IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 19, 2021, accepted January 25, 2021, date of publication February 1, 2021, date of current version February 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056007

HARMONY: A Human-Centered Multimodal Driving Study in the Wild

ARASH TAVAKOLI[1], SHASHWAT KUMAR[1], XIANG GUO[1], VAHID BALALI[2], MEHDI BOUKHECHBA[1], (Member, IEEE), AND ARSALAN HEYDARIAN[1]

[1]Link Lab, Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA 22903, USA
[2]Department of Civil Engineering and Construction Engineering Management, California State University Long Beach, Long Beach, CA 90840, USA

Corresponding author: Arsalan Heydarian (ah6rx@virginia.edu)

ABSTRACT Effective shared autonomy requires a clear understanding of driver's behavior, which is governed by multiple psychophysiological and environmental variables. Disentangling this intricate web of interactions requires understanding the driver's state and behaviors in different real-world scenarios, longitudinally. Naturalistic Driving Studies (NDS) have shown to be an effective approach to understanding the driver's state and behavior in real-world scenarios. However, due to the lack of technological and computing capabilities, former NDS only focused on vision-based approaches, ignoring important psychophysiological factors such as cognition and emotion. The main objective of this paper is to introduce HARMONY, a human-centered multimodal naturalistic driving study, where driver's behaviors and states are monitored through (1) in-cabin and outside video streams (2) physiological signals including driver's heart rate and hand acceleration (IMU data), (3) ambient noise, light, and the vehicle's GPS location, and (4) music logs, including song features such as tempo. HARMONY is the first study that collects long-term naturalistic facial, physiological, and environmental data simultaneously. This paper summarizes HARMONY's goals, framework design, data collection and analysis, and the on-going and future research efforts. Through a presented case study, we first demonstrate the importance of longitudinal driver state sensing through using Kernel Density Estimation Methods. Second, we leverage the application of Bayesian Change Point detection methods to demonstrate how we can identify driver behaviors and responses to the environmental conditions by fusing psychophysiological information with features extracted from video streams.

INDEX TERMS Naturalistic driving study, physiological sensing, driver state detection, shared-autonomy, contextual awareness, human-in-the-loop systems.

I. INTRODUCTION

Although Autonomous Vehicles (AV) are improving at a very fast rate, it is predicted that through shared autonomy, humans will be involved in driving decision making for the foreseeable future [1], [2]. Shared autonomy is a promising approach where the human driver is kept in the loop to enhance situational awareness, response time in unsafe conditions, and trust in AV [1]. In principle, AV can act as an expert driver, deferring execution to the human user only in challenging scenarios. However, deferring execution while the human driver is in a sub-optimal state (e.g., stressed, sleepy, intoxicated) can be hazardous. Thus, it is essential for AV to accurately assess and respond to the driver's state and behavioral changes in real-time and according to each individual driver profile [3], [4].

Furthermore, research has shown that drivers exhibit considerable variability in their behavioral profiles in different contextual settings (e.g., their comfort level with autonomy or desire to take over in certain situations) [5]. However, currently, AV uniformly responds to different contextual settings solely based on outdoor environmental conditions and independent of the driver's behavior and comfort profile [5]. In order to increase the AV's safety, comfort, and reliability in different situations, the system should be individually tailored to each driver [6]. This is a key consideration to achieve an acceptable level of shared autonomy, where personalized profiles can be generated to inform AV's decision making according to the driver's preferences and comfort levels. This concept is referred to as deep personalization [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

Personalization, in turn, requires contextual awareness. Recent studies across engineering and social sciences emphasize that autonomous systems (e.g., AV, and smart devices) need to become contextually aware of the environmental changes to better predict and respond to each user's state and behaviors [7]. Context can be defined as any information that is relevant in defining a driving situation [7], including traffic patterns, scenery, passengers and in-cabin activities, driver behaviors, emotional states, and any other information that can be used to describe a driving event. The contextual setting is comprised of internal and external factors [8]. Internal factors include factors related to human's emotional, cognitive and attention states. External factors include in-cabin ambient conditions (i.e., noise, temperature, glare, lighting, music being played) and outside conditions such as traffic density, road and weather conditions, and other environmental conditions. The temporal fusion of the internal and external factors represents a driving context [7]. A growing body of evidence suggests that different contextual settings have varying impacts on the driver states (e.g., emotions, attention, and cognition) and behaviors (e.g., speed patterns and hard brakes) [7]. For instance, road environment, weather condition, traffic density, driver's activities, and even the background music have shown to affect driver's state, and driving behaviors [7], [9]–[13]. As a result, to develop personalized profiles for each specific driver, we need to accurately monitor the internal and external activities and identify how changes within the environment may impact driver states and behaviors.

Over the past 15 years, the research community has identified that there is a need for collecting multimodal driving data through naturalistic studies. Examples include the ''The 100-Car Naturalistic Study,'' conducted by Virginia Tech Transportation Institute (VTTI) [14], the ''European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment (UDRIVE)'' [15], and the MIT Advanced Vehicle Technology Study [16]. Such studies have provided significant insights into how different in-cabin and outdoor conditions may impact driving behaviors and states. For instance, [17] emphasized that internal factors such as driver's distraction, aggression, emotions, and secondary tasks play an important role in accident prevalence. Similarly, previous studies have pointed out that external factors such as weather conditions [18], and road geometry and design [19] impact driver's state and behavior. However, most of these Naturalistic Driving Studies (NDS) rely on features collected from video cameras capturing in-cabin and outdoor conditions. Although video streams are extremely informative, they mostly provide insights about external factors. In fact, research suggests many internal factors and states (e.g., driver's cognition) cannot be accurately detected through using only cameras [20]. For instance, a driver might smile when being frustrated, leading to a misleading inference about the driver's state [21]. To the best of our knowledge, none of the existing longitudinal NDS have fused the features extracted from the driver's physiological measures with those extracted from the video streams.

Among the internal factors, previous studies have noted that collecting physiological data in longitudinal naturalistic driving studies was not feasible due to the lack of technological advancement that exists today (e.g., wearable devices that can provide physiological and behavioral states of a person) [22]. As a result, the majority of the past driving studies that included physiological sensing were conducted in short-term controlled experimental studies (e.g., [23]). Not having access to datasets that provide internal factors together with environmental attributes has caused many driver state detection models to rely on data that were either not from real-world studies or were not tested in real-time. For instance, [24] developed a dynamic Bayesian model to contextualize the driving behavior based on different environmental, vehicular, and driver-specific conditions. In their model, they used different attributes of the environment such as noise and temperature, with driver's psychophysiological measures such as eyelid movements, and behavioral attributes such as lane maintenance to detect different states of fatigue, drunk, reckless, and normal conditions. The term psychophysiological refers to psychological states such as emotional responses (e.g., anger, frustration, and happiness), cognitive load, and distraction that can be measured through changes in human physiology responses (e.g., heart rate, skin temperature, and skin conductance) [25]. However, they highlighted the data required for validating their model does not currently exist, and they had to rely on previous literature for retrieving probability conditions of different driver states under various environmental conditions [24].

Over the past few years, the advancements in the field of ubiquitous computing have accelerated very quickly. Currently, over 900 million wearable devices are being used worldwide on a daily basis [26]. The application of these devices spans over a variety of fields such as mental health monitoring and interventions [27], physical health and activity monitoring and training [28], [29], sleep monitoring and intervention [30], and insurance and policy purposes [31]. Additionally, recently wearable devices are also being utilized in driver state recognition research area. Although these studies were mostly conducted in controlled settings, they provided insight into the application of wearable devices in driving research. For instance, [32] have used Microsoft armbands to detect driver's drowsiness in a virtual driving environment. Reference [33] have used wearable devices for detecting driver's fatigue, stress, and abnormal conditions in a driving simulator environment. [34] have found that conventional wearable devices in a driving simulator environment can be used for driver's drowsiness detection. Another study has used wearable devices for detecting heart rate (HR) changes in different road and weather conditions in naturalistic settings and found out significant changes in HR among different conditions of the city versus highway and rainy versus clear weather [35]. These findings, although

being in controlled environments, provides evidence on the effectiveness of using such devices in driving environments for detecting driver's state.

These advancements have not only been in the areas of physiological sensing and wearable devices. Over the past decade, there have been significant improvements in computer vision and machine learning approaches, where we can now accurately detect specific features and behaviors of drivers from the in-cabin videos while detecting objects [36] and outside conditions through the outdoor videos [37]. However, since the majority of existing NDS were introduced over a decade ago, many of the existing datasets do not include these modalities of data such as driver's pose features, gaze patterns, and objects in the environment. As a result of these improvements, we can now utilize (1) wearable devices to monitor driver's states and internal changes and (2) advanced computer vision and machine learning algorithms to collect the external factors.

In this paper, we introduce HARMONY: a human-centered multimodal naturalistic driving study framework, along with its underlying principles, data collection, and processing procedures. HARMONY includes data collected from the roadway and in-cabin videos, Global Positioning System (GPS), driver's Heart Rate (HR) data as well as raw photoplethysmogram (PPG) readings, hand acceleration and movement, the in-cabin ambient conditions such as light level, noise levels, and the music being played while driving. This information is collected longitudinally from naturalistic driving scenarios. In the following sections, we first review the previous work in naturalistic driving studies and identify the existing gaps in research. We then outline the aims of HARMONY and discuss our methodology for data collection and processing. To further demonstrate the applicability of our framework, we present a case study in which we demonstrate the benefits of using longitudinal physiological signals to capture drivers' behavioral variability. Lastly, we discuss the current limitations and the future road map of HARMONY.

II. BACKGROUND

In this section, we first review the different categories of previous driver-in-the-loop studies; specifically, section *II-A* provides an overview of conducted studies through driving simulators, section *II-B* focuses on on-road controlled studies, and section *II-C* evaluates the existing Field Operational Tests (FOT) and NDS. In section *II-D*, we provide a detailed overview of the pros and cons of these studies and identify some existing gaps in current NDS studies addressed by the proposed HARMONY framework.

A. DRIVING SIMULATOR

Studies performed in driving simulators are beneficial in understanding causal analysis as different factors can be controlled [38]. Previous studies have coupled driving simulators with physiological sensors to detect driver states in different contextual settings such as detecting states of being awake versus sleeping in four-hour driving epochs [39], cognitive distraction when the lead vehicle abruptly breaks [40], response to automation takeover when the take over request is offered through different sources [41], and driver's performance and possible cognitive distraction when being exposed to different traffic signs [42]. Driving simulators are safe tools for conducting driving experiments in different environmental conditions, such as crash events, or evaluating drivers' emotional states [43]. Additionally, by using driving simulators, it is viable to collect modalities of data that are not feasible in a real-world setting such as brain activity signals (e.g., EEG) [41]. However, due to its controlled nature, driving simulators cannot be used to evaluate longitudinal behavioral changes. Thus, driving simulator studies cannot be used to capture real-world changes in driver behaviors and states given various contextual settings. Furthermore, a recent review has depicted that one-third of the review corpus that used driving simulators as a means of capturing driving behaviors in different conditions have achieved no validity in reproducing in real-world conditions [44].

B. ON-ROAD CONTROLLED STUDIES AND FIELD OPERATIONAL TESTS (FOT)

In contrast to simulator studies, on-road controlled studies, FOTs, and NDSs utilize participants driving real vehicles in naturalistic settings. The major difference is that in an on-road controlled study, the experimenter has a higher level of control on specific variables of interest [45]. For example, in an on-road controlled study, participants are asked to drive on the same route as other drivers while driving in the same experimentally equipped vehicle (in contrast to driving their own personal vehicle). The on-road studies are typically conducted for a short period of time (up to a few hours) [22]. Furthermore, these studies are typically accompanied by an observer/experimenter, which may impact the driver's behavior as he/she feels they are being observed and monitored [46].

A number of studies have evaluated the impact of different road-way conditions on driver state and behaviors through on-road controlled studies. For instance, [47], [48] used physiological responses to assess moderate levels of mental load. In this study, the authors have used the same vehicle, and participants drove through the same route. The authors suggest physiological variability can be used for distinguishing different levels of cognitive load, such as differences in driving in a city as compared to a highway environment. In another study, [11] collected EEG data from six participants while driving in a sensor-equipped Ford Escape 2015. Through their proposed framework, the authors were able to detect secondary task engagement while driving with 99% accuracy. The tasks included phone conversation, texting, answering questions, spelling, and listening to music. The participants were driving around the same time frame (2-5 pm), and an observer was present in the backseat for labeling the data as the participant was driving the vehicle. Reference [48] used cameras and chestbands for monitoring physiological measures such as ECG and respiration waves to detect important

driving events automatically. In this study, participants drove a regular sedan (Acura TLX) through a pre-defined route. In their study, authors have used an unsupervised learning method to cluster different physiological signals into categories of normal, event, and noise with a high recall rate of 75%. The detected events in the physiological signals were associated with certain driver state and behavior events of interest, such as frequent lane switching, last-minute maneuvers, being angry, frustrated, and excited. In a similar study, to monitor how driver's emotions might vary in a driving condition, [23] monitored 34 participants for 50 minutes through capturing physiological factors through the Empatica E4 wearable device, cameras to capture both in-cabin and outside conditions. Through this study, the authors discovered human-vehicle interaction (i.e., navigation, changing radio settings, cruise control) could cause the highest number of negative emotions, among other reasons.

Overall, since on-road controlled studies are conducted with real vehicles and roadways, they provide a more realistic condition compared to driving simulators. However, these studies are still conducted in controlled environments (e.g., specific road type), usually include an experimenter/observer, which may impact participant's behaviors, and lack the longitudinal aspect of naturalistic studies, where driver's behaviors can be monitored across similar and different contextual settings.

A more generalized type of on-road controlled study is FOT. These studies are conducted to assess one specific factor or function, such as a new driving assisting system or a specific intervention in real-world driving scenarios longitudinally [22], [49]. These studies are generally closer to the real-world unconstrained driving situation and have higher external validity compared to on-road controlled study [22]. However, as they are focused on a specific factor of interest, they still include some level of constraints. For instance, they might use some pre-equipped vehicles, which can lead to different behaviors by the participant. Examples of FOT can be found in [50]–[52].

C. NATURALISTIC DRIVING STUDIES

NDS are conducted in real-world conditions, and they intend to capture how contextual factors may impact driver's behaviors and states [14], [15], [22], [53]. NDS has high external validity as it is performed in real-life scenarios [22]. In these studies, the participants' vehicles are instrumented with different sensing and monitoring technologies to collect both internal and external factors while driving in naturalistic conditions. These devices may include cameras, Onboard Diagnostic (OBD) readers, GPS units, and other data acquisition systems that can collect information from both the in-cabin and outside environment [15]. In contract to on-road controlled studies, there are no observers/experimenters in NDS, and participants are asked to perform their daily routine activities [15], [16], [53]. Additionally, NDS is always accompanied by uncontrollable real-life noise [22]. As a result, to capture accurate information about participants' driving behaviors and account for real-world noise, NDS must be conducted longitudinally. Compared to driving simulator and on-road controlled studies, NDS is time consuming, resource extensive, and costly. However, these studies provide a holistic understanding of how internal and external factors influence driver's behaviors and responses in different contextual settings.

Although there is a limited number of NDS to date, they have been conducted across different countries (e.g., United States, Europe, Canada, China, Japan, and Australia). The first large scale NDS is the "The 100-Car Naturalistic Study," conducted by the Virginia Tech Transportation Institute (VTTI) in 2005. In this study 241 primary and secondary drivers were monitored for 12-13 months. 100-car includes 2,000,000 vehicle miles and approximately 43,000 hours of data. The primary goal was to provide information with regards to crash and pre-crash data from both the environment and vehicle sensors. The data acquisition system in this study included five channels of digital video, longitudinal and lateral kinematic information, lane-keeping measure, a GPS unit, and a headway detection system for providing information on leading or following vehicles [14]. The majority of the participants for this study were recruited from the northern Virginia and Washington DC area. This dataset has provided significant insights into changes in driving behaviors and states such as variations in driver's attention, and drowsiness [9], overtaking maneuvers [54], and even differences in driving behaviors among different age and sex groups [55].

The next large-scale naturalistic study that was conducted in the United States is the Second Strategic Highway Research Program (SHRP 2), conducted across Indiana, Pennsylvania, Florida, New York, North Carolina, and Washington, starting from 2006 and ending in 2015. The goal of this study was to understand the driver's performance and behavior in traffic safety (e.g., road departure, offset left-turn lanes, driver inattention, and rear-end collisions on congested freeways). The data included in-cabin and outside video streams, eye forward tracking, passive alcohol sensor, lane detection, vehicle accelerometer and gyroscope measures, GPS, forward radar, light level sensor, and infrared illumination. This dataset includes information from more than 3,400 drivers, with 5,400,000 trips spanning over 80 million km (an average of around 1,600 trips per participant) [56], [57]. The SHRP 2 study has provided a set of rich dataset to the research community, including multiple crash events that were not available to this extent before, insights on driver's behavior, and performance analysis in different contextual settings and across different age groups. For instance, [58] provided an overview of the impact of different weather conditions on driver's lane-keeping ability. Through analyzing the SHRP 2 dataset, they concluded heavy rain could significantly increase the standard deviation of lane position. Another study has analyzed the relationship between the crash and near-crash events with the driver's glance patterns based on the SHRP 2 data [59]. By performing a prevalence analysis on the glance regions, the authors

identified factors such as driver's eyes positioning (e.g., on or off the road) prior to the crash event as well as the driver's uncertainty of a driving situation (e.g., arriving to an intersection) to be significant predictors of the crash and near-crash events [59].

A similar study was conducted in 2012 in Europe, titled the European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment (UDRIVE). The goal of this study was to analyze driver's behavior with a focus on both improving safety and identifying new approaches to make a more sustainable road transportation system. The research questions span across different roadway studies, including driver risky behaviors and causes of accidents, day-to-day driving behaviors, causes of roadway distraction and inattention, interactions with vulnerable road users, and eco-driving [60]. This study collected 87,871 hours of data from 48 trucks (41,389 hours), 186 vehicles (45,591 hours), and 47 powered two-wheelers (891 hours). The data collection has been conducted in six different countries across the United Kingdom, Netherlands, Spain, Poland, France, and Germany. It includes features extracted from video streams, CAN interface data collector, GPS, accelerometer, and acceleration/speed sensors [15]. UDRIVE helped researchers define a driving style indicator based on secondary task engagement (e.g., phone usage), as well as analyzing regional differences among different drivers. [61], [62].

Following a similar framework and setup as the SHRP 2 study, other NDS studies have been conducted across China, Canada, and Japan. The ''Chinese Naturalistic Driving Study'' [63] was conducted in Shanghai in 2012-2015. This study collected data from 60 drivers over three years. The study collected 161,055 km of driving data. This study was conducted by the Tongji University, General Motors, and the Virginia Tech Transportation Institute (VTTI). This study has provided insights into the naturalistic driving behavior of Chinese drivers. For instance, through this dataset, [64] discovered intersection types can change the driver's scanning behavior. In another study, five existing car-following models were calibrated and validated using this dataset [63]. Other studies have also evaluated more specific behaviors, such as cut-in behavior [65]. Two similar studies were conducted in Canada and Japan. The NDS in Canada took place in Saskatoon and Saskatchewan in 2013 and concluded in 2015 [66]. In this study, researchers collected over 2,000,000 trips from 149 drivers driving different categories of vehicles. Specifically, this study includes a ''Truck Study'' that monitors 30 trucks in the western Canada region (i.e., Saskatchewan, Alberta, British Columbia). Similar to SHRP 2 study, the primary research question in this study was to identify how different user-related and/or environmental factors may impact the rate of crash and near-crash events in Saskatchewan. The major difference between this study and the SHRP 2 is the change in the camera views and positioning [66]. Additionally, this study mainly focused on trucks and heavy vehicles compared to SHRP 2 were larger pool of vehicles and drivers were recruited. A study conducted by Japan's Automobile Research Institute in 2006 [67] focused on evaluating driving behaviors by observing 60 participants driving 35, cc-class wagons and 25 small sedans. The study collected five modalities of data, including camera, GPS, audio, kinematic sensors, and OBD vehicle parameters. Similar to previous studies, one of the main goals of this project was to detect factors contributing to crashes.

Another NDS conducted in Canada is titled as ''Candrive'' and only focused on elderly drivers [68]–[70]. This study collected GPS and vehicle data from 928 participants of 70 years or older for up to four years. This study analyzed multiple research questions specifically for elderly drivers, such as the impact of medical conditions on driving behaviors and to further define a clinical criteria for unsafe drivers based on their health issues. For instance, based on this study, [71] provided a framework on assessing driving ability among elderly using factors such as driving characteristics (e.g., type of road), driver's actions (e.g., lane changes), and driving conditions (e.g., time of day). This study was further extended to New Zealand and Australia, in which the researchers collected data from 300 elder drivers in Melbourne, Australia and Wellington, New Zealand [70]. The Australian Naturalistic Driving Study's (ANDS) goal was to monitor 360 drivers (180 from New South Wales and 180 from Victoria) to similarly identify reasons behind crashes and near-crash events [72]. It includes cameras, GPS, vehicle dynamics (e.g., speed, brake, and turn signal sensors), and machine vision sensors' data. This study was different from the previous ones in a few ways: (1) the participants in this study were experienced drivers, excluding the novice drivers; (2) the study duration was four months for each participant; and (3) the study utilized newer driving assisting systems technologies such as such as Mobileye and Seeing Machines to detect distraction and drowsiness events [72]. Using ANDS, [73] assessed the patterns of secondary task engagement during everyday driving scenarios, and [74] developed a visualization platform to presents how different modalities simultaneously vary over time.

The most recent NDS is the MIT Advanced Vehicle Technology Study conducted in 2018 [16]. This study was conducted with the aim of (1) collecting large-scale real-world driving data with high definition videos to build and train deep learning based in-cabin and outside perception systems and (2) enhancing the understanding of human-automation interaction. In contrast to other NDS, this study only includes semi-autonomous vehicles (mostly TESLA), and is conducted in the Boston metropolitan area in the United States. This includes data collected from the Inertial Measurement Unit (IMU) sensors, GPS, CAN, and cameras. To date, this study has collected more than 15,610 days of data, with over 511,638 miles of driving semi-autonomous vehicles from 122 participants (according to the most recent publication on this work [16]). This study introduces a framework aimed at enhancing semi-autonomous vehicle safety and reliability by building better shared-automated systems [75]. A summary of the previous studies can be viewed in Table 1.

TABLE 1. This table provides an over the location, number of participants, modalities and quantities of collected data in the on-going and previous naturalistic driving studies conducted across the world.

Study Characteristics					Vehicle Sensing							Environmental Sensing						Human Sensing										
Study Title	Location	Number of Participants	Amount of Data	Duration	GPS	Speed	Compass	CAN	Acc	Gyroscope	Magnetometer	Radar	Outside Camera	Temperature	Noise	Light	Mobileye	Heart Rate	Linear Hand ACC	Hand ACC	Hand Gyroscope	Music Preferences	PPG	Computer Vision Gaze	Computer Vision Pose	Eye Forward Monitor	Alcohol Sensor	Seeing Machines
100-car study	Northern VA, Washington DC, United States	241	2,000,000 vehicle miles, 43,000 hours of data, 12 - 13 months data collection period per vehicle	2005-2006	✓	✓	-	✓	✓	-	-	✓	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SHRP-2	Eastern region of the US	3400	5,400,000, 49.5 million miles	2006-2015	✓	✓	-	✓	✓	✓	-	✓	✓	-	-	✓	-	-	-	-	-	-	-	-	-	✓	✓	-
Japanese Driving Study	Japan	60	-	2006-2008	✓	-	-	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
UDRIVE	United kingdom, Spain, Netherlands, Poland, France, and Germany	48 trucks, 186 vehicles, and 47 powered two wheelers	87871 total hours of collected data, including 41389 hours of truck, 45591 hours of vehicle, and 891 hours of powered two wheelers	2012-2017	✓	✓	-	✓	✓	-	-	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-
Chinese NDS	China	60	total mileage of 161,055 km	2012-2015	✓	✓	-	✓	✓	✓	-	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-	-		-	-
Canadian NDS	Canada	140	Over 2,000,000 trips	2013-2015	✓	✓	-	✓	✓	✓	-	✓	✓	-	-	✓	-	-	-	-	-	-	-	-	-	✓	✓	-
Candrive	Canada, New Zealand, and Australia	928 from Canada, and 300 from Australia and New Zealand	Up to seven years of data collection	2009-2015	✓	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ANDS	Australia	360	1,512,630 km	2015 - Present	✓	✓	-	✓	✓	✓	-	✓	✓	✓	✓	-	✓	-	-	-	-	-	-	-	-		✓	✓
MIT-AVT	Greater Boston area - United States	122	511,638 miles	2015 - Present	✓	✓	-	✓	✓	✓	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HARMONY	Virginia	21	At least 1 month of driving data per participant	2019 - Present	✓	✓	-	-	-	-	-	-	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-

D. EXISTING GAPS

Reviewing the existing NDS suggests, these studies have provided a holistic overview of how different internal and external factors vary over time. As a result, the research community has significantly benefited from these studies, and different researchers across the world have been able to further evaluate specific drivers' behaviors, responses, and states in specific contextual settings. Due to their significant impact and importance, which is because of their high internal validity, the number of NDS has increased over the past two decades. The majority of NDS focus on evaluating the effect of outdoor environments and external factors on driving behaviors, and there is a limited number of studies that have collected and analyzed physiological features and driver's internal factors. Additionally, in the existing NDS, driver sensing is solely conducted through features and behaviors extracted through in-cabin videos. To better grasp the need for using multimodal driver sensing, in this section we discuss the limitations of existing studies in solely relying on in-cabin video streams for monitoring driver's states and behaviors and highlight how new advancements in wearable technologies and ubiquitous computing can address these limitations.

The most important drawback of video streams are their inability to provide the underlying complicated state of the driver. Driving is a dynamic task that can be highly affected by the driver's psychophysiological state, such as the driver's underlying emotional and cognitive situation. Previous studies suggest a driver's performance can be significantly impacted by his/her emotional states; for instance, strong emotions such as positive valence leads to better takeover readiness for drivers [76]. Recent studies in the applications of affective computing in driving, and psychology and emotion sciences provide evidence that we cannot only rely on facial features to detect a person's emotional states [21], [77]. For instance, in driving research, [21] demonstrates the limitation of video features in inferring driver's state. This study attempted to classify driver frustration using videos, audio, and a combination of both. As our frustration might not necessarily appear in our facial expressions, a combination of audio and video resulted in a higher accuracy in detecting frustration. An example of such situation is when our frustration shows up as a smile, which can be mistakenly thought of as a joyful event if we only relied on video data. In addition to driving research, recent studies in psychology suggest that

analyzing emotions should not be taken out of context [77]. In other words, the context defines the emotional state that we are in and might be the reason why we feel sad, happy, or angry given certain environmental conditions [77], [78]. As a result, to achieve a trustworthy and acceptable level of shared-autonomy, future AV needs to be able to detect and validate driver's states through features extracted from different sensing modalities.

Features extracted from video streams have also been utilized to detect and classify driver's cognitive load [79], possible distraction [80], and drowsiness [81]. For instance, [79] have used features extracted from the driver's eyes to monitor driver's cognitive load in a typical n-back task while driving. In their study, the authors have achieved 88.1% accuracy in detecting low, medium, and high cognitive load from 92 participants driving semi-autonomous vehicles in the wild. Additionally, by using videos, multiple studies have attempted to detect driver's secondary tasks such as interacting with the center-dash system (e.g., to change radio station), or speaking on the phone, which might also be indicative of certain levels of possible distraction [80], [82]. Although we can extract and classify driver behaviors and states through video features, we still struggle in classifying the situation when visual cues are not indicative of driver's states. Examples include situations that these visual cues vary among participants for the same state (e.g., people vary in expressing their frustration [21]) or the driver might have hidden psychological states (e.g., under cognitive load from a prior task such as a phone call before entering the vehicle). Such cases require other modalities of data to capture the driver's cognitive load. For instance, a recent review on cognitive load estimation in driving provided eight different measures as candidates for assessing a driver's cognitive load. These measures include electroencephalography and event-related potentials, optical imaging, HR and HR variability, blood pressure, skin conductance, electromyography, thermal imaging, and pupillometry [20]. Although a recent study has attempted to estimate some of these features from video streams (e.g., estimating HR from video [83]), most of them are currently measured through physiological sensors that are not accessible while solely relying on videos. Thus, to estimate the cognitive load of the driver we need to explore beyond features extracted from video streams and identify how other devices can be complementary in measuring the driver and passenger(s) psychophysiological states during different driving events. These examples collectively suggested that videos can be insufficient, and even to some extent misleading if they are the only source that we are relying upon for driver's state recognition.

Studies also quantified driving behaviors such as lane-keeping ability or breaking patterns as a means to detect the driver's physiological state. Behavioral techniques to detect the hidden states of the driver, may not always provide a full understanding of the driver's state. For instance, [84] indicated drivers' cardiac measures can be affected by secondary tasks (e.g., n-back tasks) while leaving driving behavioral metrics such as lane-keeping unchanged.

Moreover, it should be noted that videos are both computationally expensive and privacy-intrusive. Relying upon videos for driver's state recognition often requires high amount of video data in which many attributes of driver's personal behaviors are exposed to the research team or the automaker manufacturers. People often do not feel comfortable being monitored for an extensive amount of time via video cameras. Furthermore, to perform real-time processing of video streams is computationally expensive and requires in-vehicle GPU units. However, physiological measures retrieved through wearable devices have proven to be very helpful in detecting human's state and behaviors [85], provide feedback and interventions [86] while being less privacy intrusive and having a much lower computational cost. For instance, recent human factor research studies have shown that hand acceleration data can be used to detect multiple human activities such as walking, jogging, and biking with a high level of accuracy [87]. Another study has found that physiological measures recorded in a driving simulator were indicative of a driver's psychophysiological state in automated and semi-automated driving mode [88]. Such applications can be extended to in-cabin real-world activities, which can then be helpful in detecting task engagements (e.g., phone usage, eating), safety-related issues (e.g., distraction, drowsiness) [89], [90], and emotions and cognitive load [76], [88].

Lastly, most of the previous major NDS did not collect detailed ambient in-cabin external factors data such as (1) noise level, (2) light level, and (3) the music that is played in-cabin. Previous studies through controlled experimental setups have shown the effect of these in-cabin external factors on driver's state and behaviors [91]–[94]. For instance, a recent review suggests traffic noise can be a cause for a psychological disorder, and mental stress [94]. Another study has analyzed the effect of three different conditions of ambient light (i.e., no light, blue, and orange light) on the driver's behavior. This study found out that the presence of ambient light can enhance driving performance (i.e., lane-keeping) [93].

Additionally, research suggests music can have a significant impact on a person's behaviors, emotions, and physiological states [91]. Previous research suggests listening to music can cause distraction, which increases the risks of accidents, especially in high demanding scenarios, such as unexpected red rear brake lights and complex peripheral signals [10], [12]. On the other hand, music can also be used to mitigate the impact of emotional states on the driver's risk acceptance and safety considerations while driving [91]. Additionally, more recent studies have shown various music features have a different impact on people's behaviors, although no consensus has been reached so far. Furthermore, the findings from no-driving relate studies can contradict with driving studies. For example, in a no-driving study, the high tempo is found to significantly increase HR while listening to the music as compared with silence [95]. However, in a car-following study, the physiological measurement did not differ

in conditions with and without music, even with different volume levels [96]. Other music features, like the music genre, were also found to affect driving performance [97]. Similarly, [98]–[100] mentioned that the mitigation effect of certain music types differs in human factors (age, gender, emotional states, preferences, and familiarity) and environmental factors (the time of the day, location, the current tasks, presence of other passengers). While existing studies provide significant insight into the effect of such in-cabin external factors on driver's state and behaviors, almost all of them rely on experimental controlled setups (e.g., playing a pre-defined music list by the researcher) and have not been performed in fully naturalistic driving environments.

III. HARMONY GOALS

As outlined in Table 1, the HARMONY framework addresses the identified gaps by integrating physiological sensing and machine vision to contextualize driving experiences. To create the HARMONY framework, wearable devices, cameras, and ambient sensors are utilized to track and monitor different external and internal factors. The data extracted from these devices are used to contextualize different driving events and their corresponding user or group-specific behaviors and responses. The specific goals of HARMONY are to (1) introduce an NDS framework to collect and monitor driver's psychophysiological states in addition to the changes in in-cabin and outdoor environments and (2) analyze the changes in driver states and behaviors and train machine learning models to automatically detect and classify different driving behaviors, states, and events.

A. GOAL 1: AN NDS FRAMEWORK FOR COLLECTING AND ANALYZING DRIVER's PSYCHOPHYSIOLOGICAL STATE TOGETHER WITH THE ENVIRONMENT

The first goal of HARMONY is to provide insights into driver's states and behaviors by introducing a framework that is naturalistic, longitudinal, and scalable. To achieve this, HARMONY uses off-the-shelf sensing devices for data collection and integrates state-of-the-art computer vision and machine learning algorithms to extract detailed contextual information from cameras, wearable devices, and ambient sensors. Specifically, HARMONY is designed to longitudinally monitor and collect the driver's physiology, movements, location, along with the ambient conditions in the vehicle (i.e., light and noise levels, and in-cabin music features). The proposed NDS framework then utilizes ''virtual sensors'' to extract different contextual features from the cameras and wearable devices. Additionally, HARMONY monitors the ambient noise levels as well as the music being played while driving as the main characteristic of the in-cabin environment.

B. GOAL 2: ANALYZE THE CHANGES IN DRIVER STATES AND BEHAVIORS AND BUILD PREDICTIVE DRIVER STATE MODELS

Previous research defines context as any relevant information that can be used to characterize a situation of an entity [101], [102]. This can include a place, one or a series of events, as well as the user of the application [101]. Thus, the second goal of HARMONY is to provide detailed information for defining driving situations by moving beyond only collecting the visually available information (e.g., outside or in-cabin conditions). Through fusing physiological, environmental, and behavioral factors, HARMONY aims to create predictive models to identify driver's states such as cognitive load, attentiveness, and task-engagement with higher reliability and to a greater detail compared to existing methods. Wearable devices and ubiquitous computing have been used in multiple disciplines due to their strong potential in providing rich information about the user while being less intrusive and computationally expensive. By integrating multiple sensors in one device, wearables can decode human physiological states as well as provide information on human activities and environmental events. In driving applications, in addition to revealing human's underlying state, such devices can be coupled with other modalities of data to better enhance predictive models generated by HARMONY for driver activity and engagement recognition. This is an important advancement to be included in current and future AV for retrieving the driver's state in real-time so that vehicles can rely on the driver's attention in the events of failure. In this way, HARMONY paves the way for future safe autonomous vehicles that are reliable and do not invade users' privacy.

IV. THE PROPOSED FRAMEWORK

To achieve the identified goals, as shown in Fig. 1 the HARMONY framework provides a holistic view of driving events by integrating driver's psychophysiogical factors (internal factors) with the changes in outside and in-cabin conditions (external factors). In the following subsections, we first provide an overview of the sensors and algorithms utilized in HARMONY (section IV-A). Then we provide an overview of our data processing and fusion techniques, which are implemented for retrieving events of interest with respect to the environment and the driver from the collected sensor readings (section IV-B). Lastly, we discuss the details on participant recruitment and our on-going data collection efforts (section V).

A. DATA COLLECTION

In HARMONY, data is collected through three sources: (1) driver, (2) vehicle, and (3) ambient sensors through a network of physical and virtual sensors. Physical sensors refer to commercially available devices that collect raw information such as video streams and participants' biometrics. Virtual sensors refer to feature extraction and event detection algorithms as well as APIs that provide a deeper layer of information about the contextual settings. Subsection IV-A1 and IV-A2 provide more details about the utilized physical and virtual sensors, respectively.

1) PHYSICAL SENSORS

HARMONY only collects data from two physical sensors: a dash camera and a wearable device (smartwatch). By extracting information from different sensors that are integrated

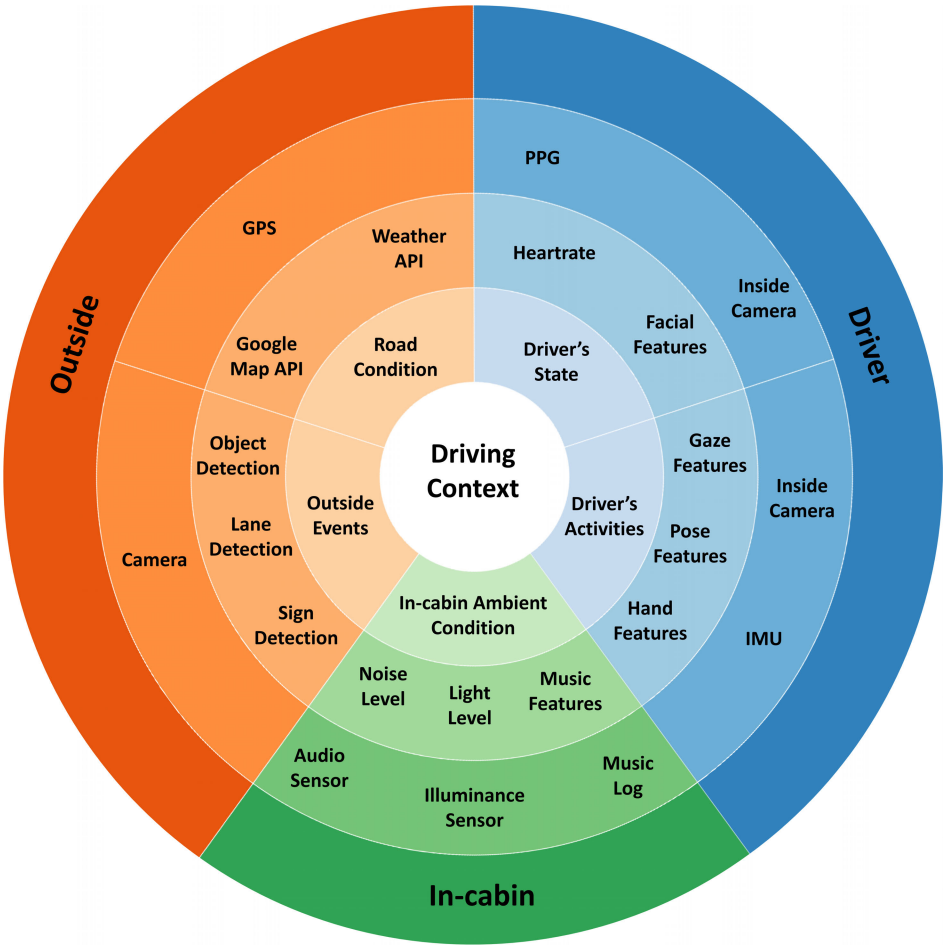

FIGURE 1. Data collection framework. Using this framework we can better contextualize driving scenarios and provide multimodal data while being low-cost.

into each of these devices, we collect raw video streams of in-cabin and outside environments as well as biometric information from the drivers.

a: CAMERA

A dual-dash camera is used to collect both inside and outside environment information simultaneously, with relatively high storage that can be used for longitudinal data collection. After testing a number of commercially available dash cameras, for the first round of data collection, we utilized the BlackVue DR750S-2CH dash camera. This camera specifically includes: (1) up to 256 GB with an SD card memory (approximately 25 hours of driving), (2) GPS device to track vehicle's location, and (3) synced with the global time retrieved through GPS, allowing the camera always provide the correct current timing (this feature is critical to synchronize the timestamps of events between the physical devices). The camera does not have an LCD, which decreases the chances of distraction by the LCD for participants. This also may reduce participants' sense of being monitored by observers/researchers. Moreover, this camera has the option of disabling the audio recording, which is required as per the Institutional Review Board (IRB) approval for this study. In addition, the utilized dual-dash camera provides the speed of

FIGURE 2. Sample view of the in-cabin camera (A), BlackVUe dash camera (B), sample view of the outside cabin dash camera (C).

the vehicle. The recorded videos are stored at 30 frames per second (fps), Full HD resolution in 3-minutes segments. Each 3-minutes segment of driving is saved as a joint video of the inside and outside environments. A view of the camera can be seen in Fig. 2.

b: SMARTWATCH

To collect data with frequency and properties of interest, HARMONY uses an android smartwatch that is equipped with the "SWear" app (Fig. 3 - A), an in-house app designed for collecting long-term sensor data from smartwatches [103]. SWear is available on Android store [104] and is designed to smooth the process of data collection on smartwatches by adding the ability to control each sensor's data collection frequency to the desired sensing regime. For HARMONY 1.0, SWear records HR [1 HZ],

FIGURE 3. Sample view of the android smartwatch (A), homepage of the app (B), status page of the app (C), setting page of the app (D).

hand acceleration [10 HZ], audio amplitude (noise level) [1/60 HZ], light intensity [1/60 HZ], location [1/600 HZ], gravity [10 HZ], Compass [1 HZ], Altitude [1 HZ], Magnetometer [10 HZ] and gyroscope [10 Hz] data. The user interface of the app on the device can be seen in Fig. 3. Participants are required to start/stop the data collection for every session of driving (3 - B). The smartwatch saves every segment of driving data locally and records the participant's ID, and the number of saved files on the app's status tab (Fig. 3 - C). Every participant is required to sync the watch with their own personal phones. Every two weeks, the participant were requested to transfer their data to our system by one-click on the upload icon on the settings tab (3 - D) to transfer data through wifi to a secure Amazon Web Service server for further storage and analysis. The watch syncs it's time periodically from the companion mobile phone, it always provides the current synced global time. Many new features have been recently added to SWear to further facilitate data collection such as automatic data sync to the cloud and adaptive sensing by automatically enabling data collection when a driving activity is detected.

2) VIRTUAL SENSORS

In addition to the two devices, multiple virtual sensors are utilized to retrieve information about the driving situation by either using the outputs of physical sensors or cloud information. Virtual sensors in our system are APIs, computer vision algorithms, as well as in-house developed event-detection algorithms. APIs include music platform API, Google Maps API, and Weather API. HARMONY collects the log of the music that the participant is listening through either the music application API used by the participant while driving (e.g., Spotify, Pandora, YouTube) or through connecting all the music applications to a unifying platform such as Last.FM account, where the log of participant's music can be retrieved. This log includes the music title, duration, lyrics, and the time that was played. After one time setup at the beginning of the project, the music data collection does not require further intervention in subsequent experiments. Additionally, the timestamp of all of the music is based on the same global time as the other devices.

Additionally, HARMONY collects the weather and road data by using two APIs of Google Map [105], and OpenWeatherMap [106]. These two APIs provide speed limits, user's route, as well as weather conditions for a set of GPS data retrieved through the smart wearable. Moreover, using multiple computer vision algorithms, detailed representations of both in-cabin and outside environment are retrieved. These algorithms help retrieve the driver's gaze, pose, facial emotions, as well as objects in the field of view (i.e., both in-cabin and outside videos). The computer vision algorithms performed on the dataset are detailed out later in section IV-B.

B. DATA PROCESSING

We first retrieve the information of every file that is recorded in our system. This information includes the file properties of the videos (e.g., duration, start time, end time), information retrieved from in-cabin and outdoor video streams (e.g., number of passengers), and the associated (time-stamped) physiological data files (if available). The data processing takes place through multiple scripts coded in Python, which will be detailed out in the subsections below.

1) VIDEO INFORMATION

Video information is retrieved automatically through FFMPEG video packages in Python [107]. By using the file information on each video, we retrieve the name of the video, participant ID, duration of each video, start time, and end time. Also, using the sunset and sunrise time, we assess if a video is recorded in day or night. This is important as many of the computer vision applications are not feasible to retrieve data on dark/night-time videos. This information is then used to produce the time frames epochs that need to be created from modalities of HARMONY (i.e., smart wearable and APIs).

2) GAZE, POSE, AND FACIAL FEATURES

Each in-cabin video is analyzed using three software: OpenFace [108], OpenPose [109], and Affectiva [110]. OpenFace analyzes faces in the in-cabin videos and outputs facial landmarks, head pose, gaze direction in 3D, and gaze angles in horizontal and vertical directions. OpenPose detects skeleton joints in each video. Finally, using the Affectiva module on iMotion software, multiple emotional metrics are retrieved, including 2D and 1D emotional measures. The 2D emotional measure includes valence and engagement. Valence is the extent that a person's facial emotion is negative or positive, ranging from -100 to $+100$. Engagement is the extent that a person reveals their specific emotion using their facial muscles ranging from 0 to 100 with not showing emotion as 0. Additionally, six basic categorical emotions (1D space) are also retrieved using Affectiva. These include sadness, anger, happiness, contempt, surprise, and fear. These results are saved into a CSV file associated with each video epoch.

3) SMARTWATCH EPOCHS

As recording the smartwatch data requires the participant to start and stop the watch for every scenario, sometimes the participant either forgets to record or does not have the watch in the car. Thus for every driving video, there is a need for an assessment to determine whether a physiological file exists

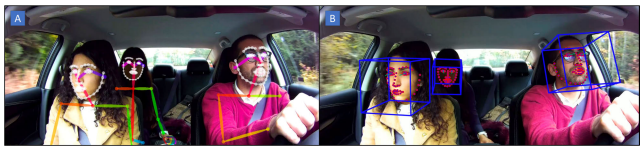

FIGURE 4. OpenPose sample output providing driver's skeleton key point (A), OpenFace sample output providing driver's facial and gaze measures in the wild (B).

or not. By using the retrieved time frames from the videos, the same epochs of smartwatch data (e.g., HR, and hand acceleration) are extracted and saved to a CSV file for each specific participant.

4) MUSIC FEATURES

The next step is to use the collected music log as described in section IV, to retrieve different music/song features being played while driving for each participant. By feeding the music log of the participant to the Spotify API, the audio features, including energy, danceability, instrumentalness, acousticness, liveness, loudness, valence, tempo, and lyrics are extracted if the music is vocal (podcast and talk shows are not included). When the lyrics are not available in Spotify's database, the system searches for the lyrics in alternative databases, such as PyLyrics [111] and Genius [112].

5) TRIP INFORMATION

Having access to longitudinal data requires us to recognize the purpose, duration, and differences or similarities of trips driven by participants. We use the similarity between the locations of the GPS data points to retrieve the trips that are similar to each other. In this way, repeated trips for each participant can be identified and used for future analysis. Additionally, we feed the GPS locations of each trip to Google API and identify (1) the route that the person takes every day, (2) the speed limit in that route for each data point, (3) the number of intersections within that route, and (4) the snap to the road on that specific route. This information is included in HARMONY as we expect this information can further improve how we contextualize the outside environment using different modalities of data (e.g., Google API and features extracted from video streams).

6) DRIVER's SPEED

The vehicle's speed is collected through the camera. The speed of the car is collected with a 1 Hz frequency and is saved to a CSV file associated with the participant and the trip. GPS on the camera lets us decrease the number of devices that are already in the vehicle to enhance the participant's comfort while using such devices. Speed of the vehicle can also be collected through CAN devices or the GPS on the smartwatch with higher frequency if needed.

7) EVENT DETECTION

After retrieving all the contextual elements, we use these measurements to retrieve different events happening in driving.

A

Triggers (Each Trigger should be added here and assigned a code)	Category	description
reverse	driving	driving in reverse gear
eating	food	
drinking	food	
holding the handle	driving	the handle on top of the driver's door
holding phone	phone/w	holding phoen in hands but not working with it
talking to passenger	passenger	engaged in a conversation with the passenger. It has levels from -3 to 3. negative and positive defines the emotional content that is visible. Number defines the amount that the driver is talking in the conversation. For instance +3 is a joyful conversation that driver is talking a lot. 0 is not considered.

B

Video	Timestamp Begin	Timestamp End	Event	Category
2019_0624_095026_160B.MP4	0:00:00	0:00:07	reverse	driving
2019_0624_095026_160B.MP4	0:00:10	0:00:29	eating	food
2019_0624_095026_160B.MP4	0:00:23	0:00:25	checking mirror driver	mirror
2019_0624_095026_160B.MP4	0:00:28	0:00:29	checking mirror driver	mirror
2019_0624_095026_160B.MP4	0:00:25	0:00:29	drinking	food
2019_0624_095026_160B.MP4	0:00:32	0:00:33	checking mirror driver	mirror
2019_0624_095026_160B.MP4	0:00:42	0:00:57	reaching to food	food
2019_0624_095026_160B.MP4	0:00:57	0:01:02	eating	food
2019_0624_095026_160B.MP4	0:01:02	0:01:17	reaching to food	food
2019_0624_095026_160B.MP4	0:01:36	0:01:42	checking mirror middle	mirror
2019_0624_095026_160B.MP4	0:01:17	0:01:39	eating	food

FIGURE 5. Annotation table, sample view of the trigger codes (A), sample annotation for a video (B).

These events provide a multimodal labeled dataset that can be used to train different models (e.g., deep learning models) for classification and prediction (e.g., driver's state recognition) of different activities and behaviors. The event detection is performed both manually and automatically. The manual event detection is performed on a random subset of the data from each participant to create ground truth and training datasets. The automatic event detection is performed on all the collected data.

a: MANUAL EVENT DETECTION

The manual process recognizes events and actions in three categories: the environment, driver's state, and driver's actions. To perform the detailed annotation, we ask the annotator(s) to view both frontal and in-cabin videos simultaneously. The annotator(s) is provided with a table that includes already known actions, environmental situation, and driver's state. A sample view of the annotation table can be viewed in Fig. 5. The detail of each group of annotations is provided below:

- **Environment**: in this category, all the environmental-related instances will be annotated. For instance, each outside video is annotated based on road type (i.e., driving in a city street, one-lane to six-lane highway, and parking lot), weather condition, presence of other specific road users (i.e., bike, bus, trucks, and pedestrians), passing by an intersection, and traffic patterns and density. Each inside video is also annotated based on the presence of passenger(s) and light intensity (binary of being dark or not dark).
- **Task**: driver's tasks, including both primary (i.e., directly related to driving) and secondary tasks (i.e., not related to driving), are manually identified and annotated. The primary tasks include performing change lane and u-turn, checking mirror (one task for each mirror), fastening and unfastening seat belt, and the secondary tasks include eating, drinking, working with phone, talking on the phone, holding the phone, checking the speed stack, working with the center stack, talking with the passenger(s), dancing to the music, the placement of hands on the steering wheel (i.e., both hands, one hand, and none), opening, and closing window.
- **State**: driver's state indicators such as specific facial expressions are recorded in this category. This will help

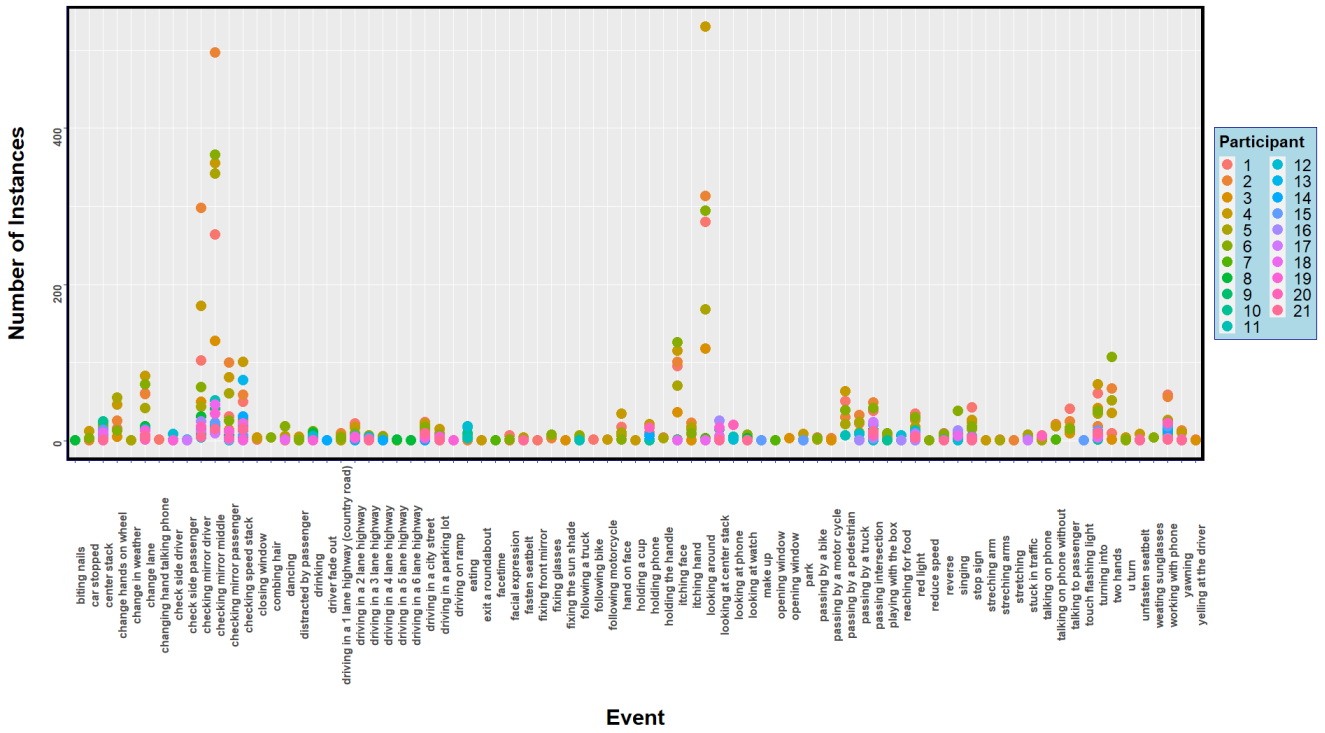

FIGURE 6. Number of the annotated instances for part of the data that only video is available to date.

with understanding sudden changes in driver's states that are visible in the videos and comparing them with driver's physiological measures, outputs of facial analysis software, and environmental conditions. This annotation includes sudden happiness, sadness, anger, being bothered by glare or sudden change in light intensity, and excessive sweating.

Each task is annotated from the second that it is visibly started in the video until the moment that there is no sign of that task in the video. Fig. 6 and 7 depict the number of instances per participant that has been annotated up until now. The annotation is an on-going task and the most recent result for this section of the dataset can be viewed on [113].

b: AUTOMATIC CLASSIFICATION AND EVENT DETECTION

after performing manual event detection, we also analyze the videos using already developed deep learning based algorithms such as object [114] and sign detection [115], and in-house lane and lane-change detection. These algorithms help us retrieve events that include certain objects in the field of views such as passing by a bike, a truck, a bus, an intersection, or a pedestrian, or detecting certain food object in the cabin, number of people in the vehicle, and holding a phone or other objects. These events then help us analyze driver's behaviors and states in specific situations such as changing lanes, passing a cyclist, or arriving a yellow light at an intersection.

V. PARTICIPANT RECRUITMENT

To collect participant's data, we received an approval from the University of Virginia's (UVA) Institutional Review Board for Social & Behavioral Sciences. For the first phase of the study, 21 individuals who were either a student and faculty from UVA or were working professionals within the state of Virginia. Participants were required to have a valid driving license and owned a personal vehicle. Participants were instructed on how to use the equipment, their right, and obligation, as well as details on how to charge their smartwatch and upload the collected data periodically. Each participant owns all of their data, meaning that they were first requested to review all of their videos, delete any or all segments of the videos prior to providing them to the research team. The videos that include people who we do not have consent form were deleted and not used in this dataset. Each participant is assigned a participant ID for identification. Each participant received $50 for every 30 hours of complete data (i.e., data from both smartwatch and camera). 21 participants (11 females and 10 males), between the ages 21 to 33 have joined the study as of November 2020.

Fig. 8 demonstrates the locations of all the collected data to date. The most recent map of the collected data can be viewed in [116]. The data has been mostly collected from eastern and northeastern regions of the United States, including states of Virginia, Pennsylvania, Delaware, West Virginia, Indiana, Illinois, Ohio, Vermont, New Hampshire, Maine, and New York. This data collection started in June 2019 and is currently on-going. It should be noted that the participants in phase I are from the state of Virginia, thus most of the trips are generated in the state of Virginia. Additionally, details of the collected data such as frequency of current data collection for each sensor and duration, can be viewed on Table 2.

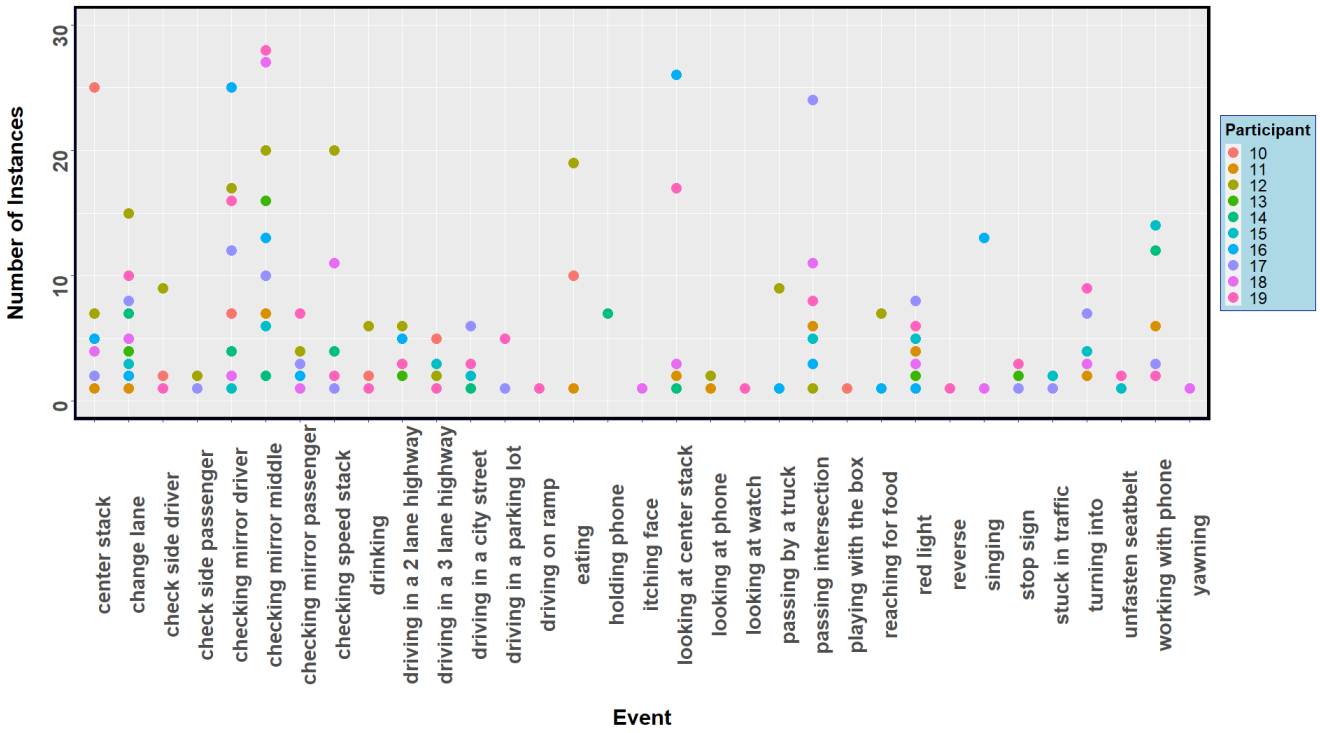

FIGURE 7. Number of the annotated instances for part of the data that all modalities are available, to date. Note that these include less data as not every video has an available physiological data attached to it.

FIGURE 8. GPS location of the collected data to date. Although participants are mostly from the state of Virginia, the data includes the roads from northeastern regions of the country.

VI. CASE STUDY

In this section, we present a case study to highlight the importance of collecting physiological data in addition to video streams in longitudinal naturalistic studies. We provide evidence on large variability in a participant's HR when driving through the same route on different days. Furthermore, we demonstrate the minimum number of days required for capturing statistical variability among a participant's physiological data while driving. Furthermore, we show how a subset of HARMONY's dataset can be used to detect driver's state changes. We append a sample dataset that includes two trips of around two hours in a vehicle equipped with HARMONY. This trip has been chosen specifically due to its diversity in providing different road types (i.e., city versus highway), weather conditions (i.e., sunny, cloudy, and rainy), and daylight variations (i.e., evening, night, and daylight). The sample dataset can be accessed via the following Open Science Framework (OSF) provided in [117].

This sample dataset includes the camera recording of the in-cabin conditions and the outside environment, in addition to the smart wearable sensing data (i.e., PPG, HR, hand IMU, and light level). Using the appended data and by specifically relying on the driver's HR, we demonstrate the relationship between the driver's physiological measures and changes in

TABLE 2. Details of collected data to date through HARMONY framework.

Device/software	Sensor	Type	Frequency	Model	Collected Data to Date (Hours)	Additional Info
Smart Watch	Hand Acceleration	Physical	10 Hz	Fossil	150	-
	Hand Gyroscope	Physical	10 Hz	Fossil	150	-
	Heart Rate	Physical	1 Hz	Fossil	150	-
	Light	Physical	1/60 Hz	Fossil	150	-
	Noise	Physical	1/60 Hz	Fossil	150	-
	Location	Physical	1/600 Hz	Fossil	150	-
Dash Camera	Inside Camera	Physical	30 fps	Blackvue DR750S-2CH	380	Full HD 1080P resolution
	Outside Camera	Physical	30 fps	Blackvue DR750S-2CH	380	Full HD 1080P resolution
	Speed	Physical	1 Hz	Blackvue DR750S-2CH	380	-
OpenFace	Gaze	Virtual	30 fps	-	380	-
OpenPose	Pose	Virtual	30 fps	-	380	-
Imotion	Facial Emotion	Virtual	30 fps	-	380	-

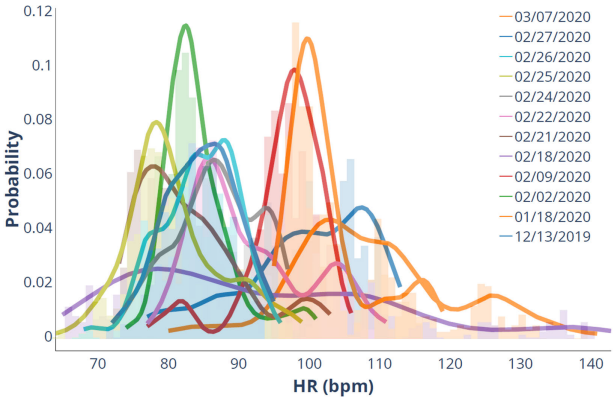

FIGURE 9. A participant's HR distribution while driving through the same route for multiple days. Note that the distribution changes on a daily basis, pointing to a need in collecting long-term longitudinal data.

outdoor and in-cabin activities/events (section VI-B). In this section, we also show how change point detection methods such as Bayesian Change Point (BCP) can be utilized in detecting important peaks in HR data in relation to changes in in-cabin and outdoor conditions.

A. HOW MUCH PHYSIOLOGICAL DATA IS NEEDED?

Previous studies have indicated different driving conditions may influence driver's physiological factors; however, many of these studies were utilizing short-term physiological sensing and conducted in controlled experimental studies [23], [47], [48]. With short-term behavioral and physiological data, we may not be able to properly capture the underlying variability in a driver's state changes. This is mainly because a human's physiological measures are dynamic and can be impacted by many different factors. People's HR, for instance, has different baselines and distributions on different days or after certain activities (e.g., going on a run). To ensure that enough data is collected in a noisy natural environment, the experimenter first needs to confirm that the captured data demonstrates the underlying variability in variables of interest. In other words, every trip of driving data can have a different underlying distribution, which requires the experimenter to collect enough trips that can capture the summation of as many distributions throughout different days of driving. Fig. 9 shows the distribution of HR data collected from one of our participants driving from home to the workplace throughout different days, for a period of mid-December 2019 to early March 2020. We have specifically chosen this example to demonstrate that distributions can be very different, even if the outside context is somewhat the same. To find out the minimum number of days required for collecting participant physiological and behavioral data, we used the Kernel Density Estimation (KDE) to assess the variation in distributions [118].

Previous research has used KDE to assess the amount of data needed for an NDS provided in [118]. We can apply the same method on the HARMONY dataset to calculate the estimation of the kernel density of the data and assess its variability with respect to adding more data. If adding more data on a daily basis causes the distribution to change significantly, it means that more data is needed. We are interested in finding the saturation point at which adding more data does not cause the overall distribution to change significantly. If a few consecutive days of driving in a naturalistic setting can provide enough physiological data, then the kernel density estimation should not change significantly as more data points are being added.

In order to estimate the underlying distribution function of a given sample of data, KDE can be used for its robustness in estimating the kernel in a non parametric fashion [118]. Through this method, the estimation for the probability distribution can be calculated as shown in equation 1 [118], [119]. Considering we have a sample sequence of data $\{X_i\}_1^n$:

$$\hat{f}(x) = n^{-1} \sum_{1}^{n} h^{-1} K(x - X_i) \tag{1}$$

where h is the bandwidth and K is the kernel function. The estimate of the probability distribution is a function of bandwidth (h). The bandwidth value can result in variations in different levels of smoothness of the estimation, with higher values providing a smoother estimate. When estimating a probability distribution through this method, we can compare the similarities of two distributions using the Kullback-Liebler (KL) divergence index to identify how much information is lost when we approximate one distribution with another. KL is calculated as:

$$KL(\hat{f}(x;n)||\hat{g}(x;m)) = \int [\hat{f}(x;n) * log(\frac{\hat{f}(x;n)}{\hat{g}(x;m)})] \tag{2}$$

where $\hat{f}$ and $\hat{g}$ are two different kernel density estimates. Now, when having two estimates of kernel density based on n and $n+m$ datapoints, equation (2) becomes:

$$KL(\hat{f}(x;n+m)||\hat{f}(x;n)) = \int [\hat{f}(x;n+m) * log(\frac{\hat{f}(x;n+m)}{\hat{f}(x;n)})] \tag{3}$$

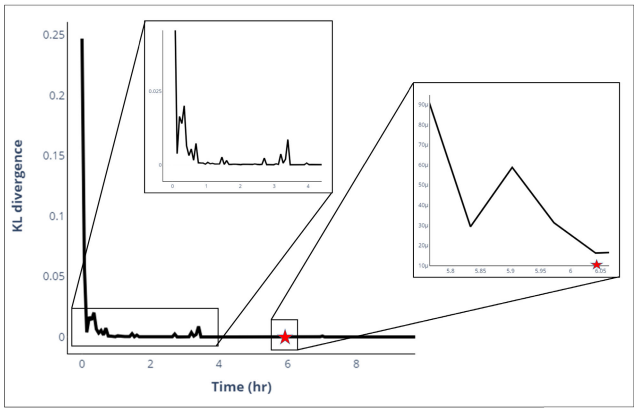

FIGURE 10. Kernel density estimation analysis for 3 months of HR data. The KL divergence algorithm demonstrates that after approximately 3 months of data collection (6 hours), the difference between estimated kernels reaches below 1e-5.

If the KL-divergence value is below a specified ϵ, then the two density functions are considered to be similar, meaning that adding more data will not add more information to the statistical variability of the factor of interest. In order to show why we cannot always rely on short-term driving behavior studies and why NDS are needed to also include physiological data longitudinally, we have applied the above method to the HR data collected from one of our participants over the course of three months from mid-December to mid-March. We have specifically chosen this participant as the data provides consecutive days of driving through the same local areas within the city. This helps us decrease the variability among external factors. Note that the driving duration varies day by day, and weeks of data collection is required to provide hours of driving data. For our problem, we have used a Gaussian Kernel, with a bandwidth of 0.2. We have gradually added m amount of data (every 250 seconds of data, which is roughly equal to one day of driving data) to the KDE function and estimated the KL-divergence between the consecutive estimates until the difference is below an $\epsilon = 1e - 5$. Fig. 10 demonstrates the difference between the kernel density estimations as more data is being added. As shown on the graph, this happens at around 6 hours of data which is approximately equal to 2 months of physiological data while daily driving for this participant.

In this example, we demonstrated that a short period of driver's state measures is not enough for capturing real variability in the data. Additionally, the results show even driving through the same route can be accompanied by very different distributions of the driver's HR levels. These differences indicate the baselines of HR in different days can vary and cause difficulties in making inferences if it is not collected for a long enough duration. Additionally, it should be considered that in a naturalistic setting, other factors can also affect the variables of interest, which can increase the required data collection period. For instance, in our case of analyzing HR data, activities prior to driving, emotional responses, number of passengers in the vehicle, and the traffic density can all affect the driver's HR distributions. Thus it is important to confirm the collected data captures the statistical variability.

B. MODELING DRIVING EVENTS USING PASSIVE SENSING

The second part of this case study demonstrates how driving events can be monitored and analyzed using passive sensors. Specifically, we show how using HR readings can identify different underlying state changes of the driver when specific in-cabin or outdoor events take place. We first identify the locations of abrupt changes in the HR readings and then identify the reasons behind the abrupt changes through the events that take place in a given time frame. To achieve this, we have annotated parts of the HARMONY 1.0 dataset based on the annotation scheme detailed out in section IV-B7 and analyzed the consistency between these events and fluctuations in passive sensor data (i.e., HR, accelerometer). We apply Bayesian Change Point (BCP) detection methods to detect the abrupt changes in HR data.

In this subsections below, we first provide an overview of the BCP method, then discuss the reasons behind each change point detected in the HR.

1) BAYESIAN CHANGE POINT ANALYSIS

We consider the problem of detecting changes in the HR as a Bayesian change point detection problem. This approach allows for easy quantification of uncertainty and integration of priors. We leverage Barry and Hartigan's [120] Bayesian change point model for this analysis. This model assumes there is an unknown partition ρ of the data in the contiguous regime, such that within each regime, the HR remains the same. The new external event typically happens between two blocks when the HR goes up. The model also assumes an independent normal distribution for each block.

Let us assume we have n HR data points $\{X_1, \ldots, X_n\}$. We will use X_{ij} to refer to the observations between indices i and j. Let $\rho = (U_1, \ldots, U_n)$ indicate a partition of the time series into non overlapping HR regimes. We use a Boolean array of change points to denote the regimes. At each time step, if U_i takes a value 1, we have a new HR regime (possibly due to an external event); else we remain in the same regime.

We are interested in the posterior density $f(\rho|X)$. By Baye's theorem, this can be written as

$$f(\rho|X) \propto f(X|\rho)f(\rho) \tag{4}$$

Prior cohesion density: Let p denote the probability of getting a change point at each location. We assume this probability to be the same at each location. If we assume that there are b partitions, the prior cohesion density can be written as

$$f(\rho|p) = p^{b-1}(1-p)^{n-b} \tag{5}$$

The joint density of observations and parameters given ρ is a product of densities of different blocks over the blocks in ρ. Let us consider a single block. If we assume that the data in this block is generated by a gaussian with mean θ

TABLE 3. Different driving events detected by the HR change point detection.

Time (UTC)	Reason	Category
23:34:05	blocked - changing lane	Lead Vehicle
23:47:00	changing lane - vehicle passing	Following Vehicle
23:49:50	car on the side	Side Vehicle
23:55:24	yellow light expecting	Intersection
0:01:45	car on the side	Side Vehicle
0:03:17	blocked - changing lane	Lead Vehicle
0:06:38	unknown - possibly a vehicle on the back	Following Vehicle
0:07:40	unknown - possibly a vehicle on the back	Following Vehicle
0:20:02	blocked - changing lane	Lead Vehicle
0:26:02	decreasing speed - traffic density - high	Traffic
0:27:00	Motorcycle passing by right	Cyclist
0:29:57	car following distance decreased	Lead Vehicle
0:31:38	changing lane - vehicle passing	Side Vehicle
0:34:36	car on shoulder	Side Vehicle
0:37:00	blocked	Lead Vehicle
0:39:40	merging in highway	Primary Task
0:42:10	distracted by phone	Secondary Task
0:45:30	yellow light expecting	Intersection
0:47:57	decreasing speed - traffic density - high	Traffic
0:50:07	yellow light expecting	Intersection
0:52:00	merging in highway	Primary Task
0:57:35	merging in highway	Primary Task
1:44:00	lead vehicle changing lane - decrease speed	Lead Vehicle
1:47:53	unknown - possibly phone	Secondary Task
1:51:19	decreasing speed - traffic density - high	Traffic
2:01:46	secondary task - drinking	Secondary Task
2:05:42	trying to do routing - confused	Primary Task
2:09:57	traffic density - high - city	Traffic

and variance σ^2. Let the prior density of θ be a gaussian with mean μ_0 and variance σ_0^2

$$f(X_{ij},\theta) = \Pi f(X_k|\theta)f(\theta)$$
$$f(X_{ij}) = \int \Pi f(X_k|\theta)f(\theta)d\theta \tag{6}$$

The above integral can be simplified to the expression below

$$f(X_{ij}) = (\frac{1}{2\pi\sigma^2})^{(j-i)/2}(\frac{\sigma^2}{\sigma_0^2+\sigma^2})^{1/2}exp(V_{ij}) \tag{7}$$

where

$$V_{ij} = -\frac{\sum_{l=i+1}^{l=j}(X_l-\hat{X}_{ij})^2}{2\sigma^2} - \frac{(j-i)(\hat{X}_{ij}-\mu_0)^2}{2(\sigma^2+\sigma_0^2)} \tag{8}$$

and $\hat{X}_{ij}$ is the mean of the observations in the partition. However $f(X_{ij})$ still depends on the parameters $\mu_0, \sigma^2, \sigma_0^2$. Defining $w = \frac{\sigma^2}{\sigma_0^2+\sigma^2}$ and choosing the following priors for the parameters:

$$f(\mu_0) = 1, \quad -\infty \le \mu_0 \le \infty$$
$$f(p) = 1/p_0, \quad 0 \le p \le p_0$$
$$f(\sigma^2) = 1/\sigma^2, \quad 0 \le \sigma^2 \le \infty$$
$$f(w) = 1/w_0, \quad 0 \le w \le w_0 \tag{9}$$
$$f(X|\rho,\mu_0,w) = \int_0^\infty 1/\sigma^2 \prod_{ij\epsilon P} f(X_{ij})d\sigma^2 \tag{10}$$

After integrating out μ_0 and w, This can be simplified to the indefinite integral below. We refer the readers to [120] for the full derivation.

$$f(X|\rho) \propto \int_0^{w_0} \frac{w^{(b-1)/2}}{(W+Bw)^{(n-1)/2}}dw, \tag{11}$$

where

$$\hat{X} = \sum_{i=1}^{n} X_i/n, \quad B = \sum_{ij\epsilon P}(j-i)(\hat{X}_{ij}-\hat{X})^2,$$
$$W = \sum_{ij\epsilon P}\sum_{l=i+1}^{l=j}(X_l-\hat{X}_{ij})^2 \tag{12}$$

Similarly, after integrating out the change probability p, the prior cohesion density thus can be written as

$$f(\rho) \propto \int_0^{p_0} p^{b-1}(1-p)^{n-b}dp \tag{13}$$

To calculate the posterior distribution over partitions, we use Markov Chain Monte Carlo (MCMC) [121]. We define a Markov chain with the following transition rule: with probability p_i, a new change point at the location i is introduced. Here B_1, W_1 and B_0, W_0 refer to the expressions in (12) with and without the change point in location i.

$$\frac{p_i}{1-p_i} = \frac{p(U_i=1|X,U_j,j\neq i)}{p(U_i=0|X,U_j,j\neq i)}$$
$$= \frac{\int_0^{p_0} p^b(1-p)^{n-b-1}dp}{\int_0^{p_0} p^{b-1}(1-p)^{n-b}dp} x \frac{\int_0^{w_0} \frac{w^{b/2}}{(W_1+B_1w)^{(n-1)/2}}dw}{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0+B_0w)^{(n-1)/2}}dw} \tag{14}$$

This ultimately leaves us with a probabilistic model with the following two parameters p_0 and w_0. We use the package *bcp* [122] in R programming language to implement our change point analysis.

2) CHANGE POINT ANALYSIS

Fig. 11 depicts the driver's HR during the trip selected for this case study. Using BCP, we detect the specific moments that the underlying distribution of HR data changes, which in this case, the change point is associated to time points that the HR increases from its baseline value for a short amount of time. Fig. 11 - A, shows the overall time series of HR for this trip (black), together with the mean computed value from the BCP method (blue), as well the probability of detecting the change point events (red).

For each one of these change points, we have manually analyzed the video streams to find out the reason behind them. Here we hypothesize that each change point is related to an internal or external event that is accompanied by the driver's HR changes. An example of the events associated with these peaks is then demonstrated in the parts B and C of Fig. 11. Note that due to the low amount of light in the environment, some of the following events may not be detectable if we only rely on camera streams.

We performed the same analysis on approximately 2 hours of driving data randomly drawn from 9 other participants that have been collected through the HARMONY dataset. The HR change points coupled with the respective video epochs reveals these change points are happening simultaneously with certain categories of in-cabin or outside events. Specifically, the following categories of events are identified that

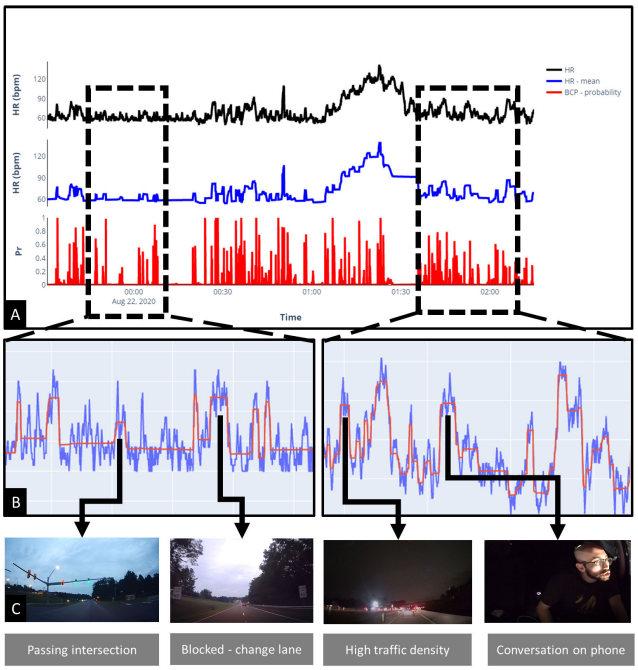

FIGURE 11. Participant's HR values together with the overlay of mean values calculated by BCP (A) the time series of participant's HR (black) together with average HR value between each two consecutive change points (blue) and the probability of detecting a change points (red) (B) the visualization of the sample change points from the videos in an arbitrary section of the time series (C) the regimes identified by BCP correspond to meaningful external events. In C1, the driver is passing through an intersection. In C2 he is blocked while trying to switch lanes, causing his HR to increase. In C3 and C4 the driver arrives at a high traffic region and is talking on the phone.

correlate with the detect HR change point events within the 9 participants:

- Lead vehicle: HR variation that is accompanied by the presence of the lead vehicle such as decreasing the speed, being blocked, abrupt change lanes, and abrupt breaking patterns.
- Arriving at an intersection: events where the driver is arriving or passing through an intersection. For instance, when the vehicle is stopped at the red light versus when the driver passes through the intersection.
- Following and side vehicles: a vehicle that follows the participant too closely or is passing the vehicle.
- Driver's tasks: this category is divided into primary and secondary tasks that drivers are engaged with while driving. Primary tasks include those directly related to driving, such as changing lane and checking mirrors. Secondary tasks include activities such as holding/talking on the phone, working with the center stack, or other non-driving related tasks.
- Traffic pattern: incidents that the driver has to decrease the vehicle's speed, or have abrupt breaking patterns due to inconsistency in traffic patterns and conditions such as arriving at a high traffic density segment on a highway.
- Other roadway users: this category includes pedestrians, motorcycles, trucks, buses, cyclists, and other roadway users that the participant passes by.

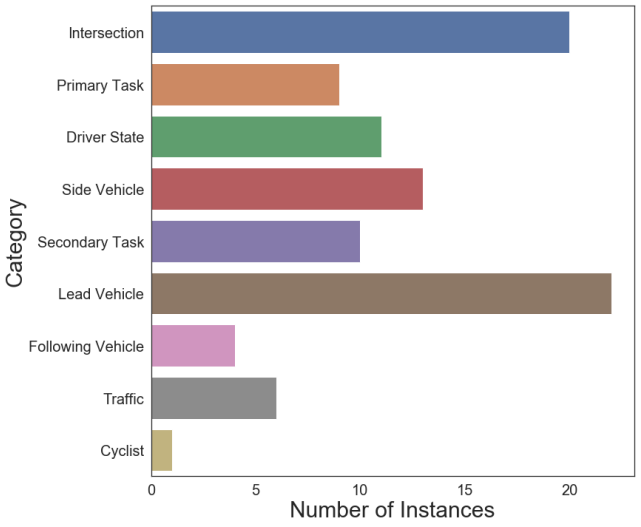

FIGURE 12. Count in each category of events happening in all 10 participants' data. Note that the two top occurring categories are intersection and lead vehicle. It's important to note that current autonomous systems have a one size fits all approach for situations when these parameters are present (e.g., a common car-following distance) which can result in different states for their users.

- Driver's state: this category includes any driver state that cannot be fully seen as an event in the in-cabin or outside video; however, there are visible changes in the participant's facial features such as the participant is smiling or frowning.

Fig. 12 provides the count per category for these events. One important note here is that similar events in the past have been shown to be the cause of emotional and stressful events. For instance, [23] mentions that the triggers of emotions can be due to traffic and driving tasks, human-computer interaction and navigation, vehicle and its equipment, and environmental factors. Furthermore, numerous studies have demonstrated that triggers of emotions can be accompanied by abrupt increases in HR values [123]. Thus it can be possible, that the trigger categories were responsible for abrupt changes in the psychophysiological state of the driver, which was then captured in HR (while not being fully visible in the vision module of the data).

By using the HARMONY framework, we have demonstrated that external events of interest can be automatically detected by applying BCP methods to the driver's physiological measures. The multimodal approach of HARMONY provides a deeper context to each external event happening in the driving scenario. Furthermore, by adding automatic computer vision techniques, HARMONY can reason more deeply about each external event detected through vision modules, such as passing an intersection, the presence of a lead vehicle, and performing a change lane.

VII. DISCUSSION AND FUTURE WORK

The vehicle industry is advancing very quickly and new technologies are introduced to automate different aspects of driving. Although these improvements are intending to

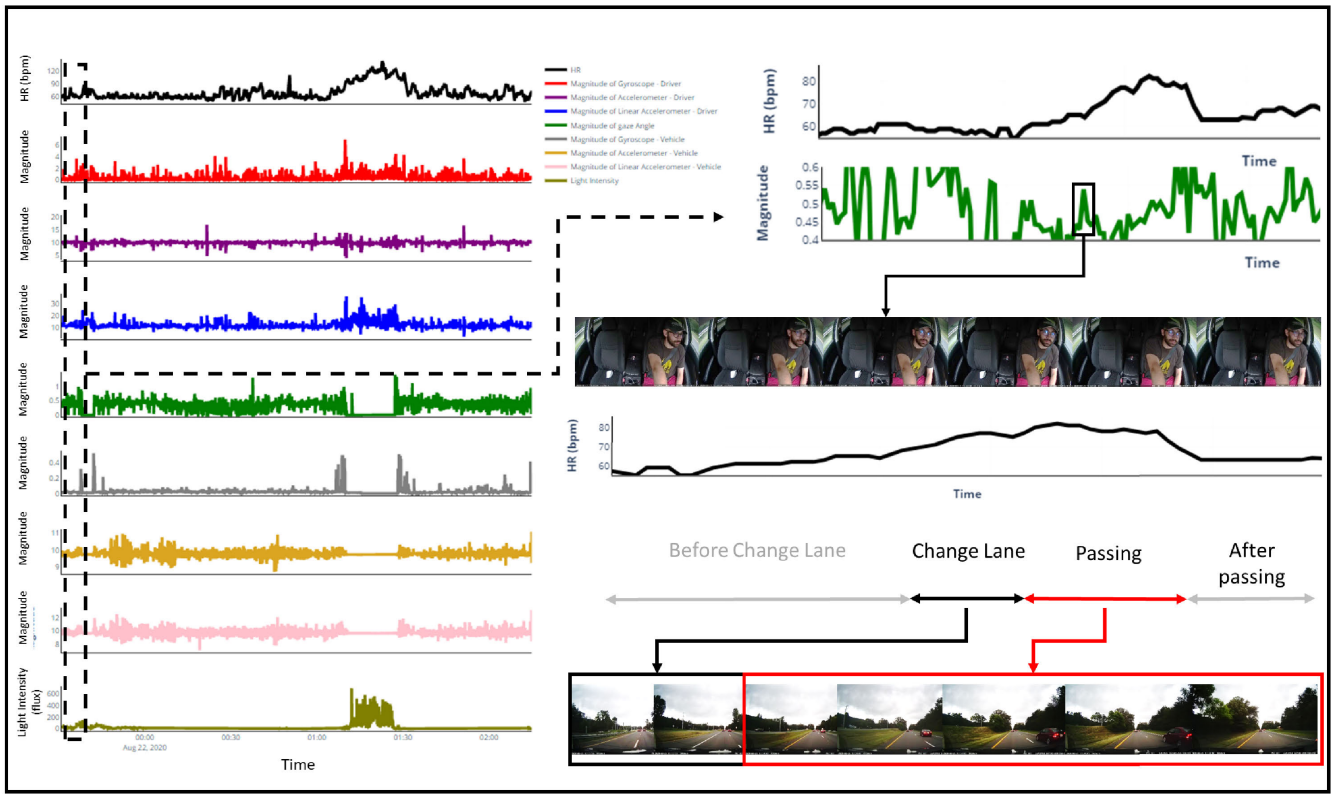

FIGURE 13. Different modalities of data in a lane change action (left). Looking deeper into the physiological data provides insight on how events like lane change can result in a prolonged increase in HR, even post removal of the initial stimulus. (right). It's important to note that it's hard to get such psychophysiological insights based purely on modalities like vision. The video for this event is provided as a demo for the paper.

enhance the driver's safety and comfort, they still lack an understanding of driver states and behaviors. In order to achieve an acceptable level of shared-autonomy, the new improvements in the automobile industry should be able to detect and respond to the driver's states and behaviors in real-time. In this section, at first, we discuss the implications of the proposed HARMONY framework and how physiological data can be used to classify driving events, task engagement, and stress level. Such information is critical for achieving shared-autonomy in future AVs.

The first implication for AV is to use the change points to better receive feedback for their decision on the road. Currently, AV can accurately detect and classify different outside conditions such as traffic light, passing through an intersection, presence of a lead vehicle, passing a cyclist, and many other objects and road conditions. However, we might behave differently or psychologically feel differently while passing through similar or even the same road conditions as before. However, AV is not capable of detecting and classifying these driver-specific states. For instance, in the presented case study, we observe when the driver is approaching an intersection with the traffic light turning yellow, his HR data indicates a sudden change in state as the HR data elevates as soon as the traffic light is within his field of view. In another scene we observe that the driver has no sudden change in his HR when he sees an intersection with a red light from the far. This may indicate the color of the traffic light, and the traffic patterns at one intersection might cause a higher level of stress for the driver. Such information can assist AV in their motion planning as well as safety considerations such as deceleration rates and car-following behaviors specific to each driver's preference and comfort levels [5]. Additionally, the AV's decision, if not aligned with the driver's choice, can negatively affect AV's acceptance among users as compared to a preferred decision [5]. The presented HARMONY framework aims to highlight the importance of driver-in-the-loop naturalistic studies, where driving experiences are contextualized based on in-cabin and outdoor conditions as well as the driver's behaviors and psychophysiological states.

Another implication of the presented work is to identify the psychophysiological effect of each event on the driver. For instance, whether an AV should decide to pass a lead vehicle or increase its distance with it. Preferring each one of these decisions can be different among different drivers and contextual settings (e.g., traffic density). This means that passing a vehicle can be affecting the driver's state differently than increasing the distance, and this variation can also be different among different drivers. To better grasp these characteristics, we provide a closer look at one of the change points events within the presented case study dataset. Fig. 13 depicts a change event that is accompanied by an abrupt peak in the HR data. In this event, the driver attempts to perform a change lane due to being blocked by the vehicle ahead. After checking the mirror (green), the driver performs the lane

TABLE 4. List of the detected change points together with their underlying reason and category in a random sample drawn from 9 participants HR data.

Time (UTC)	Reason	Category	Time (UTC)	Reason	Category
19:57:05	yellow light observed	Intersection	14:58:16	blocked - change lane	Lead Vehicle
19:59:41	red light	Intersection	14:59:30	change lane	Primary Task
20:03:54	waiting for long red light	Intersection	14:59:57	blocked - change lane	Lead Vehicle
20:04:18	attempting to cross	Primary Task	15:00:48	driver fading	Driver's internal State
20:24:46	yellow light expecting	Intersection	15:02:30	passing by a truck	Side Vehicle
20:25:19	Probable internal stress with lips biting	Driver Internal State	16:06:32	lead vehicle slowing down	Lead Vehicle
21:25:27	Observing a car attempting to merge	Side Vehicle	16:07:28	blocked - change lane	Lead Vehicle
21:25:36	waiting for the car to merge	Side Vehicle	16:09:24	passing by a truck	Side Vehicle
16:38:02	internal stress	Driver Internal State	16:10:09	passing by a truck	Side Vehicle
19:30:54	yellow light expecting	Intersection	17:20:51	attempting to install phone	Secondary Task
19:31:53	looking for phone	Secondary Task	17:21:18	internal stress - visible on face	Driver Internal State
19:33:10	driver yawning	Driver Internal State	17:23:11	passing by a truck	Side Vehicle
19:48:41	receiving a happy text - visible on face	Secondary Task	17:24:04	lead vehicle slowing down	Lead Vehicle
19:51:24	stop sign	Intersection	17:26:15	lead vehicle slowing down	Lead Vehicle
21:23:13	lead vehicle stops	Lead Vehicle	17:34:52	attempting to merge	Primary Task
21:24:06	lead vehicle stops	Lead Vehicle	17:35:25	blocked - change lane	Lead Vehicle
21:25:06	waiting for long red light	Intersection	17:37:37	blocked	Lead Vehicle
17:46:48	followed by a vehicle	Following Vehicle	16:56:23	lead vehicle slowing down	Lead Vehicle
20:29:57	lead vehicle slowing down	Lead Vehicle	16:57:53	attempting to cross the road	Primary Task
16:35:54	lead vehicle slowing down	Lead Vehicle	20:00:16	yawning	Driver Internal State
16:36:36	red light	Intersection	20:59:24	more intense singing	Driver Internal State
16:37:47	yellow light expecting	Intersection	14:28:13	engaged in talking	Driver Internal State
16:40:40	passing intersection	Intersection	14:49:48	passing by a truck	Side Vehicle
16:41:07	yellow light expecting	Intersection	22:14:39	stuck in traffic	Traffic
21:22:43	passing by a truck	Side Vehicle	18:53:20	blocked - lead vehicle slows down	Lead Vehicle
16:50:05	distracted by phone	Secondary Task	12:50:25	stuck in traffic	Traffic
16:51:21	distracted by phone	Secondary Task	12:51:27	passing intersection	Intersection
16:55:57	distracted by phone	Secondary Task	21:32:33	red light	Intersection
16:57:46	yawning	Driver Internal State	20:49:17	passing intersection	Intersection
17:11:32	starting to sing	Driver Internal State	20:49:58	passing intersection	Intersection
14:45:51	lead vehicle slowing down	Lead Vehicle	20:49:27	merging	Primary Task
14:48:00	attempting to grab an item	Secondary Task	20:53:08	blocked - change lane	Lead Vehicle
15:50:15	driver moving - feeling uncomfortable	Driver Internal State	20:53:43	increasing speed to pass a car	Side Vehicle
16:01:42	observing intersection sign	Intersection	21:18:57	yellow light expecting	Intersection

change. The HR of the driver stays elevated throughout the whole passing action. Afterward, when the driver completely passes the vehicle, and there is no vehicle in the front, the HR starts to decrease to the baseline value. The important point here is that events such as being blocked by a vehicle can keep the driver's HR elevated for a long period of time after the event ends, which is not recognizable using video feeds at all. In this example, the gaze signal shows the mirror checking (green) but does not show further information about the driver's stress levels, which is positively correlated with the changes in HR.

It is important to note that in this paper, we only focused on demonstrating how the HR data can be utilized to detect different driving event changes. However, by coupling HR measures with the other data streams from the wearable devices such as the IMU, light, and noise level information, it is possible to more accurately detect and classify the driver's engagement levels, activities, and emotional states. These data streams are also included in the exiting HARMONY dataset.

Additionally, through our case study, we have demonstrated that stability in the underlying distribution of data collected from one participant was achieved after more than two months of data collection. Note that here we have only analyzed the variability of driver's HR data when commuting through similar routes. Driving through different routes as well as analyzing other modalities of data with the same method will result in an even longer required data collection. This matter emphasizes the importance of longitudinal data collection for driver state detection.

One important aspect of our work is the presence of valid and accurate groundtruth for the data, which is currently done manually. The task of manual annotation is time-consuming, resource extensive, and somewhat tedious. To address these issues, we have employed three major strategies. First, as shown in the presented case study, we are utilizing machine learning techniques to fuse video features and wearable data to train classifiers for detecting different driving behaviors, states, and activities. As our on-going and future efforts, we will develop open-source classifiers that can identify safety-related driving behaviors and events. Such classifiers can also be applied on existing NDS datasets to automatically detect certain events or activities from the video streams previously collected from in-cabin and outside driving activities. Additionally, we have employed a group of annotators to increase the manual process speed. Finally, we have enhanced the annotation process by choosing among videos that have certain characteristics such as higher change points in the physiological data.

Similar to previous NDS, the HARMONY dataset currently includes data from roadways in Virginia and the northeast of the United States. As future work, we plan to expand the HARMONY dataset by recruiting participants across the different states in the United States. Additionally, our

participants are currently in the age group of 20-35 years old. When considering human physiological measures and driving behaviors, it is important to have a diverse pool of participants (age, gender, race, socio-economic level, etc.) to better capture individual differences. In the future phases of HARMONY, we will expand our participant recruitment pool to increase the diversity among our participant population. Moreover, as part of the next phase of HARMONY, we will include semi-automated vehicles such as commercially available TESLA vehicles to test our developed framework and resulting models in different levels of shared-autonomy. Lastly, we intend to integrate emerging edge computing devices such as the NVIDIA Jetson family to collect and analyze data from different modalities in real-time. Such embedded systems can significantly improve data collection and processing capabilities within NDS, such as HARMONY.

VIII. CONCLUSION

This paper introduces the HARMONY framework for conducting multimodal longitudinal driving studies in the wild. In contrast to the previous studies, HARMONY takes advantage of ubiquitous computing techniques to collect changes in driving contexts such driver's physiological and emotional states or changes in environmental conditions. The proposed framework integrates low-cost and commercially available devices as well as open-source machine learning algorithms to contextualize driving experiences. HARMONY is setup to collect driver's preferences as well as environmental attributes that cannot be retrieved from outside/inside cabin videos through different physical and virtual sensors. Through implementing HARMONY, we can better reason about different in-cabin and outside driving events. Furthermore, by providing the driver's physiological measures, HARMONY delves into the effect of each contextual element on the driver's state. This study paves the way for helping AVs better understand the driver's state, which can be used to provide better feedback in a shared autonomy approach.

APPENDIX

Table 4 provides all of the events that were detected through annotating HR change points from 9 participants driving data together with their underlying reasons, and category.

ACKNOWLEDGMENT

The authors would like to thank the UVA Link Lab and Commonwealth Cyber Initiative (CCI) for providing support and resources to enable this project. Additionally, they would like to thank Megan Lin for helping with data cleaning, annotation, and processing. Also, they are thankful to the UVA Institutional Review Board for their continuous support and feedback.

REFERENCES

[1] L. Fridman, "Human-centered autonomous vehicle systems: Principles of effective shared autonomy," 2018, *arXiv:1810.01835*. [Online]. Available: http://arxiv.org/abs/1810.01835

[2] J. Terken and B. Pfleging, "Toward shared control between automated vehicles and users," *Automot. Innov.*, vol. 3, no. 1, pp. 53–61, Mar. 2020.

[3] L. Feng, C. Wiltsche, L. Humphrey, and U. Topcu, "Synthesis of human-in-the-loop control protocols for autonomous systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 450–462, Apr. 2016.

[4] G. Li, Y. Yang, T. Zhang, X. Qu, D. Cao, B. Cheng, and K. Li, "Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102820.

[5] S. Y. Park, D. J. Moore, and D. Sirkin, "What a driver wants: User preferences in semi-autonomous vehicle decision-making," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–13.

[6] V. Butakov and P. Ioannou, "Personalized driver/vehicle lane change models for ADAS," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4422–4431, Oct. 2015.

[7] R. Fernandez-Rojas, A. Perry, H. Singh, B. Campbell, S. Elsayed, R. Hunjet, and H. A. Abbass, "Contextual awareness in human-advanced-vehicle systems: A survey," *IEEE Access*, vol. 7, pp. 33304–33328, 2019.

[8] B. Azari, C. Westlin, A. B. Satpute, J. B. Hutchinson, P. A. Kragel, K. Hoemann, Z. Khan, J. B. Wormwood, K. S. Quigley, D. Erdogmus, and J. Dy, "Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience," *Sci. Rep.*, vol. 10, no. 1, pp. 1–17, 2020.

[9] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," Nat. Highway Traffic Saf. Admin., Washington, DC, USA, Tech. Rep. DOT HS 810 594, 2006.

[10] W. Brodsky and Z. Slor, "Background music as a risk factor for distraction among young-novice drivers," *Accident Anal. Prevention*, vol. 59, pp. 382–393, Oct. 2013.

[11] V. Alizadeh and O. Dehzangi, "The impact of secondary tasks on drivers during naturalistic driving: Analysis of EEG dynamics," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 2493–2499.

[12] W. Brodsky, "A performance analysis of in-car music engagement as an indication of driver distraction and risk," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 55, pp. 210–218, May 2018.

[13] G. Li, W. Lai, X. Sui, X. Li, X. Qu, T. Zhang, and Y. Li, "Influence of traffic congestion on driver behavior in post-congestion driving," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105508.

[14] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, "An overview of the 100-car naturalistic study and findings," *Nat. Highway Traffic Saf. Admin., Paper*, vol. 5, p. 400, Jun. 2005.

[15] R. Eenink, Y. Barnard, M. Baumann, X. Augros, and F. Utesch, "Udrive: The European naturalistic driving study," in *Proceedings of Transport Research Arena*. Paris, France, IFSTTAR, 2014.

[16] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, A. Patsekin, J. Kindelsberger, L. Ding, S. Seaman, A. Mehler, A. Sipperley, A. Pettinato, B. D. Seppelt, L. Angell, B. Mehler, and B. Reimer, "MIT advanced vehicle technology study: large-scale naturalistic driving study of driver Behavior and interaction with automation," *IEEE Access*, vol. 7, pp. 102021–102038, 2019.

[17] T. A. Dingus, F. Guo, S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, and J. Hankey, "Driver crash risk factors and prevalence evaluation using naturalistic driving data," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 10, pp. 2636–2641, Mar. 2016.

[18] A. Das, A. Ghasemzadeh, and M. M. Ahmed, "Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data," *J. Saf. Res.*, vol. 68, pp. 71–80, Feb. 2019.

[19] S. L. Hallmark, S. Tyner, N. Oneyear, C. Carney, and D. McGehee, "Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data," *J. Saf. Res.*, vol. 54, p. 17, Sep. 2015.

[20] G. Prabhakar, A. Mukhopadhyay, L. Murthy, M. Modiksha, D. Sachin, and P. Biswas, "Cognitive load estimation using ocular parameters in automotive," *Transp. Eng.*, vol. 2, Dec. 2020, Art. no. 100008.

[21] I. Abdic, L. Fridman, D. McDuff, E. Marchi, B. Reimer, and B. Schuller, *Driver Frustration Detection From Audio and Video in the Wild*, vol. 9904. Cham, Switzerland: Springer, 2016, p. 237.

[22] O. Carsten, K. Kircher, and S. Jamson, "Vehicle-based studies of driving in the real world: The hard truth?" *Accident Anal. Prevention*, vol. 58, pp. 162–174, Sep. 2013.

[23] S. Zepf, M. Dittrich, J. Hernandez, and A. Schmitt, "Towards empathetic car interfaces: Emotional triggers while driving," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–6.

[24] S. Al-Sultan, A. H. Al-Bayatti, and H. Zedan, "Context-aware driver behavior detection system in intelligent transportation systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4264–4275, Nov. 2013.

[25] M. Z. Baig and M. Kavakli, "A survey on psycho-physiological analysis & measurement methods in multimodal systems," *Multimodal Technol. Interact.*, vol. 3, no. 2, p. 37, May 2019.

[26] H. Tankovska. (Sep. 2020). *Global Connected Wearable Devices 2016–2022*. [Online]. Available: https://www.statista.com/statistics/487291/global-connected-wearable-de%vices/

[27] S. S. Coughlin and J. Stewart, "Use of consumer wearable devices to promote physical activity: A review of health intervention studies," *J. Environ. Health Sci.*, vol. 2, no. 6, pp. 1–6, 2016.

[28] M. Kos and I. Kramberger, "A wearable device and system for movement and biometric data acquisition for sports applications," *IEEE Access*, vol. 5, pp. 6411–6420, 2017.

[29] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31715–31728, 2018.

[30] Y. J. Jeon and S. J. Kang, "Wearable sleepcare kit: Analysis and prevention of sleep apnea symptoms in real-time," *IEEE Access*, vol. 7, pp. 60634–60649, 2019.

[31] I. Raber, C. P. McCarthy, and R. W. Yeh, "Health insurance and mobile health devices: Opportunities and concerns," *Jama*, vol. 321, no. 18, pp. 1767–1768, 2019.

[32] H. Lee, J. Lee, and M. Shin, "Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," *Electronics*, vol. 8, no. 2, p. 192, Feb. 2019.

[33] M. Choi, G. Koo, M. Seo, and S. W. Kim, "Wearable device-based system to monitor a Driver's stress, fatigue, and drowsiness," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 3, pp. 634–645, Mar. 2018.

[34] T. Kundinger, P. K. Yalavarthi, A. Riener, P. Wintersberger, and C. Schartmüller, "Feasibility of smart wearables for driver drowsiness detection and its potential among different age groups," *Int. J. Pervas. Comput. Commun.*, vol. 16, no. 1, pp. 1–23, Jan. 2020.

[35] A. Tavakoli, M. Boukhechba, and A. Heydarian, "Personalized driver state profiles: A naturalistic data-driven study," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.* Cham, Switzerland: Springer, 2020, pp. 32–39.

[36] G. Li, Y. Yang, and X. Qu, "Deep learning approaches on pedestrian detection in hazy weather," *IEEE Trans. Ind. Electron.*, vol. 67, no. 10, pp. 8889–8899, Oct. 2020.

[37] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, "A deep learning based image enhancement approach for autonomous driving at night," *Knowl.-Based Syst.*, vol. 213, Feb. 2021, Art. no. 106617.

[38] M. Jeon, "Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods," in *Emotions and Affect in Human Factors and Human-Computer Interaction*. Amsterdam, The Netherlands: Elsevier, 2017, pp. 3–26.

[39] J. Wörle, B. Metz, C. Thiele, and G. Weller, "Detecting sleep in drivers during highly automated driving: The potential of physiological parameters," *IET Intell. Transp. Syst.*, vol. 13, no. 8, pp. 1241–1248, Aug. 2019.

[40] E. Nilsson, C. Ahlström, S. Barua, C. Fors, P. Lindén, B. Svanberg, S. Begum, M. U. Ahmed, and A. Anund, "Vehicle driver monitoring: Sleepiness and cognitive load," Swedish Nat. Road, Transp. Res. Inst. (VTI), Linköping, Sweden, Tech. Rep. 2013/0296-8.2, 2017.

[41] E. Pakdamanian, N. Namaky, S. Sheng, I. Kim, J. A. Coan, and L. Feng, "Toward minimum startle after take-over request: A preliminary study of physiological data," in *Proc. 12th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, Sep. 2020, pp. 27–29.

[42] N. Lyu, L. Xie, C. Wu, Q. Fu, and C. Deng, "Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in China," *Int. J. Environ. Res. Public Health*, vol. 14, no. 2, p. 203, Feb. 2017.

[43] W. Li, Y. Cui, Y. Ma, X. Chen, G. Li, G. Guo, and D. Cao, "A spontaneous driver emotion facial expression (DEFE) dataset for intelligent vehicles," 2020, *arXiv:2005.08626*. [Online]. Available: http://arxiv.org/abs/2005.08626

[44] R. A. Wynne, V. Beanland, and P. M. Salmon, "Systematic review of driving simulator validation studies," *Saf. Sci.*, vol. 117, pp. 138–151, Aug. 2019.

[45] A. Ziakopoulos, D. Tselentis, A. Kontaxi, and G. Yannis, "A critical overview of driver recording tools," *J. Saf. Res.*, vol. 72, pp. 203–212, Feb. 2020.

[46] D. Oswald, F. Sherratt, and S. Smith, "Handling the hawthorne effect: The challenges surrounding a participant observer," *Rev. Social Stud.*, vol. 1, no. 1, pp. 53–73, 2014.

[47] H. Wiberg, E. Nilsson, P. Lindén, B. Svanberg, and L. Poom, "Physiological responses related to moderate mental load during car driving in field conditions," *Biol. Psychol.*, vol. 108, pp. 115–125, May 2015.

[48] N. Li, T. Misu, and A. Miranda, "Driver Behavior event detection for manual annotation by clustering of the driver physiological signals," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 2583–2588.

[49] Y. Barnard, S. Innamaa, S. Koskinen, H. Gellerman, E. Svanberg, and H. Chen, "Methodology for field operational tests of automated vehicles," *Transp. Res. Procedia*, vol. 14, pp. 2188–2196, Jan. 2016.

[50] R. Ervin, J. Sayer, D. LeBlanc, S. Bogard, M. Mefford, M. Hagan, Z. Bareket, and C. Winkler, "Automotive collision avoidance system field operational test report: Methodology and results," Nat. Highway Traffic Saf. Admin., U.S. Dept. Transp., Washington, DC, USA, Tech. Rep. DOT HS 809 900, 2005.

[51] D. LeBlanc, "Road departure crash warning system field operational test: Methodology and results. Volume 1: Technical report," Univ. Michigan, Ann Arbor, Transp. Res. Inst., Ann Arbor, MI, USA, Tech. Rep., 2006.

[52] J. Sayer, D. LeBlanc, S. Bogard, D. Funkhouser, S. Bao, M. L. Buonarosa, and A. Blankespoor, "Integrated vehicle-based safety systems field operational test: Final program report," U.S. Dept. Transp. Res., Innov. Technol. Admin. ITS Joint Program Office, Washington, DC, USA, Tech. Rep. FHWA-JPO-11-150 Tech. Rep., 2011.

[53] I. van Schagen and F. Sagberg, "The potential benefits of naturalistic driving for road safety research: Theoretical and empirical considerations and challenges for the future," *Procedia Social Behav. Sci.*, vol. 48, pp. 692–701, Jan. 2012.

[54] R. Chen, K. D. Kusano, and H. C. Gabler, "Driver behavior during overtaking maneuvers from the 100-car naturalistic driving study," *Traffic Injury Prevention*, vol. 16, no. sup2, pp. S176–S181, Oct. 2015.

[55] J. Montgomery, K. D. Kusano, and H. C. Gabler, "Age and gender differences in time to collision at braking from the 100-car naturalistic driving study," *Traffic Injury Prevention*, vol. 15, no. sup1, pp. S15–S20, Sep. 2014.

[56] K. L. Campbell, "The SHRP 2 naturalistic driving study: Addressing driver performance and Behavior in traffic safety," *Tr News*, no. 282, pp. 30–35, Nov. 2012.

[57] T. A. Dingus, J. M. Hankey, J. F. Antin, S. E. Lee, L. Eichelberger, K. E. Stulce, D. McGraw, M. Perez, and L. Stowe, "Naturalistic driving study: Technical coordination and quality control," SHRP 2, Chicago, Il, USA, Tech. Rep. S2-S06-RW-1, 2015.

[58] A. Ghasemzadeh and M. M. Ahmed, "Drivers' lane-keeping ability in heavy rain: Preliminary investigation using SHRP 2 naturalistic driving study data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2663, no. 1, pp. 99–108, Jan. 2017.

[59] T. Victor, M. Dozza, J. Bärgman, C.-N. Boda, J. Engström, C. Flannagan, J. D. Lee, and G. Markkula, "Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk," Nat. Academy Sci., Washington, DC, USA, Tech. Rep. S2-S08A-RW-1, 2015.

[60] Y. Barnard, F. Utesch, N. van Nes, R. Eenink, and M. Baumann, "The study design of UDRIVE: The naturalistic driving study across europe for cars, trucks and scooters," *Eur. Transp. Res. Rev.*, vol. 8, no. 2, p. 14, Jun. 2016.

[61] J. Bärgman, N. van Nes, M. Christoph, R. Janssen, V. Heijne, O. Carsten, M. Dotzauer, F. Utesch, E. Svanberg, M. Pereira Cocron, and F. Forcolin, "The udrive dataset and key analysis results," Eur. Union's 7th Framework Programme Res., Technol. Develop. Demonstration, The Netherlands, Tech. Rep. 41.1, 2017.

[62] L. Guyonvarch, T. Hermitte, F. Duvivier, C. Val, and A. Guillaume, "Driving style indicator using UDRIVE NDS data," *Traffic Injury Prevention*, vol. 19, no. 1, pp. S189–S191, Feb. 2018.

[63] M. Zhu, X. Wang, A. Tarko, and S. Fang, "Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study," *Transp. Res. C, Emerg. Technol.*, vol. 93, pp. 425–445, Aug. 2018.

[64] G. Li, Y. Wang, F. Zhu, X. Sui, N. Wang, X. Qu, and P. Green, "Drivers' visual scanning behavior at signalized and unsignalized intersections: A naturalistic driving study in China," *J. Saf. Res.*, vol. 71, pp. 219–229, Dec. 2019.

[65] X. Wang, M. Yang, and D. Hurwitz, "Analysis of cut-in behavior based on naturalistic driving data," *Accident Anal. Prevention*, vol. 124, pp. 127–137, Mar. 2019.

[66] J. M. Hankey, "Canadian naturalistic driving study," Council Deputy Ministers Responsible Transp. Highway Saf., Ottawa, ON, Canada, Tech. Rep., 2014. [Online]. Available: https://vtechworks.lib.vt.edu/bitstream/handle/10919/53968/Hankey-2014.pdf?sequence

[67] N. Uchida, M. Kawakoshi, T. Tagawa, and T. Mochida, "An investigation of factors contributing to major crash types in japan based on naturalistic driving data," *IATSS Res.*, vol. 34, no. 1, pp. 22–30, Jul. 2010.

[68] S. C. Marshall, K. G. Wilson, M. Man-Son-Hing, I. Stiell, A. Smith, K. Weegar, Y. Kadulina, and F. J. Molnar, "The canadian safe driving study—Phase i pilot: Examining potential logistical barriers to the full cohort study," *Accident Anal. Prevention*, vol. 61, pp. 236–244, Dec. 2013.

[69] S. C. Marshall, M. Man-Son-Hing, M. Bédard, J. Charlton, S. Gagnon, I. Gélinas, S. Koppel, N. Korner-Bitensky, J. Langford, B. Mazer, A. Myers, G. Naglie, J. Polgar, M. M. Porter, M. Rapoport, H. Tuokko, B. Vrkljan, and A. Woolnough, "Protocol for candrive II/ozcandrive, a multicentre prospective older driver cohort study," *Accident Anal. Prevention*, vol. 61, pp. 245–252, Dec. 2013.

[70] J. Langford, J. L. Charlton, S. Koppel, A. Myers, H. Tuokko, S. Marshall, M. Man-Son-Hing, P. Darzins, M. Di Stefano, and W. Macdonald, "Findings from the candrive/ozcandrive study: Low mileage older drivers, crash risk and reduced fitness to drive," *Accident Anal. Prevention*, vol. 61, pp. 304–310, Dec. 2013.

[71] F. Knoefel, B. Wallace, R. Goubran, and S. Marshall, "Naturalistic driving: A framework and advances in using big data," *Geriatrics*, vol. 3, no. 2, p. 16, Mar. 2018.

[72] A. Williamson, R. Grzebieta, J. E. Eusebio, W. Y. Zheng, J. Wall, J. Charlton, M. Lenne, J. Haley, B. Barnes, A. Rakotonirainy, and J. Woolley, "The Australian naturalistic driving study: From beginnings to launch," in *Proc. Australas. Road Saf. Conf.*, 2015, pp. 1–7.

[73] R. Young, "Revised odds ratio estimates of secondary tasks: A reanalysis of the 100-car naturalistic driving study data," in *Proc. SAE Tech. Paper Ser.*, Apr. 2015, pp. 1–46. [Online]. Available: https://saemobilus.sae.org/content/2015-01-1387/

[74] G. S. Larue, S. Demmel, M. Khakzar, A. Rakotonirainy, and R. Grzebieta, "Visualising data of the Australian naturalistic driving study," in *Proc. 8th Int. Symp. Naturalistic Driving Res.*, 2019.

[75] L. Fridman, L. Ding, B. Jenik, and B. Reimer, "Arguing machines: Human supervision of black box AI systems that make life-critical decisions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1335–1343.

[76] N. Du, F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving," *Transp. Res. C, Emerg. Technol.*, vol. 112, pp. 78–87, Mar. 2020.

[77] L. F. Barrett, "The theory of constructed emotion: An active inference account of interoception and categorization," *Social Cognit. Affect. Neurosci.*, vol. 12, no. 1, pp. 1–23, 2017.

[78] K. Hoemann, Z. Khan, M. J. Feldman, C. Nielson, M. Devlin, J. Dy, L. F. Barrett, J. B. Wormwood, and K. S. Quigley, "Context-aware experience sampling reveals the scale of variation in affective experience," *Sci. Rep.*, vol. 10, no. 1, pp. 1–16, Dec. 2020.

[79] L. Fridman, B. Reimer, B. Mehler, and W. T. Freeman, "Cognitive load estimation in the wild," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–9.

[80] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5379–5390, Jun. 2019.

[81] C. Zhang, X. Wu, X. Zheng, and S. Yu, "Driver drowsiness detection using multi-channel second order blind identifications," *IEEE Access*, vol. 7, pp. 11829–11843, 2019.

[82] L. Li, B. Zhong, C. Hutmacher, Y. Liang, W. J. Horrey, and X. Xu, "Detection of driver manual distraction via image-based hand and ear recognition," *Accident Anal. Prevention*, vol. 137, Mar. 2020, Art. no. 105432.

[83] J. Zou, Z. Li, and P. Yan, "Automatic monitoring of Driver's physiological parameters based on microarray camera," in *Proc. IEEE Eurasia Conf. Biomed. Eng., Healthcare Sustainability (ECBIOS)*, May 2019, pp. 15–18.

[84] J. K. Lenneman and R. W. Backs, "Cardiac autonomic control during simulated driving with a concurrent verbal working memory task," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 51, no. 3, pp. 404–418, Jun. 2009.

[85] L. Shu, Y. Yu, W. Chen, H. Hua, Q. Li, J. Jin, and X. Xu, "Wearable emotion recognition using heart rate data from a smart bracelet," *Sensors*, vol. 20, no. 3, p. 718, Jan. 2020.

[86] J. Costa, F. Guimbretière, M. F. Jung, and T. Khalid Choudhury, "BoostMeUp: A smartwatch app to regulate emotions and improve cognitive performance," *GetMobile, Mobile Comput. Commun.*, vol. 24, no. 2, pp. 25–29, Sep. 2020.

[87] C. Chatzaki, M. Pediaditis, G. Vavoulas, and M. Tsiknakis, "Human daily activity and fall recognition using a smartphone's acceleration sensor," in *Proc. Int. Conf. Inf. Commun. Technol. Ageing Well e-Health*. Cham, Switzerland: Springer, 2016, pp. 100–118.

[88] N. Du, X. J. Yang, and F. Zhou, "Psychophysiological responses to takeover requests in conditionally automated driving," 2020, *arXiv:2010.03047*. [Online]. Available: http://arxiv.org/abs/2010.03047

[89] B.-L. Lee, B.-G. Lee, and W.-Y. Chung, "Standalone wearable driver drowsiness detection system in a smartwatch," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5444–5451, Jul. 2016.

[90] L. Liu, C. Karatas, H. Li, S. Tan, M. Gruteser, J. Yang, Y. Chen, and R. P. Martin, "Toward detection of unsafe driving with wearables," in *Proc. Workshop Wearable Syst. Appl. WearSys*, 2015, pp. 27–32.

[91] M. Jeon, "A systematic approach to using music for mitigating affective effects on driving performance and safety," in *Proc. ACM Conf. Ubiquitous Comput. UbiComp*, 2012, pp. 1127–1132.

[92] L. Zhang, J. Kang, H. Luo, and B. Zhong, "Drivers' physiological response and emotional evaluation in the noisy environment of the control cabin of a shield tunneling machine," *Appl. Acoust.*, vol. 138, pp. 1–8, Sep. 2018.

[93] M. Hassib, M. Braun, B. Pfleging, and F. Alt, "Detecting and influencing driver emotions using psycho-physiological sensors and ambient light," in *Proc. IFIP Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, 2019, pp. 721–742.

[94] O. Hahad, J. H. Prochaska, A. Daiber, and T. Muenzel, "Environmental noise-induced effects on stress hormones, oxidative stress, and vascular dysfunction: Key factors in the relationship between cerebrocardiovascular and psychological disorders," *Oxidative Med. Cellular Longevity*, vol. 2019, pp. 1–13, Nov. 2019.

[95] E. van Dyck, J. Six, E. Soyer, M. Denys, I. Bardijn, and M. Leman, "Adopting a music-to-heart rate alignment strategy to measure the impact of music and its tempo on human heart rate," *Musicae Scientiae*, vol. 21, no. 4, pp. 390–404, Dec. 2017.

[96] A. B. Ünal, D. de Waard, K. Epstude, and L. Steg, "Driving with music: Effects on arousal and performance," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 21, pp. 52–65, Nov. 2013.

[97] H. Wen, N. N. Sze, Q. Zeng, and S. Hu, "Effect of music listening on physiological condition, mental workload, and driving performance with consideration of driver temperament," *Int. J. Environ. Res. Public Health*, vol. 16, no. 15, p. 2766, Aug. 2019.

[98] Y. Zhu, Y. Wang, G. Li, and X. Guo, "Recognizing and releasing drivers' negative emotions by using music: Evidence from driver anger," in *Proc. Adjunct Proc. 8th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, Oct. 2016, pp. 173–178.

[99] C. Cohrdes, C. Wrzus, M. Wald-Fuhrmann, and M. Riediger, "'The sound of affect': Age differences in perceiving valence and arousal in music and their relation to music characteristics and momentary mood," *Musicae Scientiae*, vol. 24, no. 1, pp. 21–43, 2020.

[100] A. E. Krause, A. C. North, and L. Y. Hewitt, "Music-listening in everyday life: Devices and choice," *Psychol. Music*, vol. 43, no. 2, pp. 155–170, Mar. 2015.

[101] A. K. Dey, "Understanding and using context," *Pers. Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, 2001.

[102] A. Rakotonirainy and F. Maire, "Context-aware driving behaviour model," in *Proc. 19th Int. Tech. Conf. Enhanced Saf. Vehicles*. Washington, DC, USA: National Highway Traffic Safety Administration, 2005, pp. 1–6.

[103] M. Boukhechba and L. E. Barnes, "Swear: Sensing using wearables. generalized human crowdsensing on smartwatches," in *Proc. IEEE 11th Int. Conf. Appl. Hum. Factors Ergonom.*, 2020.

[104] M. Boukhechba. *Swear—Apps on Google Play*. Accessed: Nov. 20, 2020. [Online]. Available: https://play.google.com/store/apps/details?id=uva.swear

[105] *Geo-Location Apis*. Accessed: Nov. 20, 2020. [Online]. Available: https://cloud.google.com/maps-platform/

[106] *Current Weather and Forecast*. Accessed: Nov. 20, 2020. [Online]. Available: https://openweathermap.org/

[107] *Ffmpeg*. Accessed: Nov. 20, 2020. [Online]. Available: https://ffmpeg.org/
[108] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial Behavior analysis toolkit," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 59–66.
[109] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," 2018, *arXiv:1812.08008*. [Online]. Available: http://arxiv.org/abs/1812.08008
[110] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. E. Kaliouby, "AFFDEX SDK: A cross-platform real-time multi-face expression recognition toolkit," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, May 2016, pp. 3723–3726.
[111] *Pradipta. Song Lyrics & Knowledge*. Accessed: Nov. 20, 2020. [Online]. Available: https://github.com/geekpradd/PyLyrics
[112] Genius. *Genius: A Pythonic Implementation of Lyrics*. Accessed: Nov. 20, 2020. [Online]. Available: genius.com
[113] UVABRAINLAB. *Harmony Annotation Scheme and Details*. Accessed: Nov. 20, 2020. [Online]. Available: http://uvabrainlab.com/portfolio/reason-aware-annotation-for-contextualizing-driving-scenarios/
[114] W. Abdulla. (2017). *Mask R-CNN for Object Detection and Instance Segmentation on Keras and Tensorflow*. [Online]. Available: https://github.com/matterport/Mask_RCNN
[115] Á. Arcos-García, J. A. Álvarez-García, and L. M. Soria-Morillo, "Evaluation of deep neural networks for traffic sign detection systems," *Neurocomputing*, vol. 316, pp. 332–344, Nov. 2018.
[116] UVABRAINLAB. *Harmony Most Recent Map*. Accessed: Nov. 20, 2020. [Online]. Available: http://uvabrainlab.com/portfolio/driving-behavior/
[117] UVABRAINLAB. *Harmony Case Study*. Accessed: Nov. 20, 2020. [Online]. Available: https://osf.io/zextd/
[118] W. Wang, C. Liu, and D. Zhao, "How much data are enough? A statistical approach with case study on longitudinal driving behavior," *IEEE Trans. Intell. Veh.*, vol. 2, no. 2, pp. 85–98, Jun. 2017.
[119] N.-B. Heidenreich, A. Schindler, and S. Sperlich, "Bandwidth selection for kernel density estimation: A review of fully automatic selectors," *AStA Adv. Stat. Anal.*, vol. 97, no. 4, pp. 403–433, Oct. 2013.
[120] D. Barry and J. A. Hartigan, "A Bayesian analysis for change point problems," *J. Amer. Stat. Assoc.*, vol. 88, no. 421, pp. 309–319, Mar. 1993.
[121] W. R. Gilks, "M arkov chain m onte c arlo," *Encyclopedia Biostatistics*, vol. 4. Wiley, 2005.
[122] C. Erdman and J. W. Emerson, "Bcp: AnRPackage for performing a Bayesian analysis of change point problems," *J. Stat. Softw.*, vol. 23, no. 3, pp. 1–13, 2007.
[123] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition Emotion*, vol. 23, no. 2, pp. 209–237, Feb. 2009.

ARASH TAVAKOLI received the B.Sc. degree in civil engineering the Sharif University of Technology and the M.Sc. degree in civil engineering from Virginia Tech. He is currently pursuing the Ph.D. degree with the Engineering Systems and Environment Department and the Link Lab, University of Virginia. His research interests include the intersection of transportation engineering, computer science, and psychology.

SHASHWAT KUMAR received the B.Sc. degree in computer science from the Birla Institute of Technology and Science, India. He is currently pursuing the Ph.D. degree with the Engineering Systems and Environment Department and the Link Lab, University of Virginia. His research interests include ubiquitous computing, statistical learning, and interpretable machine learning.

XIANG GUO received the B.Sc. and M.Sc. degrees in transportation engineering from Beihang University, Beijing, China. He is currently pursuing the Ph.D. degree with the Engineering Systems and Environment Department and the Link Lab, University of Virginia. His research interests include traffic safety, human factors, human performance modeling, virtual reality and mixed reality.

VAHID BALALI received the B.Sc. and M.Sc. degrees in civil engineering from the University of Tehran, Iran, in 2008 and 2011, respectively, the M.Sc. degree in construction engineering and management from Virginia Tech, in 2012, and the Ph.D. degree in civil and environmental engineering from the University of Illinois at Urbana–Champaign, in 2015. From 2015 to 2016, he was a Senior BIM and Project Controls Specialist with STV, Chicago. Since 2016, he has been an Assistant Professor with the Civil Engineering and Construction Engineering Management Department, California State University, Long Beach. He is the author of one book and more than 40 articles. His research interests include visual data sensing and analytics for AEC industry, virtual design and construction for infrastructure asset management and interoperable system integration, and smart transportation planning for sustainable infrastructure decision-making. He was a recipient of the Early Academic Career Excellence Award by CSULB, in 2020, the Top 40 under 40 by the Consulting-Specifying Engineer, in 2017, and the Top Young Professional by the Engineering News Record (ENR), in 2016.

MEHDI BOUKHECHBA (Member, IEEE) is currently an Assistant Professor with the Engineering Systems and Environment Department and the Co-Director of the Sensing Systems for Health Lab. His primary research interests include ubiquitous computing, data science, behavioral modeling, and pervasive health technologies. In recent years, he has been developing novel ubiquitous sensing platforms to understand human behaviors in the wild. His research has been focused on designing new assessment and intervention methods for multiple health conditions, such as depression, anxiety, cancer, infectious disease, and traumatic brain injury.

ARSALAN HEYDARIAN received the B.Sc. and M.Sc. degrees in civil engineering from Virginia Tech and the M.Sc. degree in system engineering and the Ph.D. degree in civil engineering from the University of Southern California (USC). He is currently an Assistant Professor with the Department of Engineering Systems and Environment and the Link Lab, University of Virginia. His research interests include user-centered design, construction, and operation of intelligent infrastructure with the objective of enhancing sustainability, adaptability, and resilience future infrastructure systems. Specifically, his research can be divided into four main research streams: intelligent built environments, mobility and infrastructure design, smart transportation, and data-driven mixed reality.