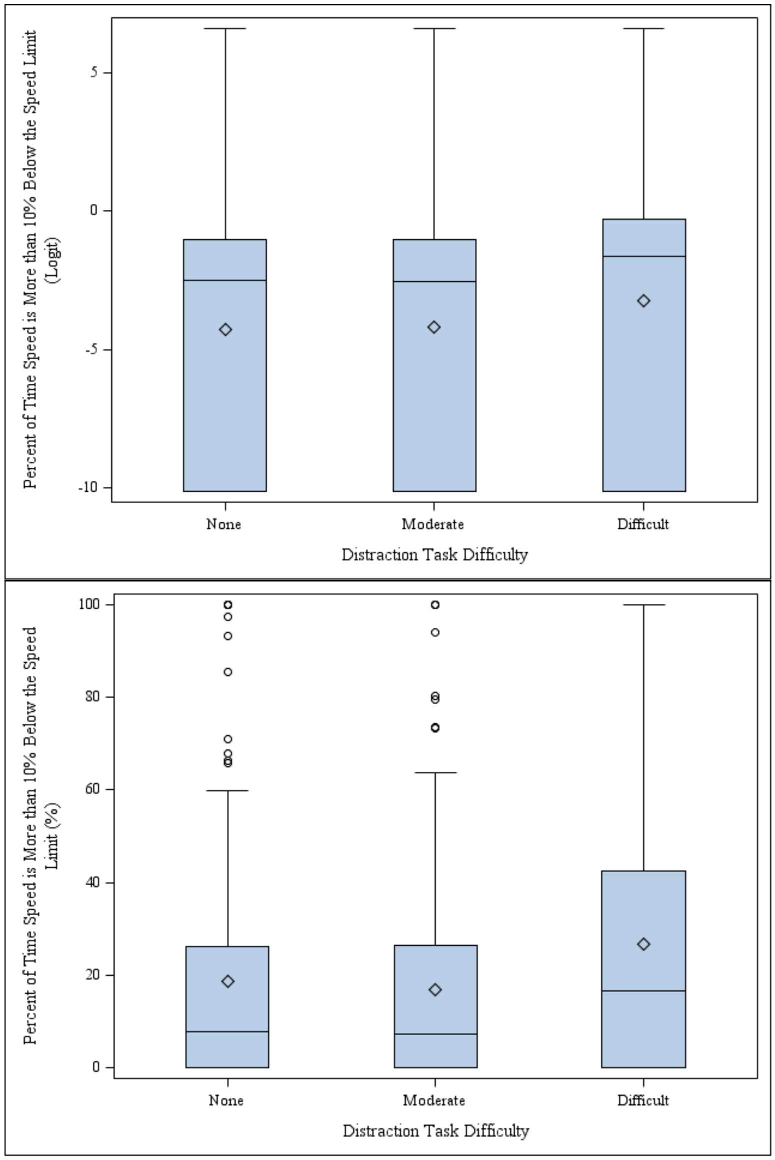

Figure A-6. Percent of time more than 10 percent below the speed limit increased with difficult distraction task. Logit transformation (above) and raw results (below)

Percent Speed High

There were significant effects associated with event ($p < .0001$), but not by task ($p = .0948$). There were significant differences between events ($p < .05$) that divided the events into four distinct groups. The roads with the lowest speed limits had the greatest percentage of time with higher speeds. The urban curves and gravel rural had the next greatest percentage, followed by the dark rural. The interstate events had the least percentage of time with high speeds.

Discussion

As the aim of this analysis was to determine if the distraction tasks resulted in changes in performance such that the data would be usable for the algorithm work, this analysis focused on changes in control of lane position and speed in the general geographic area (events) where engagement with the secondary tasks took place. Although an analysis at a greater level of granularity would likely have resulted in stronger effects, the event level was selected to provide a conservative estimate of the effects that will be present in the raw data used for the algorithm.

Overall, distraction effects were observed on both speed and lane keeping, although the lane keeping effects varied by event with some higher speed events showing changes in performance. Table A-8 summarizes the overall effects and Table A-9 provides Cohen's *d* effect sizes relative to the baseline condition by difficulty level. For lateral control, lane keeping variability increased with distraction task difficulty and resulted in more lane departures while engaged in the difficult distraction task. For longitudinal control, drivers drove slower relative to the speed limit while engaged in the difficult distraction task, resulting in a greater percentage of time driving more than 10 percent below the speed limit. These results, summarized over events in which distraction tasks occurred, suggest that the effects of the distraction tasks are sufficiently large that they are observable at this level of granularity of analysis.

Table A-8. Effect of distraction tasks on analyzed measures

Dependent Measure	**Effect of Moderate Distraction**	**Effect of Difficult Distraction**	**Interactive Effects**
SDLP	-	-	Yes
Lane Departures		-	Yes
Mean Speed		-	No
% Speed Low		-	No
% Speed High			No

Table A-9. Cohen's d effect sizes for the 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

Appendix B: Algorithm Measures

Table B-1. Raw data recorded for all algorithms

Variable	Description
CFS_Accelerator_Pedal_Position	Throttle pedal position (0-1)
CFS_Brake_Pedal_Force	Brake pedal force (lbf)
CFS_Steering_Wheel_Angle	Steering wheel angle (degrees)
SCC_Lane_Deviation_Lane	Current lane ID
SCC_Lane_Deviation	Lane position (ft)
SCC_Lane_Deviation_Width	Width of current lane (ft)
SCC_Lane_Depart_Warn_LeftDist	Distance from left front corner of vehicle to left lane edge (ft)
SCC_Lane_Depart_Warn_RightDist	Distance from right front corner of vehicle to right lane edge (ft)
SCC_Lane_Depart_Warn_Heading	Current heading angle in the lane (degrees)
SCC_OwnVeh_Curvature	Radius of curvature of road (ft)
VDS_Veh_Speed	Vehicle speed (mph)
VDS_Chassis_CG_Accel_Fwd	Longitudinal vehicle acceleration (ft/s/s)
VDS_Chassis_CG_Accel_Right	Lateral vehicle acceleration (ft/s/s)
StrDes	Double exponentially smoothed steering wheel angle
StrdDes	Double exponentially smoothed steering wheel angle rate
Events	Current event number
Order	Current event ordinal value (count)
isLaneDeparture	Is the vehicle currently departed from the lane? (binary)
isLaneChange	Is the vehicle currently changing lanes (binary)
PRC17s	Percent road center over previous 17 seconds (%)
AttendD	AttenD buffer output (seconds)
Speed_Limit	Speed limit (mph)
Steer_Reverse_VLarge	Very large steering reversals [Gap size = 6°, LPF cutoff = 0.6 Hz](Anttila & Luoma, 2005)
Steer_Reverse_Large	Large steering reversals [Gap size = 3°, LPF cutoff = 0.6 Hz]
Steer_Reverse_Small	Small steering reversals [Gap size = 0.1°, LPF cutoff = 2 Hz]
TLC	Time to lane crossing (seconds) (Mattsson, 2007)
TLC_Minima	Time to lane crossing minima (Östlund et al., 2005)
Lane_Deviation_Modified	Modified lateral position variation (Östlund et al., 2005) [Note: misnames in files as Lane_Departure_Modified]
Road_Demand_Metric	Road demand metric (0 to 100) (Dingus et al., 1989)
Track_State	Eye tracking state
Gaze_Qual_L	Left eye gaze quality
Gaze_Qual_R	Right eye gaze quality
Gaze_Rot_X	Gaze rotation on X axis
Gaze_Rot_Y	Gaze rotation on Y axis
Gaze_Proj_Horiz	Horizontal gaze projected onto camera plane

Variable	Description
Gaze_Proj_Vert	Vertical gaze projected onto camera plane
OncenterGaze	On center gaze, normalized using radius of 8° (see Victor et al., 2005). Has value of one at the ellipse boundary
isOnCenter	Is the on center gaze measure less than one?
Gaze_Horiz_0	Calibrated horizontal gaze center
Gaze_Vert_0	Calibrated vertical gaze center
Gaze_Location	Coded gaze location in world (uses numbers instead of names)
Gaze_Front_Cone	Is the gaze inside a 90° front cone? (used for AttenD)
Hpos_Conf	Head position confidence
Hpos_Filt_X	Filtered head position on X axis
Hpos_Filt_Y	Filtered head position on Y axis
Hpos_Filt_Z	Filtered head position on Z axis
Hrot_Filt_E0	Head rotation Euler angle 0
Hrot_Filt_E1	Head rotation Euler angle 1
Hrot_Filt_E2	Head rotation Euler angle 2
Head_Horiz_Rot_0	Calibrated horizontal head rotation center
Head_Vert_Rot_0	Calibrated vertical head rotation center
Head_Horiz_Pos_0	Calibrated horizontal head position center
Head_Vert_Pos_0	Calibrated vertical head position center
Head_Depth_Pos_0	Calibrated fore/aft head position center
Head_Horiz_Proj	Horizontal head rotation projected onto camera plane
Head_Vert_Proj	Vertical head rotation projected onto camera plane

Table B-2. Measures aggregated into 1 second segments, common among all algorithms

Measure	Description
subject_first	Participant number for drive
scenario_first	Scenario version of drive
attendd_max	Maximum AttenD buffer value in segment
str_des_mean	Mean value of double exponential smoothed steering wheel angle in segment
str_des_sd	Standard deviation of double exponential smoothed steering wheel angle in segment
strd_des_mean	Mean value of double exponential smoothed steering wheel angle rate in segment
strd_des_sd	Standard deviation of double exponential smoothed steering wheel angle rate in segment
throttle_mean	Mean throttle pedal position in segment
throttle_sd	Standard deviation of throttle pedal position in segment
brake_mean	Mean brake pedal force in segment
brake_sd	Standard deviation of brake pedal force in segment
steer_mean	Mean steering wheel angle in segment
steer_sd	Standard deviation of steering wheel angle in segment
lanedev_mean	Mean lane deviation in segment
lanedev_sd	SDLP in segment
curvature_max	Maximum road radius of curvature in segment
speed_mean	Mean vehicle speed in segment
speed_sd	Standard deviation of vehicle speed in segment
Ax_mean	Mean longitudinal acceleration
Ax_sd	Standard deviation of longitudinal acceleration
Ax_p2p	Peak-to-peak longitudinal acceleration over segment
Ay_mean	Mean lateral acceleration
Ay_sd	Standard deviation of lateral acceleration
Ay_p2p	Peak-to-peak lateral acceleration over segment
events_mode	Mode of event number in segment
prc17_max	Maximum of percent road center measure over segment
speedlimit_mode	Mode of speed limit over segment
srrvlarge_count	Number of very large steering reversals in segment
srrlarge_count	Number of large steering reversals in segment
srrsmall_count	Number of small steering reversals in segment
tlc_mean	Mean time to lane crossing
tlc_min	Minimum time to lane crossing
lanedevmod_mean	Mean modified lateral position variation
lanedevmod_sd	Standard deviation of modified lateral position variation
q_max	Maximum road demand metric over segment
headhoriz_mean	Mean projected horizontal head direction
headhoriz_sd	Standard deviation of projected horizontal head direction

Measure	Description
headvert_mean	Mean vertical projected vertical head direction
headvert_sd	Standard deviation of projected vertical head direction
headconf_mean	Mean head confidence value
headconf_max	Maximum head confidence value
headconf_pct	Percent of time with high head confidence in segment

DOT HS 813 564c
September 2026

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA

16175c-090226-v3