Driver State Modeling through Multimodal Naturalistic Driving Data

---

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Arash Tavakoli

August 2022

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Arash Tavakoli

This Dissertation has been read and approved by the examing committee:

Advisor: Arsalan Heydarian

Advisor:

Committee Member: Mehdi Boukhechba

Committee Member: T. Donna Chen

Committee Member: Steven M. Boker

Committee Member: Brian L. Smith

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

August 2022

To ***mom***, for she taught me to be brave

To ***dad***, for he taught me to move even in the darkness

To ***Kimia***, for she taught me to be revolutionary

To ***Nazanin***, for she inspires me to become a better version of myself each and every day

To ***My Family in Iran***, for I am reminded of how lucky I am to have them every day

To ***Hamid, Kian, and Arman***, for their friendship is what pushes me forward every day

To ***Mustafa***, for his friendship is what reminds me of home everyday.

To ***Paul***, for he taught me not to take life too seriously

To ***Aref***, for he helped me see the light when I needed it

To ***Shashwat***, for he pushed me to think deeper

To ***Link Lab***, for it gave me a second family

To ***Professor Arsalan***, for he taught me to never give up

I would like to express my gratitude to the committee members and collaborators who supported me through this interesting journey. Thank you **Professor Heydarian** for always supporting me, having my back, and inspiring me to improve. You are truly an amazing advisor, a friend, and a person that I can always talk to. Every day I find myself incredibly lucky for emailing you in the first place. Thank you for taking a chance on me. It has been such a fun and joyful experience working with you, and I am forever grateful for knowing you. Thank You!

Thank you **Professor Boukhechba** for your support during the past few years, technical discussions, and hands-on collaborations. I always enjoy working with you, learning from you, and also playing soccer with you. Thank You!

Thank you **Professor Boker**. Words cannot describe how lucky I feel to have known you. Talking to you about my work and learning about the other aspects of it has been such an inspiration to me for the past few years. Thank You!

Thank you **Professor Chen** for your questions. Other than all of your support in the past years, the questions you asked really helped me see other angles of my work. Both in the Qualification and Proposal defense exams, you asked me questions that really made me think deeper about my research. Thank You!

Thank you **Professor Smith** for helping me better understand the next steps of my career. Talking to you has helped me both in better seeing the different angles of my work and also to better find a career path. I am very grateful to have known you. I would like to also mention that I am very grateful to be part of the ESE department. You and the rest of the faculty have made such an amazing department for me to be able to be creative and work towards what I really enjoy. Thank You!

Thank you **Professor Balali** for helping me and supporting my research for the past few years. Working and collaborating with you has been one of the very reasons that I enjoyed my journey as a Ph.D. student and I am forever indebted to your support. Thank You!

Lastly, I am grateful to the members of the **BRAIn Lab at UVA**, especially, Xiang, Mahsa, Alan, Siavash, Amir, and Beatrice. Without the help of these amazing people I would not have stood where I am right now. Thank You!

Contents

Chapter 1

Introduction and Summary of Dissertation

This chapter is mainly focused on providing an introduction and executive summary of the dissertation. This includes discussing the importance of personalization in shared-autonomy, multimodal sensing, and integrating conventional wearable devices into the current naturalistic driving study frameworks. We then provide a summary of all the following chapters by discussing the specific research questions and gaps addressed in each chapter, as well as the main findings. Sections 1.1 and 1.2 are

1.1 Shared-autonomy and Personalization of Autonomous Vehicles

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 (17; 38). 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 (17). 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 (15; 29).

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) (31). 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 (31). In order to increase the AV's safety, comfort, and reliability in different situations, the system should be individually tailored to each driver (8). 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"

(17).

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 (16). Context can be defined as any information that is relevant in defining a driving situation (16), 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 (4). 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 (16). 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) (16). 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 (22; 7; 3; 6; 16; 26). As a result, to develop personalized profiles for each specific driver, we need to accurately monitor the internal and external events and identify how changes within the environment may impact driver states and behaviors.

1.2 Importance of multimodal sensing

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) (30), the "European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment (UDRIVE)" (14), and the MIT Advanced Vehicle Technology Study (18). Such studies have provided significant insights into how different in-cabin and outdoor conditions may impact driving behaviors and states. For instance, (13) 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 (12), 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 (32). For instance, a driver might smile when being frustrated, leading to a misleading inference about the driver's state (1). 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) (9). As a result, the majority of the past driving studies that included physiological sensing were conducted in short-term controlled experimental studies (e.g., (39)). 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, (2) 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 here 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) (5). 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 without using real data (2).

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 (34). The application of these devices spans over a variety of fields such as mental health monitoring and interventions (11), physical health and activity monitoring and training (23; 20), sleep monitoring and intervention (21), and insurance and policy purposes (33). 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, (25) have used Microsoft armbands to detect driver's drowsiness in a virtual driving environment. (10) have used wearable devices for detecting driver's fatigue, stress, and abnormal conditions in a driving simulator environment. (24) have found that conventional wearable devices in a driving simulator environment can be used for driver's drowsiness de-

tection. 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 (27) and outside conditions through the outdoor videos (28). 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 analyze the external factors.

1.3 Dissertation Objectives

This dissertation is focused on designing and implementing a novel multimodal longitudinal naturalistic driving study, HARMONY by taking advantage of new technologies such as wearable devices and computer vision techniques. Based on the HARMONY dataset collected from 22 participants, this dissertation proposes different modeling schemes to detect (1) drivers' state, (2) behavior, and (3) changes in driver's state under the dynamic change in the external conditions. More specifically, this dissertation (1) leverages supervised learning to detect driver's state and behavior as well as driving environment attributes by using wearable data; (2) implements unsupervised learning to find the psychophysiological response of drivers' within different driving behaviors, which can be further used for guiding autonomous vehicles to take human-centered actions; (3) implements Bayesian Change Point detection method and linear mixed effect modeling together with computer vision algorithms to understand the level of association of certain infrastructural elements (e.g., intersections), and environmental attributes (e.g., presence of vulnerable road users) with the changes in drivers' heart rate; and (4) leverages state-space latent variable modeling framework to understand the changes in driver's state (i.e., stress level and workload) under the dynamic perturbations of external environment (i.e., changes in number of vehicles as a proxy for traffic density and hand movement as a proxy for task demands). A summary of each chapter is provided below.

1.3.1 Chapter 2: Literature Review on Naturalistic Driving Studies

The first part of this dissertation is mainly concerned with developing and implementing a novel naturalistic driving study (NDS) framework, thus it is required to first review the past work in this arena of research. This chapter provides a summary of previous NDS, as well as highlights how we can overcome the existing gaps through integrating off-the-shelf wearable and human sensing techniques. This chapter specifically addresses the following research question:

Research Question: What are the gaps in the current naturalistic driving study frameworks?

1.3.2 Chapter 3: HARMONY, A Human-Centered Multimodal Driving Study in the Wild

The main objective of this chapter 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. Chapter 3 summarizes HARMONY's goals, framework design, data collection and analysis, validation, and ongoing and future research efforts. The result of this chapteris published in (37). This chapter specifically addresses the following research question:

Research Question 1: What are the main considerations in designing a multimodal human-centered driving study framework in the wild?

1.3.3 Chapter 4: Multimodal Driver State and Behavior Detection

In previous chapters, we have pointed out the (1) potential shortcomings and concerns when solely relying on video streams for driver state and behavior detection; and (2) the utility of novel off-the-shelf wearable devices for driver state and behavior detection. To address these gaps, this chapter first focuses on using supervised learning for classifying drivers' state, behaviors, and environmental attributes using passive sensing. The outcome of this project is to detect certain driver activities such as using phone through supervised methods. While supervised learning provides valuable tools for classification and detection of drivers' state, it suffers from the manual annotation burden for building ground truth. Therefore, this chapter will also consider unsupervised methods to cluster drivers' state and behaviors and analyze drivers' state within each specific behavior. The outcome of this project is to detect

certain driver state such as having abnormal heart rate through unsupervised methods.

1.3.3.1 Project 1: Driver state, behavior, and environmental attributes classification using supervised techniques

In this project we are focused on developing models that can detect drivers' state by soley relying on wearable device. Integrating driver, in-cabin, and outside environment's contextual cues into the vehicle's decision making is the centerpiece of semi-automated vehicle safety. Multiple systems have been developed for providing context to the vehicle, which often rely on video streams capturing drivers' physical and environmental states. While video streams are a rich source of information, their ability in providing context can be challenging in certain situations, such as low illuminance environments (e.g., night driving), and they are highly privacy-intrusive. Data collected through passive sensing smart smartwatches are leveraged for classifying elements of driving context. Specifically, through using the data collected from 15 participants of HARMONY, and by using multiple machine learning algorithms such as random forest, driver's activities (e.g., using phone and eating), outside events (e.g., passing intersection and changing lane), and outside road attributes (e.g., driving in a city versus a highway) are classified with an average F1 score of 94.55, 98.27, and 97.86 percent respectively, through 10-fold cross-validation. The results show the applicability of multimodal data retrieved through smart wearable devices in providing context in real-world driving scenarios and pave the way for a better shared autonomy and privacy-aware driving data-collection, analysis, and feedback for future autonomous vehicles that does not rely on camera streams. The results of this project is published in the 2021 IEEE Intelligent Vehicle Symposium (36). This project specifically addresses the following questions:

Research Question 1: Do features retrieved through passive sensing techniques can provide contextual awareness to the vehicles by revealing different internal and external contextual cues through classification algorithms?

Research Question 2: What is the utility of multimodality in passive sensing for classification purposes? More specifically, to what extend the multimodal data help with enhancing the accuracy of classification?

1.3.3.2 Project 2: Driver state, behavior, and environmental attributes classification using unsupervised techniques

Naturalistic driving data (NDD) can help understand drivers' reactions to each driving scenario and then be used as a metric for taking further autonomous actions that are aligned with the drivers' comfort, feeling, and safety levels. While in the previous chapter, we have seen the application of supervised learning in driver state and behavior classification, the massive amount of data collected through NDS can become problematic for annotation and

data preparation. Needless to say that because of the existing noise in naturalistic conditions, a higher amount of data should be annotated when compared to similar situations in on-road controlled studies or driving simulators. In other words, NDD requires a high amount of manual labor to label certain drivers' states and behavioral patterns. Unsupervised analysis of NDD can be used to automatically detect different patterns from the driver and vehicle data. A framework is proposed to understand a driver's reactions towards different driving patterns unsupervised. This framework decomposes a driving scenario by using Bayesian Change Point detection methods. Similar to previous research, the Latent Dirichlet Allocation method is applied on both driver state and behavior data. Two case studies based on the HARMONY framework are presented in which vehicles were equipped to collect exterior, interior, and driver behavioral data. Four patterns of driving behaviors (i.e., harsh brake, normal brake, curved driving, and highway driving), as well as two patterns of driver's heart rate (HR) (i.e., baseline vs. abnormal high HR), and gaze entropy (i.e., low versus high), were detected in the case study data. Drivers' HR had a higher fraction of abnormal patterns during harsh brakes, accelerating, and curved driving, pointing to the need for a more conservative driving style to better fit the driver. Additionally, free-flow driving with close to zero accelerations on the highway was accompanied by more fraction of normal HR as well as a lower gaze entropy pattern pointing to route selection decisions that include more highway segments to keep the driver calmer with a lower workload. The results of this chapter are under review in Accident Analysis and Prevention. This project specifically addresses the following research questions:

Research Question 1: Are driver's behavior and state patterns detectable through unsupervised methods?

Research Question 2; What are the differences across participants in their heart rate and gaze patterns within different contextual settings?

1.3.4 Chapter 5: How Does the Driving Context Affect Driver's State?

While previous projects are guided towards detecting drivers' state using multimodal data, it is insightful to consider how the state itself changes while interacting with the dynamic changes of contextual setting. This chapter focuses on identifying the interaction between changes in the context and drivers' state. To this end, this chapter proposes three projects as follows.

1.3.4.1 Project 1: Driver's HR as a feedback to the environmental events

This project is mainly focused on understanding the interaction between environmental attributes and changes in drivers' heart rate (HR). Drivers' HR has been previously shown

to be correlated with stress level, anxiety, and negative emotions. The collected data in HARMONY is employed to show the utility of driver's HR as feedback to environmental changes. This is first done through a visual inspection of videos that are accompanied by abrupt increases in drivers' HR. The abrupt increases are detected through Bayesian Change Point Detection (BCP). Fusing the information from the in-cabin and outside videos with the HR changepoint locations from 15 participants results in understanding the reasons behind abrupt increases in HR. The preliminary analysis shows that events such as passing through intersections, closely following a lead vehicle, passing by certain road users, and objects are shown to be associated with abrupt increases in driver's HR, possibly causing stress in the driver. This project specifically answers the following research questions:

Research Question 1: Is a driver's heart rate as retrieved through conventional wearable devices indicative of external contextual changes?

1.3.4.2 Project 2: Understanding the Differences Across Environmental Attributes on How They Affect the Driver's State

In this study, by using a naturalistic driving study database, we analyze the changes in the driving scene, including road objects and the dynamical relationship between the ego vehicle and the lead vehicle with respect to changes in drivers' psychophysiological metrics (i.e., HR and facial expressions). We find that different road objects might be associated with varying levels of increase in drivers' HR as well as different proportions of negative facial emotions detected through computer vision. Our results indicate that larger vehicles on the road, such as trucks and buses, are associated with the highest amount of increase in drivers' HR as well as negative emotions. Additionally, we provide evidence that shorter distances to the lead vehicle in naturalistic driving, as well as the higher standard deviation in the distance, might be associated with a higher number of abrupt increases in drivers' HR, showing a possible increase in stress level. Lastly, our results indicate more positive emotions, more facial engagement, and a lower number of abrupt changes in HR at a higher speed of driving. This research lays the ground for designing human-centered vehicles, urban environments, and services that can understand the level to which each road object and environment might affect drivers' and passengers' well-being. This project specifically answers the following research questions.

Research Question 1: Do different perturbations differ in their effect on drivers' state changes? **Research Question 2**: Do surrounding vehicles and especially the lead vehicle plays a role in drivers' state changes?

1.3.4.3 Project 3: Modeling changes in drivers' state under the external contextual perturbations using latent variable state-space modeling framework

Driving is a multidimensional task happening in a driving context that includes the driver's internal states as well as external environmental conditions. Driver's internal states, including stress levels, and workload, interact with each other and with the external context temporally. Although multiple studies analyzed the relationship between the driving environment, workload, and driver's stress in a paired fashion, no study has analyzed all of these factors together while considering their possible interaction. Additionally, because matters such as stress level, and workloads are not measurable variables, they need to be modeled through a latent variable modeling scheme. This chapter proposes using latent-variable state-space modeling framework for driver state analysis that takes into account the interaction between the two constructs (i.e., emotion and cognition), and the holistic effect of the context on the two variables. By using latent-variable state-space models, drivers' workload and stress levels are modeled as latent variables estimated through multimodal human sensing data, under the perturbations of the environment. Through a case study of multimodal driving data collected from 11 participants of HARMONY, we show that a better maximum log-likelihood fit is retrieved in a model that considers two latent variables for driver's psychophysiological state with a covariance between latent variable as compared with a model with one latent variable. we then show that external contextual elements such as scene complexity and secondary task demands can affect driver's stress levels and workload. Additionally, the utility of state-space models in analyzing the possible lag between the two constructs of stress level and workload is discussed. This chapter, in summary, uses latent variable state space methods to demonstrate how we can identify driver states and responses to the environmental conditions by fusing psychophysiological information with features extracted from outside video streams. This project specifically addresses the following research questions:

Research Question 1: Do two separate latent variables for driver's workload and stress levels provide a better description for driver's psychophysiological state in the wild?

Research Question 2: How does the historical latent state of the driver affect his/her current state (i.e., stress and workload)?

Research Question 3: How does the association between latent constructs change throughout the driving scenario?

Research Question 4: How does the external context (i.e., task demands and traffic density) affect the driver's internal contextual state (i.e., stress level and workload)?

Bibliography

[1] Abdic, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B. and Schuller, B. (2016). Driver frustration detection from audio and video in the wild, *Springer*, Vol. 9904, Springer, p. 237.

[2] Al-Sultan, S., Al-Bayatti, A. H. and Zedan, H. (2013). Context-aware driver behavior detection system in intelligent transportation systems, *IEEE transactions on vehicular technology* **62**(9): 4264–4275.

[3] Alizadeh, V. and Dehzangi, O. (2016). The impact of secondary tasks on drivers during naturalistic driving: Analysis of eeg dynamics, *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2493–2499.

[4] Azari, B., Westlin, C., Satpute, A. B., Hutchinson, B., Kragel, P. A., Hoemann, K., Khan, Z., Wormwood, J. B., Quigley, K., Erdogmus, D. et al. (2020). Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience.

[5] Baig, M. Z. and Kavakli, M. (2019). A survey on psycho-physiological analysis & measurement methods in multimodal systems, *Multimodal Technologies and Interaction* **3**(2): 37.

[6] Brodsky, W. (2018). A performance analysis of in-car music engagement as an indication of driver distraction and risk, *Transportation research part F: traffic psychology and behaviour* **55**: 210–218.

[7] Brodsky, W. and Slor, Z. (2013). Background music as a risk factor for distraction among young-novice drivers, *Accident Analysis & Prevention* **59**: 382–393.

[8] Butakov, V. A. and Ioannou, P. (2014). Personalized driver/vehicle lane change models for adas, *IEEE Transactions on Vehicular Technology* **64**(10): 4422–4431.

[9] Carsten, O., Kircher, K. and Jamson, S. (2013). Vehicle-based studies of driving in the real world: The hard truth?, *Accident Analysis & Prevention* **58**: 162–174.

[10] Choi, M., Koo, G., Seo, M. and Kim, S. W. (2017). Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness, *IEEE Transactions on Instrumentation and Measurement* **67**(3): 634–645.

[11] Coughlin, S. S. and Stewart, J. (2016). Use of consumer wearable devices to promote physical activity: a review of health intervention studies, *Journal of environment and health sciences* **2**(6).

[12] Das, A., Ghasemzadeh, A. and Ahmed, M. M. (2019). Analyzing the effect of fog weather conditions on driver lane-keeping performance using the shrp2 naturalistic driving study data, *Journal of safety research* **68**: 71–80.

[13] Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M. and Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data, *Proceedings of the National Academy of Sciences* **113**(10): 2636–2641.

[14] Eenink, R., Barnard, Y., Baumann, M., Augros, X. and Utesch, F. (2014). Udrive: the european naturalistic driving study, *Proceedings of Transport Research Arena*, IFSTTAR.

[15] Feng, L., Wiltsche, C., Humphrey, L. and Topcu, U. (2016). Synthesis of human-in-the-loop control protocols for autonomous systems, *IEEE Transactions on Automation Science and Engineering* **13**(2): 450–462.

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

[17] Fridman, L. (2018). Human-centered autonomous vehicle systems: Principles of effective shared autonomy, *arXiv preprint arXiv:1810.01835* .

[18] Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Patsekin, A., Kindelsberger, J., Ding, L. et al. (2019). Mit advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation, *IEEE Access* **7**: 102021–102038.

[19] Hallmark, S. L., Tyner, S., Oneyear, N., Carney, C. and McGehee, D. (2015). Evaluation of driving behavior on rural 2-lane curves using the shrp 2 naturalistic driving study data, *Journal of safety research* **54**: 17–e1.

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

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

[22] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., Ramsey, D. J. et al. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

[23] Kos, M. and Kramberger, I. (2017). A wearable device and system for movement and biometric data acquisition for sports applications, *IEEE Access* **5**: 6411–6420.

[24] Kundinger, T., Yalavarthi, P. K., Riener, A., Wintersberger, P. and Schartmüller, C. (2020). Feasibility of smart wearables for driver drowsiness detection and its potential among different age groups, *International Journal of Pervasive Computing and Communications* .

[25] Lee, H., Lee, J. and Shin, M. (2019). Using wearable ecg/ppg sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots, *Electronics* **8**(2): 192.

[26] Li, G., Lai, W., Sui, X., Li, X., Qu, X., Zhang, T. and Li, Y. (2020). Influence of traffic congestion on driver behavior in post-congestion driving, *Accident Analysis & Prevention* **141**: 105508.

[27] Li, G., Yang, Y. and Qu, X. (2019). Deep learning approaches on pedestrian detection in hazy weather, *IEEE Transactions on Industrial Electronics* .

[28] Li, G., Yang, Y., Qu, X., Cao, D. and Li, K. (2020). A deep learning based image enhancement approach for autonomous driving at night, *Knowledge-Based Systems* p. 106617.

[29] Li, G., Yang, Y., Zhang, T., Qu, X., Cao, D., Cheng, B. and Li, K. (2021). Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios, *Transportation Research Part C: Emerging Technologies* **122**: 102820.

[30] Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J. and Goodman, M. (2005). An overview of the 100-car naturalistic study and findings, *National Highway Traffic Safety Administration, Paper* **5**: 0400.

[31] Park, S. Y., Moore, D. J. and Sirkin, D. (2020). What a driver wants: User preferences in semi-autonomous vehicle decision-making, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

[32] Prabhakar, G., Mukhopadhyay, A., Murthy, L., Modiksha, M., Sachin, D. and Biswas, P. (2020). Cognitive load estimation using ocular parameters in automotive, *Transportation Engineering* **2**: 100008.

[33] Raber, I., McCarthy, C. P. and Yeh, R. W. (2019). Health insurance and mobile health devices: Opportunities and concerns, *Jama* **321**(18): 1767–1768.

[34] Tankovska, H. (2020). Global connected wearable devices 2016-2022.
URL: *https://www.statista.com/statistics/487291/global-connected-wearable-devices/*

[35] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2020). Personalized driver state profiles: A naturalistic data-driven study, *International Conference on Applied Human Factors and Ergonomics*, Springer, pp. 32–39.

[36] Tavakoli, A., Kumar, S., Boukhechba, M. and Heydarian, A. (2021). Driver state and behavior detection through smart wearables, *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE.
URL: *https://doi.org/10.1109/iv48863.2021.9575431*

[37] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* **9**: 23956–23978.

[38] Terken, J. and Pfleging, B. (2020). Toward shared control between automated vehicles and users, *Automotive Innovation* **3**(1): 53–61.

[39] Zepf, S., Dittrich, M., Hernandez, J. and Schmitt, A. (2019). Towards empathetic car interfaces: Emotional triggers while driving, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

Chapter 2

Literature Review on Naturalistic Driving Studies

2.1 Summary

In order to understand the limitations of current NDS, we first review the different categories of previous driver-in-the-loop studies; specifically, section *2.2* provides an overview of conducted studies through driving simulators, section *2.3* focuses on on-road controlled studies and Field Operational Tests (FOT), and section *2.4* evaluates the existing NDS. This chapter is

2.2 Driving Simulator

Studies performed in driving simulators are beneficial in understanding causal analysis as different factors can be controlled (29). 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 (65), cognitive distraction when the lead vehicle abruptly breaks (49), response to automation takeover when the take over request is offered through different sources (51), and driver's performance and possible cognitive distraction when being exposed to different traffic signs (44). Driving simulators are safe tools for conducting driving experiments in different environmental conditions, such as crash events, or evaluating drivers' emotional states (42). 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) (51). 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 (66).

2.3 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 (74). 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) (10). 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 (50).

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, (63), (41) 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, (2) 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. (41) 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, (69) 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 stud-

ies 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 (10; 4). These studies are generally closer to the real-world unconstrained driving situation and have higher external validity compared to on-road controlled study (10). 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 (19; 36; 53).

2.4 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 (10; 59; 18; 48). NDS has high external validity as it is performed in real-life scenarios (10). 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 (18). In contract to on-road controlled studies, there are no observers/experimenters in NDS, and participants are asked to perform their daily routine activities (20; 59; 18). Additionally, NDS is always accompanied by uncontrollable real-life noise (10). 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 (48). 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 (31), overtaking maneuvers (12), and even differences in driving behaviors among different age and sex groups (47).

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) (9; 15). 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, (23) 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 (60). 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 (60).

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 (5). 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 (18). 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. (3; 24).

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" (72) 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, (39) 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 (72). Other studies have also evaluated more specific behaviors, such as cut-in behavior (61). 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 (26). 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 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 (26). 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 (56) 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 (46; 45; 34). 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, (32) 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 (34). 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 (64). 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 (64). Using ANDS, (68) assessed the patterns of secondary task engagement during everyday driving scenarios, and (35) 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 (20). 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 (20)). This study introduces a framework aimed at enhancing semi-autonomous vehicle safety and reliability by building better shared-automated systems (21). A summary of the previous studies can be viewed in Table 2.1.

2.5 Current Gaps and Challenges in NDS

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

Table 2.1: 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 ©2021 IEEE

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**	**Mobileeye**	**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	✓	✓	-	-	-	-	-	-	✓	-	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-

.

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 (17). 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 (1; 6). For instance, in driving research, (1) 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 (6). 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 (6; 30). 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 (22), possible distraction (67), and drowsiness (70). For instance, (22) 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 (67; 40). 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 (1)) 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 (52). Although a recent study has attempted to estimate some of these features from video streams (e.g., estimating HR from video (75)), 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, (38) 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 (54), provide feedback and interventions (14) 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 (11). 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 (16). 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) (37; 43), and emotions and cognitive load (17; 16; 17).

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 (28; 71; 27; 25). For instance, a recent review suggests traffic noise can be a cause for a psychological disorder, and mental stress (25). 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) (27).

Additionally, research suggests music can have a significant impact on a person's behaviors, emotions, and physiological states (28). 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 (8; 7). 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 (28). 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(58). However, in a car-following study, the physiological measurement did not differ in conditions with and without music, even with different volume levels(57). Other music features, like the music genre, were also found to affect driving performance (62). Similarly,(73; 13; 33) 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.

Bibliography

[1] Abdic, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B. and Schuller, B. (2016). Driver frustration detection from audio and video in the wild, *Springer*, Vol. 9904,

Springer, p. 237.

[2] Alizadeh, V. and Dehzangi, O. (2016). The impact of secondary tasks on drivers during naturalistic driving: Analysis of eeg dynamics, *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2493–2499.

[3] Bärgman, J., van Nes, N., Christoph, M., Jansen, R., Heijne, V., Carsten, O. and Fox, C. (2017). The udrive dataset and key analysis results, *Technical report*, Tech. rep. 2017.: 10. 26323/UDRIVE.

[4] Barnard, Y., Innamaa, S., Koskinen, S., Gellerman, H., Svanberg, E. and Chen, H. (2016). Methodology for field operational tests of automated vehicles, *Transportation research procedia* **14**: 2188–2196.

[5] Barnard, Y., Utesch, F., van Nes, N., Eenink, R. and Baumann, M. (2016). The study design of udrive: the naturalistic driving study across europe for cars, trucks and scooters, *European Transport Research Review* **8**(2): 14.

[6] Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization, *Social cognitive and affective neuroscience* **12**(1): 1–23.

[7] Brodsky, W. (2018). A performance analysis of in-car music engagement as an indication of driver distraction and risk, *Transportation research part F: traffic psychology and behaviour* **55**: 210–218.

[8] Brodsky, W. and Slor, Z. (2013). Background music as a risk factor for distraction among young-novice drivers, *Accident Analysis & Prevention* **59**: 382–393.

[9] Campbell, K. L. (2012). The shrp 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety, *Tr News* (282).

[10] Carsten, O., Kircher, K. and Jamson, S. (2013). Vehicle-based studies of driving in the real world: The hard truth?, *Accident Analysis & Prevention* **58**: 162–174.

[11] Chatzaki, C., Pediaditis, M., Vavoulas, G. and Tsiknakis, M. (2016). Human daily activity and fall recognition using a smartphone's acceleration sensor, *International Conference on Information and Communication Technologies for Ageing Well and e-Health*, Springer, pp. 100–118.

[12] Chen, R., Kusano, K. D. and Gabler, H. C. (2015). Driver behavior during overtaking maneuvers from the 100-car naturalistic driving study, *Traffic injury prevention* **16**(sup2): S176–S181.

[13] Cohrdes, C., Wrzus, C., Wald-Fuhrmann, M. and Riediger, M. (2020). "the sound of affect": Age differences in perceiving valence and arousal in music and their relation to music characteristics and momentary mood, *Musicae Scientiae* **24**(1): 21–43.

[14] Costa, J., Guimbretière, F., Jung, M. F. and Khalid Choudhury, T. (2020). Boostmeup: A smartwatch app to regulate emotions and improve cognitive performance, *GetMobile: Mobile Computing and Communications* **24**(2): 25–29.

[15] Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K. E., McGraw, D., Perez, M. and Stowe, L. (2015). *Naturalistic driving study: Technical coordination and quality control*, number SHRP 2 Report S2-S06-RW-1.

[16] Du, N., Yang, X. J. and Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving, *arXiv preprint arXiv:2010.03047* .

[17] Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K. and Yang, X. J. (2020). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving, *Transportation research part C: emerging technologies* **112**: 78–87.

[18] Eenink, R., Barnard, Y., Baumann, M., Augros, X. and Utesch, F. (2014). Udrive: the european naturalistic driving study, *Proceedings of Transport Research Arena*, IFSTTAR.

[19] Ervin, R., Sayer, J., LeBlanc, D., Bogard, S., Mefford, M., Hagan, M., Bareket, Z. and Winkler, C. (2005). Automotive collision avoidance system field operational test report: methodology and results, *Technical report*.

[20] Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Patsekin, A., Kindelsberger, J., Ding, L. et al. (2019). Mit advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation, *IEEE Access* **7**: 102021–102038.

[21] Fridman, L., Ding, L., Jenik, B. and Reimer, B. (2019). Arguing machines: Human supervision of black box ai systems that make life-critical decisions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0.

[22] Fridman, L., Reimer, B., Mehler, B. and Freeman, W. T. (2018). Cognitive load estimation in the wild, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, p. 652.

[23] Ghasemzadeh, A. and Ahmed, M. M. (2017). Drivers' lane-keeping ability in heavy rain: preliminary investigation using shrp 2 naturalistic driving study data, *Transportation research record* **2663**(1): 99–108.

[24] Guyonvarch, L., Hermitte, T., Duvivier, F., Val, C. and Guillaume, A. (2018). Driving style indicator using udrive nds data, *Traffic injury prevention* **19**(sup1): S189–S191.

[25] Hahad, O., Prochaska, J. H., Daiber, A. and Muenzel, T. (2019). Environmental noise-induced effects on stress hormones, oxidative stress, and vascular dysfunction: key factors in the relationship between cerebrocardiovascular and psychological disorders, *Oxidative Medicine and Cellular Longevity* **2019**.

[26] Hankey, J. M. et al. (2014). Canadian naturalistic driving study.

[27] Hassib, M., Braun, M., Pfleging, B. and Alt, F. (2019). Detecting and influencing driver emotions using psycho-physiological sensors and ambient light, *IFIP Conference on Human-Computer Interaction*, Springer, pp. 721–742.

[28] Jeon, M. (2012). A systematic approach to using music for mitigating affective effects on driving performance and safety, *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 1127–1132.

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

[30] Katie, H., Zulqarnain, K., Feldman, M. J., Catie, N., Devlin, M., Dy, J., Barrett, L. F., Wormwood, J. B. and Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience, *Scientific Reports (Nature Publisher Group)* **10**(1).

[31] Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., Ramsey, D. J. et al. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

[32] Knoefel, F., Wallace, B., Goubran, R. and Marshall, S. (2018). Naturalistic driving: a framework and advances in using big data, *Geriatrics* **3**(2): 16.

[33] Krause, A. E., North, A. C. and Hewitt, L. Y. (2015). Music-listening in everyday life: Devices and choice, *Psychology of music* **43**(2): 155–170.

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

[35] Larue, G. S., Demmel, S., Khakzar, M., Rakotonirainy, A. and Grzebieta, R. (2019). Visualising data of the australian naturalistic driving study.

[36] LeBlanc, D. (2006). Road departure crash warning system field operational test: methodology and results. volume 1: technical report, *Technical report*, University of Michigan, Ann Arbor, Transportation Research Institute.

[37] Lee, B.-L., Lee, B.-G. and Chung, W.-Y. (2016). Standalone wearable driver drowsiness detection system in a smartwatch, *IEEE Sensors journal* **16**(13): 5444–5451.

[38] Lenneman, J. K. and Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task, *Human factors* **51**(3): 404–418.

[39] Li, G., Wang, Y., Zhu, F., Sui, X., Wang, N., Qu, X. and Green, P. (2019). Drivers' visual scanning behavior at signalized and unsignalized intersections: A naturalistic driving study in china, *Journal of safety research* **71**: 219–229.

[40] Li, L., Zhong, B., Hutmacher Jr, C., Liang, Y., Horrey, W. J. and Xu, X. (2020). Detection of driver manual distraction via image-based hand and ear recognition, *Accident Analysis & Prevention* **137**: 105432.

[41] Li, N., Misu, T. and Miranda, A. (2016). Driver behavior event detection for manual annotation by clustering of the driver physiological signals, *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2583–2588.

[42] Li, W., Cui, Y., Ma, Y., Chen, X., Li, G., Guo, G. and Cao, D. (2020). A spontaneous driver emotion facial expression (defe) dataset for intelligent vehicles, *arXiv preprint arXiv:2005.08626* .

[43] Liu, L., Karatas, C., Li, H., Tan, S., Gruteser, M., Yang, J., Chen, Y. and Martin, R. P. (2015). Toward detection of unsafe driving with wearables, *Proceedings of the 2015 workshop on Wearable Systems and Applications*, pp. 27–32.

[44] Lyu, N., Xie, L., Wu, C., Fu, Q. and Deng, C. (2017). Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments:

A case study of the highways in china, *International journal of environmental research and public health* **14**(2): 203.

[45] Marshall, S. C., Man-Son-Hing, M., Bedard, M., Charlton, J., Gagnon, S., Gelinas, I., Koppel, S., Korner-Bitensky, N., Langford, J., Mazer, B. et al. (2013). Protocol for candrive ii/ozcandrive, a multicentre prospective older driver cohort study, *Accident Analysis & Prevention* **61**: 245–252.

[46] Marshall, S. C., Wilson, K. G., Man-Son-Hing, M., Stiell, I., Smith, A., Weegar, K., Kadulina, Y. and Molnar, F. J. (2013). The canadian safe driving study—phase i pilot: Examining potential logistical barriers to the full cohort study, *Accident Analysis & Prevention* **61**: 236–244.

[47] Montgomery, J., Kusano, K. D. and Gabler, H. C. (2014). Age and gender differences in time to collision at braking from the 100-car naturalistic driving study, *Traffic injury prevention* **15**(sup1): S15–S20.

[48] Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J. and Goodman, M. (2005). An overview of the 100-car naturalistic study and findings, *National Highway Traffic Safety Administration, Paper* **5**: 0400.

[49] Nilsson, E., Ahlström, C., Barua, S., Fors, C., Lindén, P., Svanberg, B., Begum, S., Ahmed, M. U. and Anund, A. (2017). Vehicle driver monitoring: sleepiness and cognitive load.

[50] Oswald, D., Sherratt, F. and Smith, S. (2014). Handling the hawthorne effect: The challenges surrounding a participant observer, *Review of social studies* **1**(1): 53–73.

[51] Pakdamanian, E., Namaky, N., Sheng, S., Kim, I., Coan, J. A. and Feng, L. (2020). Toward minimum startle after take-over request: A preliminary study of physiological data, *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 27–29.

[52] Prabhakar, G., Mukhopadhyay, A., Murthy, L., Modiksha, M., Sachin, D. and Biswas, P. (2020). Cognitive load estimation using ocular parameters in automotive, *Transportation Engineering* **2**: 100008.

[53] Sayer, J., LeBlanc, D., Bogard, S., Funkhouser, D., Bao, S., Buonarosa, M. L., Blankespoor, A. et al. (2011). Integrated vehicle-based safety systems field operational test: Final program report, *Technical report*, United States. Joint Program Office for Intelligent Transportation Systems.

[54] Shu, L., Yu, Y., Chen, W., Hua, H., Li, Q., Jin, J. and Xu, X. (2020). Wearable emotion recognition using heart rate data from a smart bracelet, *Sensors* **20**(3): 718.

[55] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* **9**: 23956–23978.

[56] Uchida, N., Kawakoshi, M., Tagawa, T. and Mochida, T. (2010). An investigation of factors contributing to major crash types in japan based on naturalistic driving data, *IATSS research* **34**(1): 22–30.

[57] Ünal, A. B., de Waard, D., Epstude, K. and Steg, L. (2013). Driving with music: Effects on arousal and performance, *Transportation research part F: traffic psychology and behaviour* **21**: 52–65.

[58] Van Dyck, E., Six, J., Soyer, E., Denys, M., Bardijn, I. and Leman, M. (2017). Adopting a music-to-heart rate alignment strategy to measure the impact of music and its tempo on human heart rate, *Musicae Scientiae* **21**(4): 390–404.

[59] van Schagen, I. and Sagberg, F. (2012). The potential benefits of naturalistic driving for road safety research: Theoretical and empirical considerations and challenges for the future, *Procedia-social and behavioral sciences* **48**: 692–701.

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

[61] Wang, X., Yang, M. and Hurwitz, D. (2019). Analysis of cut-in behavior based on naturalistic driving data, *Accident Analysis & Prevention* **124**: 127–137.

[62] Wen, H., Sze, N., Zeng, Q. and Hu, S. (2019). Effect of music listening on physiological condition, mental workload, and driving performance with consideration of driver temperament, *International journal of environmental research and public health* **16**(15): 2766.

[63] Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B. and Poom, L. (2015). Physiological responses related to moderate mental load during car driving in field conditions, *Biological psychology* **108**: 115–125.

[64] Williamson, A., Grzebieta, R., Eusebio, J., Zheng, W. Y., Wall, J., Charlton, J. L., Lenné, M., Haley, J., Barnes, B., Rakotonirainy, A. et al. (2015). The australian naturalistic driving study: From beginnings to launch, *Proceedings of the 2015 Australasian Road Safety Conference*.

[65] Wörle, J., Metz, B., Thiele, C. and Weller, G. (2019). Detecting sleep in drivers during highly automated driving: The potential of physiological parameters, *IET Intelligent Transport Systems* **13**(8): 1241–1248.

[66] Wynne, R. A., Beanland, V. and Salmon, P. M. (2019). Systematic review of driving simulator validation studies, *Safety science* **117**: 138–151.

[67] Xing, Y., Lv, C., Wang, H., Cao, D., Velenis, E. and Wang, F.-Y. (2019). Driver activity recognition for intelligent vehicles: A deep learning approach, *IEEE Transactions on Vehicular Technology* **68**(6): 5379–5390.

[68] Young, R. (2015). Revised odds ratio estimates of secondary tasks: A re-analysis of the 100-car naturalistic driving study data, *Technical report*, SAE Technical Paper.

[69] Zepf, S., Dittrich, M., Hernandez, J. and Schmitt, A. (2019). Towards empathetic car interfaces: Emotional triggers while driving, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[70] Zhang, C., Wu, X., Zheng, X. and Yu, S. (2019). Driver drowsiness detection using multi-channel second order blind identifications, *IEEE Access* **7**: 11829–11843.

[71] Zhang, L., Kang, J., Luo, H. and Zhong, B. (2018). Drivers' physiological response and emotional evaluation in the noisy environment of the control cabin of a shield tunneling machine, *Applied Acoustics* **138**: 1–8.

[72] Zhu, M., Wang, X., Tarko, A. et al. (2018). Modeling car-following behavior on urban expressways in shanghai: A naturalistic driving study, *Transportation research part C: emerging technologies* **93**: 425–445.

[73] Zhu, Y., Wang, Y., Li, G. and Guo, X. (2016). Recognizing and releasing drivers' negative emotions by using music: Evidence from driver anger, *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 173–178.

[74] Ziakopoulos, A., Tselentis, D., Kontaxi, A. and Yannis, G. (2020). A critical overview of driver recording tools, *Journal of safety research* .

[75] Zou, J., Li, Z. and Yan, P. (2019). Automatic monitoring of driver's physiological parameters based on microarray camera, *2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, IEEE, pp. 15–18.

Chapter 3

HARMONY: a Human-centered Multimodal Driving Study in the Wild

[1]

3.1 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.

3.1.1 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

[1] This chapter is ©2021 IEEE. Reprinted, with permission, from (16)

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.

3.1.2 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 (7; 15). This can include a place, one or a series of events, as well as the user of the application (7). 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.

3.2 The Proposed Framework

To achieve the identified goals, as shown in Fig. 3.1 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 3.3). 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 3.4). Lastly, we discuss the details on participant recruitment and our on-going data collection efforts (section 3.5).

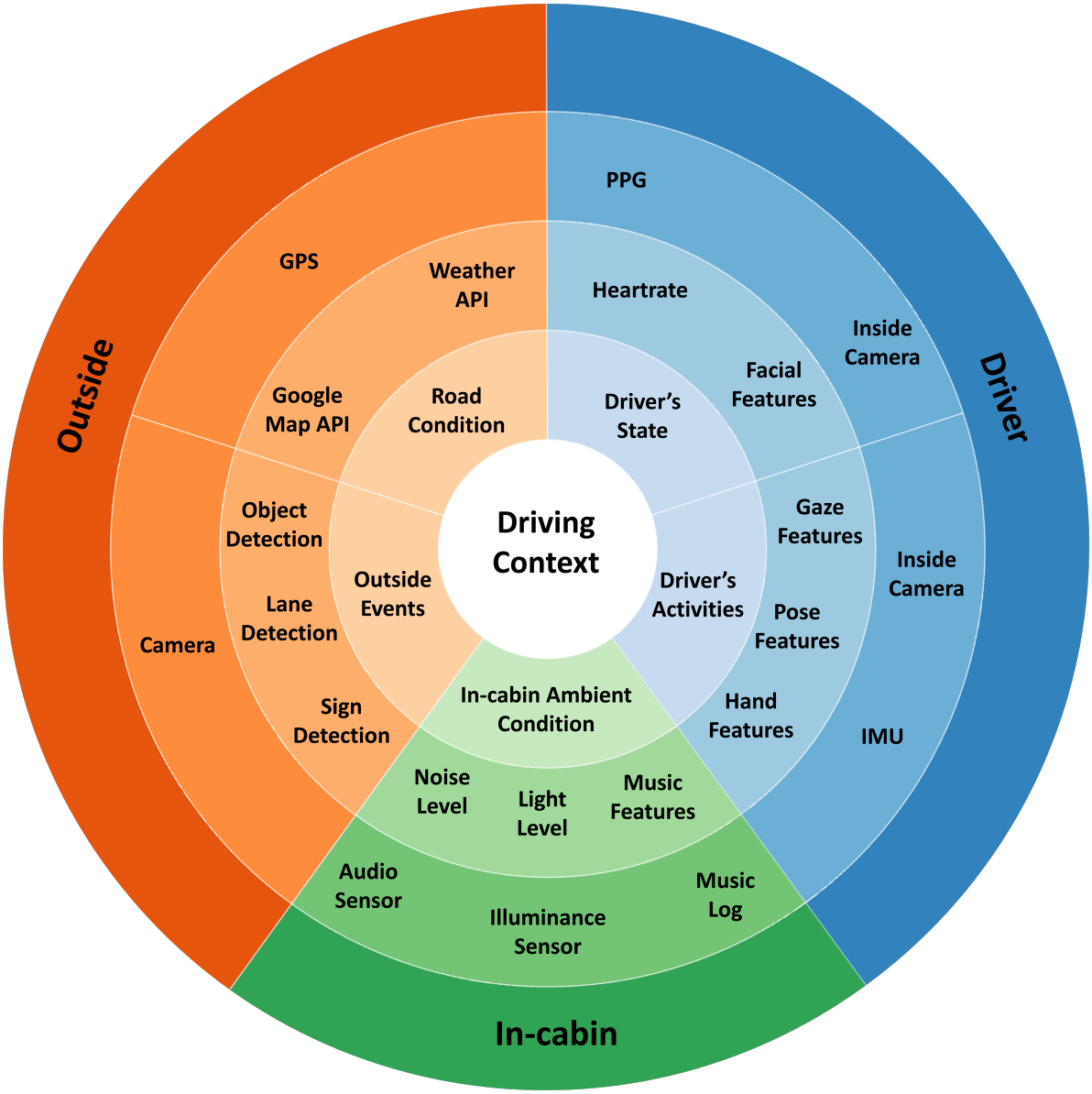


Figure 3.1: Data Collection Framework. Using this framework we can better contextualize driving scenarios and provide multimodal data while being low-cost ©2021 IEEE

3.3 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 3.3.1 and 3.3.2 provide more details about the utilized physical and virtual sensors, respectively.

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

3.3.1.1 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 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. 3.2.

3.3.1.2 Smartwatch:

To collect data with frequency and properties of interest, HARMONY uses an android smartwatch that is equipped with the "SWear" app (Fig. 3.3 - A), an in-house app designed for collecting long-term sensor data from smartwatches (5). SWear is available on Android store

Figure 3.2: sample view of the in-cabin camera (A), BlackVue dash camera (B), sample view of the outside cabin dash camera (C) ©2021 IEEE

(4) 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], 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. 3.3. Participants are required to start/stop the data collection for every session of driving (3.3 - 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. 3.3 - 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 (3.3 - 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.

3.3.2 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

Figure 3.3: sample view of the android smartwatch (A), homepage of the app (B), status page of the app (C), setting page of the app (D) ©2021 IEEE

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 (10), and OpenWeatherMap (22). 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 3.4.

3.4 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.

3.4.1 Video information

Video information is retrieved automatically through FFMPEG video packages in Python (8). 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).

3.4.2 Gaze, Pose, and Facial Features

Each in-cabin video is analyzed using three software: OpenFace (3), OpenPose (6), and Affectiva (13). 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.

3.4.3 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 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.

3.4.4 Music Features

The next step is to use the collected music log as described in section 3.2, 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 (14) and Genius (9).

3.4.5 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).

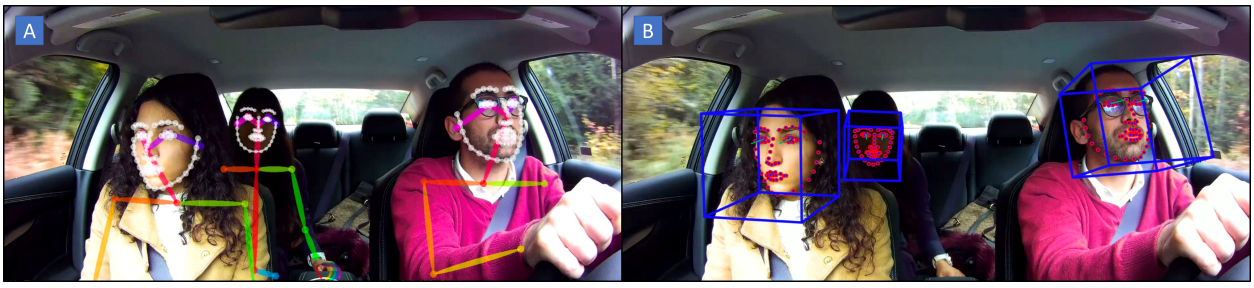


Figure 3.4: OpenPose sample output providing driver's skeleton key point (A), OpenFace sample output providing driver's facial and gaze measures in the wild (B) ©2021 IEEE

3.4.6 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.

3.4.7 Event Detection

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

3.4.7.1 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 an-

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 3.5: Annotation table, sample view of the trigger codes (A), sample annotation for a video (B) ©2021 IEEE

notator(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. 3.5. 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

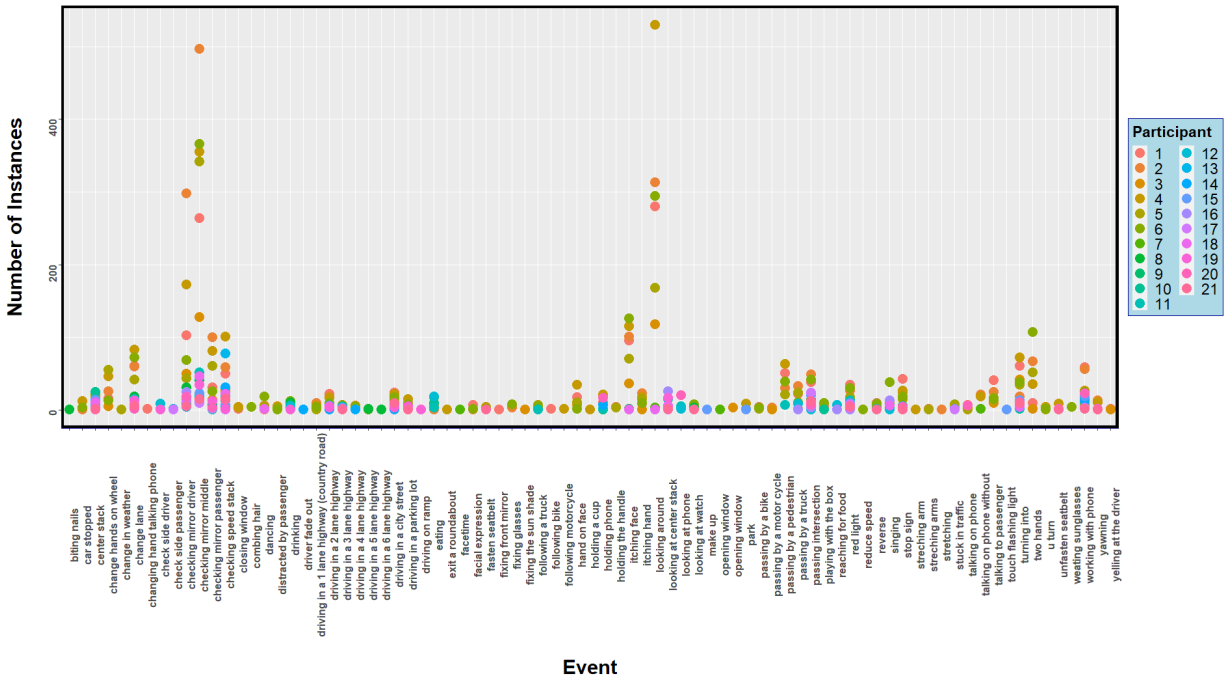


Figure 3.6: Number of the annotated instances for part of the data that only video is available to date ©2021 IEEE

category. This will help 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. 3.6 and 3.7 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 (17).

3.4.7.2 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 (1) and sign detection (2), 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

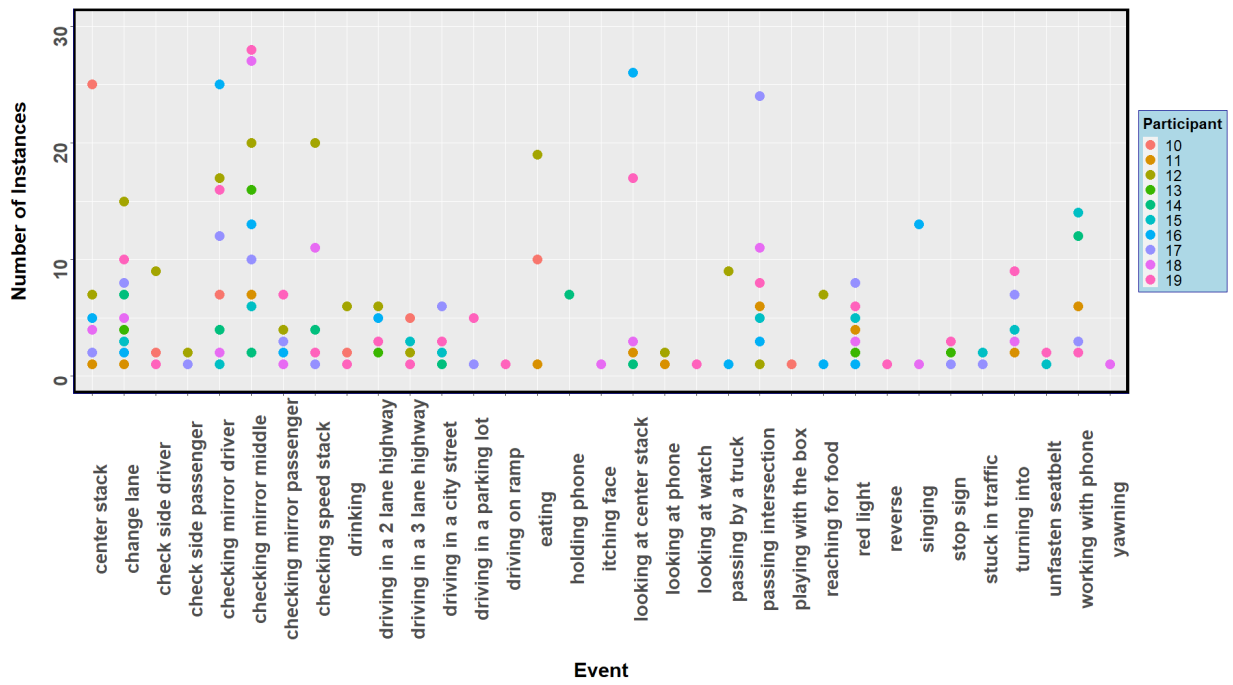


Figure 3.7: Number of the annotated instances for part of the data that all modalities are available, to date. Note that these include less data as not every video has an available physiological data attached to it. ©2021 IEEE

light at an intersection.

3.5 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. 3.8 demonstrates the locations of all the collected data to date. The most recent map of the collected data can be viewed in (18). The data has been mostly collected from eastern

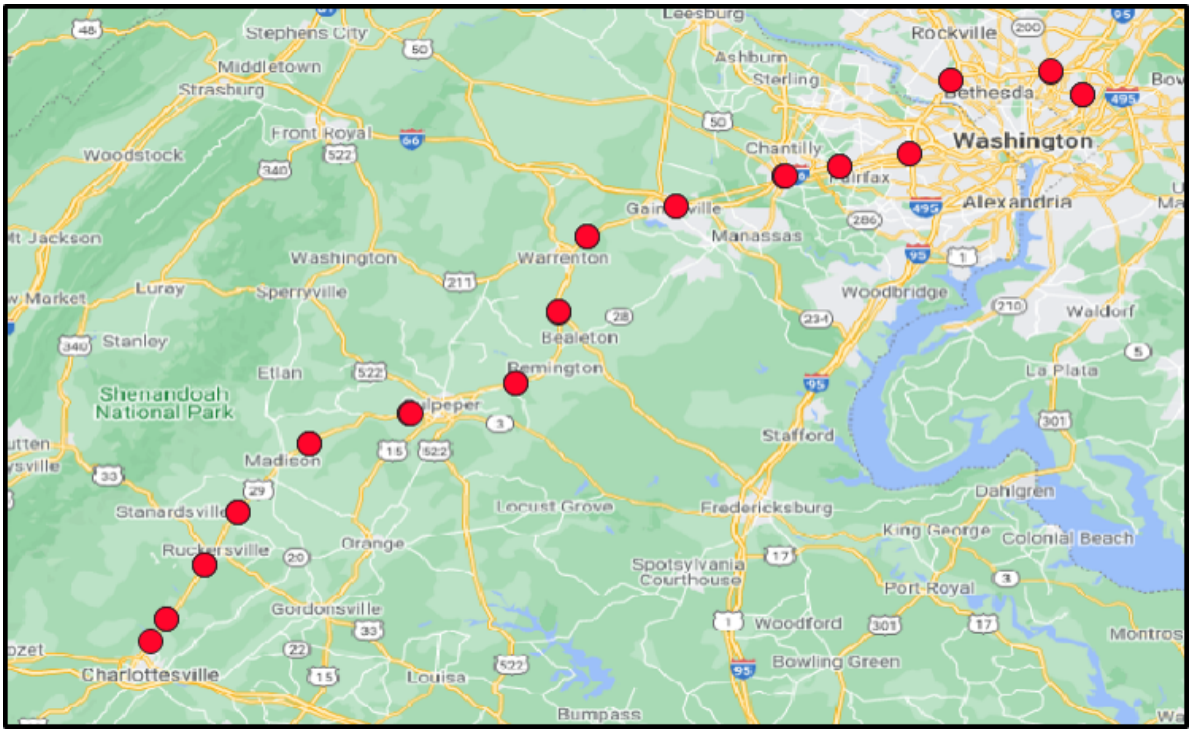


Figure 3.8: GPS location of the collected data to date. Although participants are mostly from the state of Virginia, the data includes the roads from northeastern regions of the country. ©2021 IEEE

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 3.1.

3.6 A method for analyzing the amount of data required by the system

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 (20; 12; 21). 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

Table 3.1: Details of collected data to date through HARMONY framework ©2021 IEEE

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	-

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. 3.9 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 (19).

Previous research has used KDE to assess the amount of data needed for an NDS provided in (19). 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

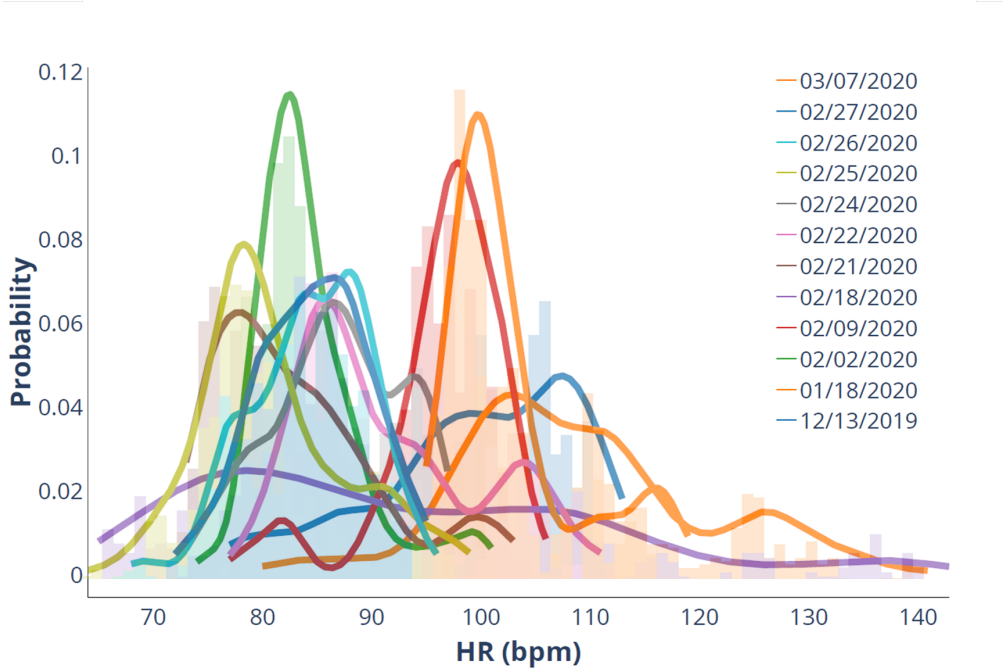


Figure 3.9: A participant's HR distribution while driving through the same route for multiple days. Note that the distribution changes on a daily basis, pointing to a need in collecting long-term longitudinal data ©2021 IEEE

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 (19). Through this method, the estimation for the probability distribution can be calculated as shown in equation 1 (19; 11). Considering we have a sample sequence of data $\{X_i\}_1^n$:

$$\hat{f}(x) = n^{-1} \sum_{1}^{n} h^{-1} K(x - X_i) \tag{3.1}$$

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:

$$KL(\hat{f}(x;n)||\hat{g}(x;m)) = \int [\hat{f}(x;n) * log(\frac{\hat{f}(x;n)}{\hat{g}(x;m)})] \tag{3.2}$$

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:

$$\begin{gathered} KL(\hat{f}(x;n+m)||\hat{f}(x;n)) = \\ \int [\hat{f}(x;n+m) * log(\frac{\hat{f}(x;n+m)}{\hat{f}(x;n)})] \end{gathered} \tag{3.3}$$

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. 3.10 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

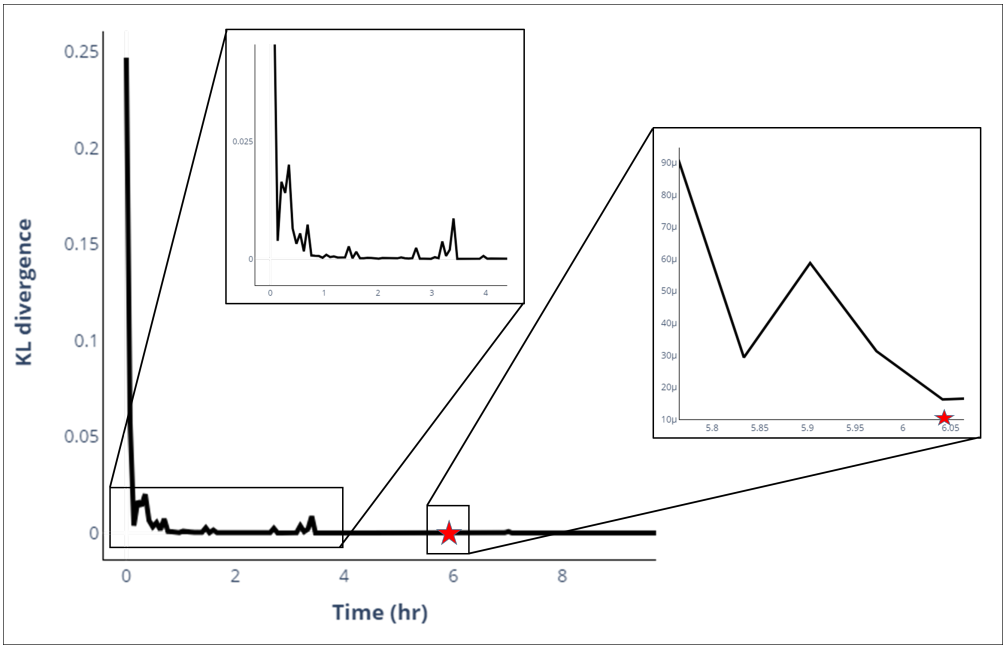


Figure 3.10: Kernel density estimation analysis for 3 months of HR data. The KL divergence algorithm demonstrates that after approximately 3 months of data collection (6 hours), the difference between estimated kernels reaches below 1e-5 ©2021 IEEE

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.

Bibliography

[1] Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow, `https://github.com/matterport/Mask_RCNN`.

[2] Arcos-Garcia, A., Alvarez-Garcia, J. A. and Soria-Morillo, L. M. (2018). Evaluation of deep neural networks for traffic sign detection systems, *Neurocomputing* **316**: 332–344.

[3] Baltrusaitis, T., Zadeh, A., Lim, Y. C. and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, pp. 59–66.

[4] Boukhechba, M. (n.d.). Swear - apps on google play.
URL: *https://play.google.com/store/apps/details?id=uva.swear*

[5] Boukhechba, M. and Barnes, L. E. (2020). Swear: Sensing using wearables. generalized human crowdsensing on smartwatches, *2019 IEEE 11th International Conference on Applied Human Factors and Ergonomics. IEEE.*

[6] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y. (2018). Openpose: realtime multi-person 2d pose estimation using part affinity fields, *arXiv preprint arXiv:1812.08008* .

[7] Dey, A. K. (2001). Understanding and using context, *Personal and ubiquitous computing* **5**(1): 4–7.

[8] *FFMPEG* (n.d.).
URL: *https://ffmpeg.org/*

[9] Genius (n.d.). Genius: A pythonic implementation of lyrics.
URL: *genius.com*

[10] *Geo-location APIs* (n.d.).
URL: *https://cloud.google.com/maps-platform/*

[11] Heidenreich, N.-B., Schindler, A. and Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors, *AStA Advances in Statistical Analysis* **97**(4): 403–433.

[12] Li, N., Misu, T. and Miranda, A. (2016). Driver behavior event detection for manual annotation by clustering of the driver physiological signals, *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2583–2588.

[13] McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J. and Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit, *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pp. 3723–3726.

[14] Pradipta (n.d.). Song lyrics knowledge.
URL: *https://github.com/geekpradd/PyLyrics*

[15] Rakotonirainy, A. and Maire, F. D. (2005). Context-aware driving behavioural model.

[16] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* **9**: 23956–23978.

[17] UVABRAINLAB (n.d.a). Harmony annotation scheme and details.
URL: *http://uvabrainlab.com/portfolio/reason-aware-annotation-for-contextualizing-driving-scenarios/*

[18] UVABRAINLAB (n.d.b). Harmony most recent map.
URL: *http://uvabrainlab.com/portfolio/driving-behavior/*

[19] Wang, W., Liu, C. and Zhao, D. (2017). How much data are enough? a statistical approach with case study on longitudinal driving behavior, *IEEE Transactions on Intelligent Vehicles* **2**(2): 85–98.

[20] Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B. and Poom, L. (2015). Physiological responses related to moderate mental load during car driving in field conditions, *Biological psychology* **108**: 115–125.

[21] Zepf, S., Dittrich, M., Hernandez, J. and Schmitt, A. (2019). Towards empathetic car interfaces: Emotional triggers while driving, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[22] *urrent weather and forecast* (n.d.).
URL: *https://openweathermap.org/*

Chapter 4

Multimodal Driver State and Behavior Detection

This chapter focuses on detecting driver's state based on multimodal data in the wild. In this chapter, two separate methods are showcased for this purpose. The first method, uses conventional machine learning frameworks to perform supervised learning in detecting driver's state, behaviors, and driving environment based on the smartwatch data[1]. The second method, then dives into unsupervised learning scheme using text analytic approaches to find driving behavior and state topics for each segment of driving.

4.1 Project 1: Multimodal Driver State and Behavior Detection Through Supervised Learning

4.1.1 Introduction

Autonomous technologies are evolving exponentially, changing the interaction between semi-automated vehicles and the human driver drastically (6; 40). Previously, advanced driving assisting systems (ADAS) (e.g., cruise control, or lane-keeping assisting system) were only used to assist the driver who acted as a supervisor; however, with the fast-paced improvements in auto-pilot systems (e.g., autopilot in TESLA vehicles), the line between the supervisor and a sole driver is starting to blur. This approach to autonomy is referred to as shared-autonomy, in which the task of driving is achieved through a collaboration between the driver and the vehicle (35). Despite their successes, these approaches still rely on the driver to take over in critical situations without the vehicle having any understanding of the current state of the driver (e.g., distracted, fatigued, stressed, or under the influence of substances). In order to address these issues, the shared autonomy approach generally includes a driver sensing module in which the driver's states and behaviors are monitored in real-time. An example of a commercial semi-automated system that has extensive driver monitoring is the Comma.ai's *openpilot* (24), which monitors the driver in real-time for driver distraction

[1]This project is ©2021 IEEE. Reprinted, with permission, from (101)

detection. Through the driver sensing module, the vehicle is provided with a contextual awareness with respect to the driver's status in real-time.

Driver sensing systems often rely on camera streams as they are very powerful in revealing driver's activities, states, and behaviors in real-time through different activity recognition (76; 60), distraction detection (48), cognitive load estimation (37), and emotion detection methods (1). However, camera streams suffer from a few hardware, driving, and privacy-related issues. First, changes in illumination (e.g., night condition), having shadows, angle of the camera, quality of videos, and even type of the camera can significantly influence their accuracy and reliability in detecting driver's state and behaviors (45; 76). Second, to cover the entire in-cabin space, multiple cameras have to be used because one camera cannot cover all in-cabin and outside conditions (83; 73). Third, cameras are highly intrusive in nature, which may make people feel uncomfortable when being constantly monitored (93). Due to privacy concerns, many times, people prefer not sharing their video/audio data, making it impossible to track their states (4). Lastly, although we have seen significant improvements in computational techniques, processing video streams and applying computer vision techniques to detect human activity and state are still relatively costly and resource extensive (43).

Although recent improvements such as enhancements in night vision capability of cameras or fusing radar systems with camera streams have overcome some of the aforementioned problems, the issue with privacy intrusion, cost, and high computational requirement still remain as main issues with cameras. An alternative solution is to use passive sensing techniques through smart wearable devices to detect driver behavior and states. These passive sensing methods can be used to (1) optimize video recording duration to the least amount possible (e.g., only collecting videos when passive sensing requires additional contextual information on the driver and environment); (2) cover situations that vision systems cannot capture (e.g., in low-light conditions); and (3) use less computation energy and cost as compared to video recording. While passive sensing has been studied in driving research, their applicability in real-world driving scenarios remains an open question, partly because these technologies are just starting to become popular and emerge as a long-term deployment solution (102).

In this study, we leverage ubiquitous sensing for classifying various elements of driving context, including driver's activity, outside events, and driving environment attributes in a fully naturalistic driving study. We first provide a summary of previous works in the driver's state and behavior recognition using wearable devices. Then we delve into our methodology for collecting and processing passive driver sensing data. We compare and contrast multiple machine learning methods on classifying driver's activities, outside events, and road attributes from non-intrusive data in real-world driving situations. Our work

facilitates the transition to using off-the-shelf wearable devices for detecting driver's state and behaviors.

4.1.2 Background Study

Recent studies have provided insights into the application of wearable devices and ubiquitous computing in driving studies. These studies primarily focused on various states of the driver such as stress (22; 99; 100), drowsiness (56; 22), distraction (14), fatigue (22), as well as different driving behaviors such as take-over readiness (74; 75; 5), driving maneuvers (e.g., turning) (62), turning and lane changes (42).

(62) used a combination of phone and wearable device to detect driving maneuvers such as steering wheel turning angle. In their study, authors first detect driver's sensor reading from the pool of driver, passenger, and vehicle sensor data with an accuracy of 98.9 %. Then, based on the driver's hand inertial measurement unit (IMU) measurements, the steering wheel angle is estimated in a stationary vehicle. (84) analyzed driver's arousal level in a driving simulator using a wearable device, worn on the participants' writs. In their study, authors used deep neural networks to detect three levels of arousal, under, normal, and over-aroused based on the driver's heart rate, skin conductance, and skin temperature. The authors were able to achieve an F-score of 0.82 for arousal detection through recruiting 11 participants. Similarly, (22) used a wrist-worn wearable device for detecting driver's stress, fatigue, and drowsiness. By using a driving simulator with 28 participants, they were able to achieve an accuracy level of 98.43 % for detecting each state. (39) used a wearable smartwatch to detect instances of driver's distraction in a driving simulator. In their study, by collecting physiological data from 16 participants and utilizing multiple machine learning algorithms such as decision trees and support vector machines to classify different driving states, they achieved an average accuracy of 89 % in distraction detection.

(14) used a wearable device and a phone for detecting driver's distraction based on driver's hand movement. In their study, authors developed an app that detects various motions of the driver's hands, such as different types of holding the steering wheel, by fusing the sensor readings from both driver's hand and the phone's IMU measurements. By using the IMU sensors on the wearable device (59) was able to detect different classes of phone, text, drink, using a marker, using a touch screen, and normal driving with an F-1 score of 0.87. (42) used a smartwatch and wearable magnetic ring along with eye-tracking glasses to detect driver's distraction and activities such as turning and lane changes. Their systems achieved 87 % precision and 90 % recall on data collected from 10 participants in an on-road controlled study. In a recent study, (56) used off-the-shelf wearable devices to detect driver's drowsiness in a driving simulator. In their study, authors have used heart rate signals for drowsiness detection and achieved an accuracy of 99.9 %. This study also pointed out the differences

between age groups for drowsiness detection. (99) used wearable data from 12 participants and compared drivers' heart rate variability in different road types (i.e., city versus highway), weather conditions (i.e., rainy versus clear), and presence of a passenger. They found that drivers are on average calmer on highways, in clear weather, and when being accompanied by a passenger.

Although these studies provided rich insights into the application of wearable devices in driving, they are mostly performed in controlled environments (i.e., driving simulators and on-road controlled studies). This can become problematic because most of these studies lack the proper ecological and external validity to allow their findings to generalize in real-life contextual settings where different real-world challenges and noise exist. Additionally, most of these studies have only leveraged the IMU sensor (i.e., accelerometer and gyroscope) on the smart wearable for driver behavior detection leaving other sensors such as heart rate (HR) and Photoplethysmography (PPG) unexplored. Furthermore, as opposed to most of the studies that used wearable devices that were designed specifically for the study of interest, we are interested in exploring ubiquitous sensors such as off-the-shelf devices (e.g., smartwatches) for driver state and behavior analysis in the wild. This is due to the fact that these devices are (1) already used by drivers and participants in their daily lives, and (2) they have multiple sensors built into them.

4.1.3 Methodology

We now present our approach. We first describe the data collection and annotation protocol. Then, we provide details on the feature exploration to test the feasibility of the classification problem. Finally, we describe our feature extraction and machine learning classification pipeline.

4.1.3.1 Data Collection

The data in this study has been collected through HARMONY, a human-centered naturalistic driving study platform (103). Here we provide an overview of the data collection and cleaning process. However, the reader is invited to refer to (103) for more detailed information on the data collection, storage, and processing within the HARMONY system. A sample data of HARMONY is made publicly available to researchers and can be retrieved from (106).

HARMONY is a framework that is leveraged to collect driver's behavioral and state measures together with environmental sensing, through physical devices such as cameras (both inside and outside), and wearable devices, as well as virtual sensors such as computer vision algorithms or APIs to extract features from the data collected by the physical sensors. Within the current HARMONY dataset and framework, 21 participants' vehicles are

equipped with a BlackVue DR-750S-2CH dash camera, which records both inside and outside of the vehicle as the car turns on and the driving scenario starts. Each driver is provided with an android smartwatch, equipped with an in-house app, namely SWear (17). SWear is available on Google Play store (16) and is designed to facilitate the process of data collection on smartwatches by adding the ability to control each sensor's data collection frequency to the desired sensing specifics and sync sensors data to the cloud. Through this, we collected drivers' hand acceleration [10Hz], gyroscope measures [10Hz], Photoplethysmography (PPG) [10Hz], ambient audio amplitude (noise level) [1/60 HZ], HR [1Hz], ambient light intensity [1/60 HZ], and GPS location [1/600 HZ]. The data collection set-up in participants vehicle can be viewed on Fig. 4.1 - A.

4.1.3.2 Data Annotation and inspection

We have chosen a random subset of the collected data for the purpose of this study from 15 participants within HARMONY. We use the in-cabin and outside facing videos for manual annotation of driver's activities, outside events, and environmental attributes for training purpose (Fig. 4.1 - A). For this study, we have divided up driver activities into working with the mobile phone, checking sides, which often includes checking mirrors, checking speed stack, checking center stack, eating/drinking food, searching for an item on the passenger's seat, touching face, and singing and dancing. Outside events consist of passing an intersection, staying behind a traffic light, being stuck in dense traffic, and changing lanes. The road conditions are divided into the city streets, 2-lane highway, 3-lane (or more) highway, parking lot, and merging ramps. Since all devices' timestamps are synced, we use the start time of these events to pull the wearable data, such as IMU, PPG, light, and noise levels. A sample of annotated data can be view on Fig. 4.1 - A. Table 4.1 shows the final count of data samples per activity as used for this study.

Additionally, we inspected the aforementioned categories prior to classification for visual observation of differences in sensor reading distribution. This helps investigate how sensing data fluctuates across these settings and lay the ground for the feasibility of classifying them through machine learning algorithms. For instance, drivers' heart rate follows different distributions when performing different activities. This can be seen on Fig. 4.2 - A where the distribution of heart rate in singing has a peak at a higher heart rate level as compared to the food-related (i.e., eating and drinking) category. Additionally, we observe differences in the distribution of IMU and light sensors as depicted on Fig. 4.2.

4.1.3.3 Feature Extraction

Working with the data extracted from commercially available wearable devices has two main challenges; First, The average data collection frequency of each modality can be different.

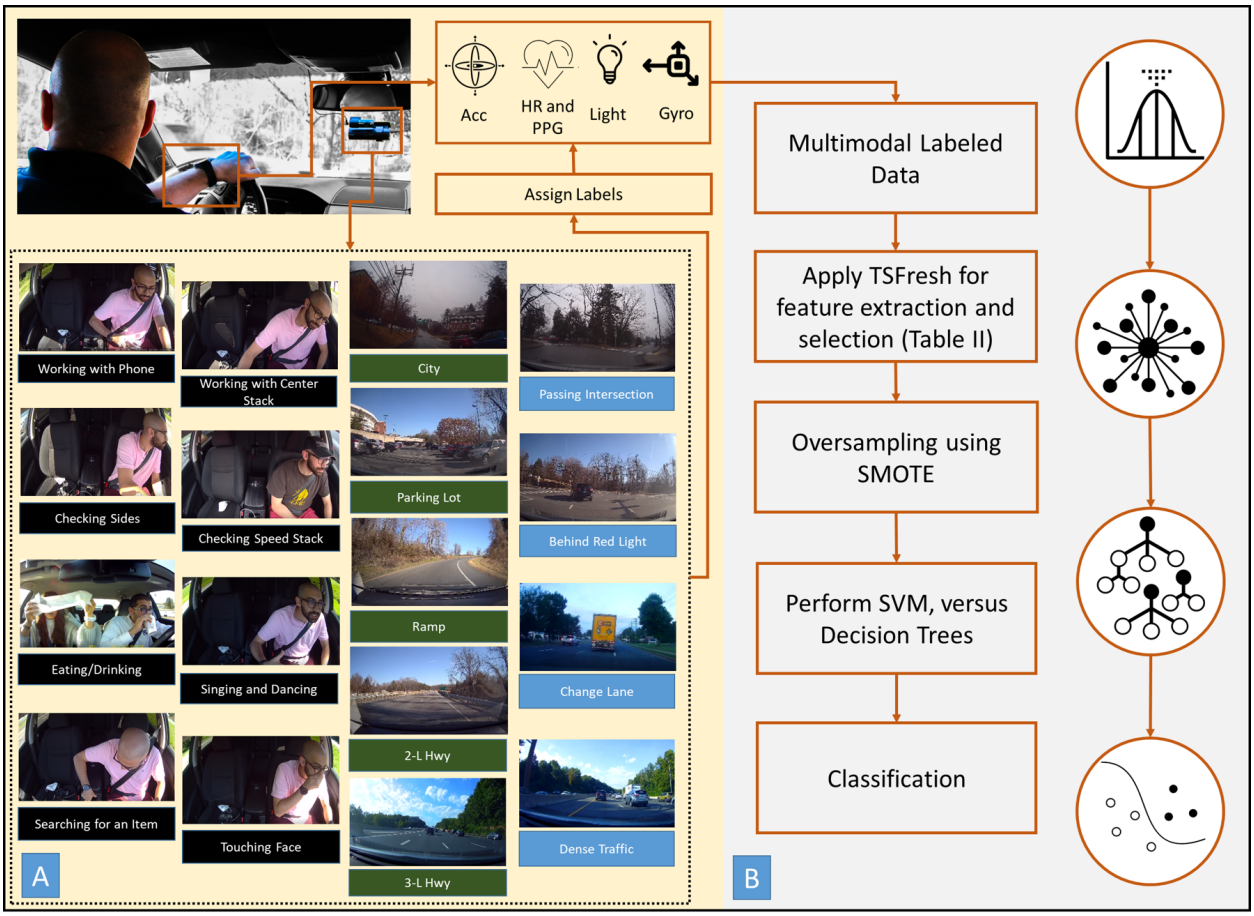


Figure 4.1: Summary of our methodology for integrating smartwatches into driver behavior analysis in the wild. We use videos to label smartwatch data ©2021 IEEE

Table 4.1: Summary of data samples used for classification ©2021 IEEE

Category	Event	No. of Samples	Average Duration (sec)
Inside Activity	Checking Sides	128000	1
Inside Activity	Eating/Drinking	27584	4
Inside Activity	Working with Center Stack	6540	3.5
Inside Activity	Checking Speed Stack	24000	1
Inside Activity	Touching Face	14500	5.3
Inside Activity	Working with Phone	47000	22
Inside Activity	Singing and Dancing	5870	22
Inside Activity	Searching for an Item	2680	10
Outside Event	Change Lane	22437	3
Outside Event	Passing an Intersection	36722	3
Outside Event	Traffic Light	42670	20
Outside Event	Stuck in Traffic	4542	60
Road Type	City Street	182614	158
Road Type	Parking Lot	12553	38
Road Type	Merging Ramp	4984	26
Road Type	2L - Highway	180295	168
Road Type	3L - Highway	104102	132

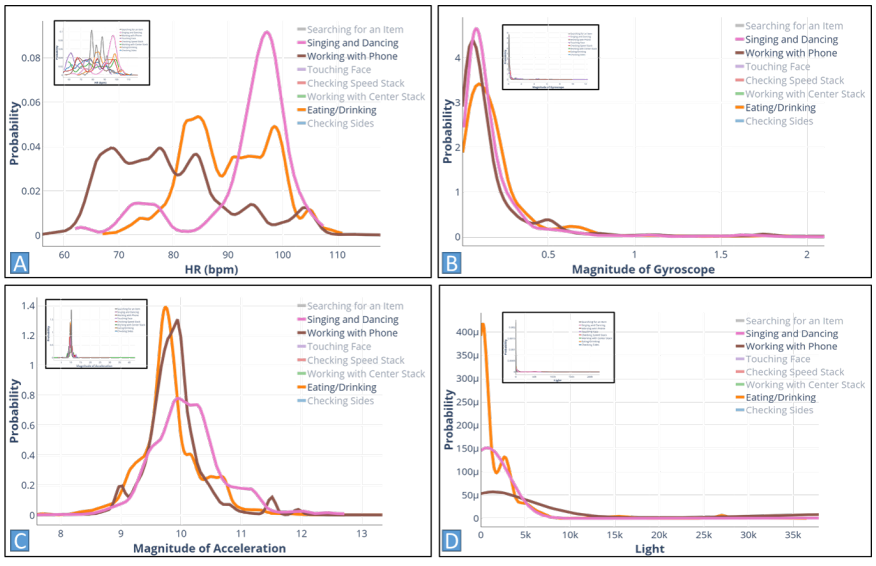


Figure 4.2: Distribution of HR (A), accelerometer (B), gyroscope (C), and light (D) data among different activities. This example shows how the distribution of each modality varies across activities and motivates its use as a predictor of drivers states. Note the distribution differences among different activities, such as shifted HR peaks for singing as compared to food-related and phone-related activities. ©2021 IEEE

Table 4.2: Summary of selected features ©2021 IEEE

Domain	Features
Time	Kurtosis, mean, standard deviation, maximum, minimum, variance, skewness, median, variation coefficient, absolute sum of changes, Benford correlation, count above mean, count below mean, first location of maximum, first location of minimum, has duplicate, has duplicate max, has duplicate min, last location of maximum, last location of minimum, longest strike above mean, longest strike below mean, mean abs change, mean change, mean second derivative central, sum of reoccurring data points, sum of reoccurring values, sum values
Frequency	Energy, power, entropy

Additionally, each sensor's data collection frequency is not constant, and it may vary throughout the time due to hardware requirements (e.g., the frequency of HR data varies between 0.8Hz to 1.1 Hz at different battery levels). Although we can change the average frequency value through SWear, the exact frequency cannot be defined prior to data collection. To solve these challenges, we resampled the data at 10 Hz, which is equal to the IMU, Light, and PPG sensor average frequency, which have the lowest sampling rate.

In order to perform feature extraction, we analyzed the duration of different activities to find the correct sliding window. Additionally, previous literature ((105)) on accelerometer-based activity recognition had stated a window overlap of 50 percent to avoid correlations between samples. We chose a sliding window of 1 second to cover all the activities duration (see Table 4.1) in our feature extraction, with 50 % overlap. In order to perform feature extraction, we leveraged the tsfresh package (23) on Python to retrieve multiple time and frequency domain features from each individual wearable sensor (i.e., Light, PPG, HR, and IMU). Additionally, we have leveraged previous literature in human and driver activity and state recognition using biosignals to select relevant features from the tsfresh library (7; 105; 75; 41; 92). Table 4.2 shows the selected features for this study.

Because the frequency of different activities varies across different driving scenarios, the dataset has an unbalanced nature with respect to different classes. For instance, in our annotated dataset, the number of the mirror checking events is noticeably higher than the phone usage activities (see Table 4.1). To solve this issue, we have used two separate methods of:

(A) balancing the classes by weighing each class to be inversely proportional to its frequency in the dataset.

(B) Oversampling based on Synthetic Minority Oversampling Technique (SMOTE) (20) to

generate new samples for the minority classes. SMOTE generates new data points from convex combinations of nearest neighbours. We apply SMOTE only on the training set. This is important for keeping the classifier unbiased towards new samples.

Using the extracted features, We built different models to classify three major groups of driver's activities, environmental events, and environmental attributes as described earlier. In each category, three different machine learning methods were used to build classifiers using a 10-fold cross-validation method. The classifiers included random forest, decision trees, and extra decision trees. We observed that the Random Forest model outperforms the other two classifiers, thus we only focus on providing the results for the Random Forest classifier. For this, the Scikit-learn package (78) was utilized.

4.1.4 Results

Fig. 4.3 demonstrates the confusion matrix for each of the three classification tasks (A: driver activity, B: outside events, and C: road types) using the highest accuracy model trained on the original data, as well as the models trained on oversampled data (e.g., A1 is the original data and A-2 shows the oversampled data).

For classifying activities, the Random Forest model outperforms the other two methods (i.e., decision trees and extra decision tress), with an average F1 Score of 90.99 % on the imbalanced data (4.3). This model does relatively poorly on the two categories of "searching for an item," and "working with the center stack display of the vehicle" (Fig. 4.3 - A-1 and A-2). The searching for an item category is mostly due to lower amount available data. The center stack category may be mistaken with the checking sides category 20 percent of the time. We suspect this is mostly due to the fact that the participants body movements are very similar when working with the center stack and checking different sides of the vehicle (e.g., checking side mirrors). To further confirm this, we perform the SMOTE oversampling on the training set and train the classifier on the oversampling data. It is recognizable that the accuracy of detecting the searching for an item category increases significantly (i.e., from 75.9 % to 88.6.3 %), while the center stack detection accuracy does not increase as much (i.e., from 73.9 % to 81.7 %) (Fig. 4.3 - A-1 and A-2). Finally, the model trained on oversampled data, on average achieves an F1 score of 94.55 %.

For classifying outside events, the Random Forest classifier outperforms the other methods (Table 4.3) with an average F1 score of 97.68 %. Using SMOTE, we can further enhance the classification F1 score to 98.26 %. Oversampling, mostly enhances change lane and passing intersection categories but does not change the overall accuracy significantly. Similarly, Random Forest outperforms other models in detecting the road type with an average F1 score of 93.62 % and 97.68 % for the original data and oversampled data respectively. Due

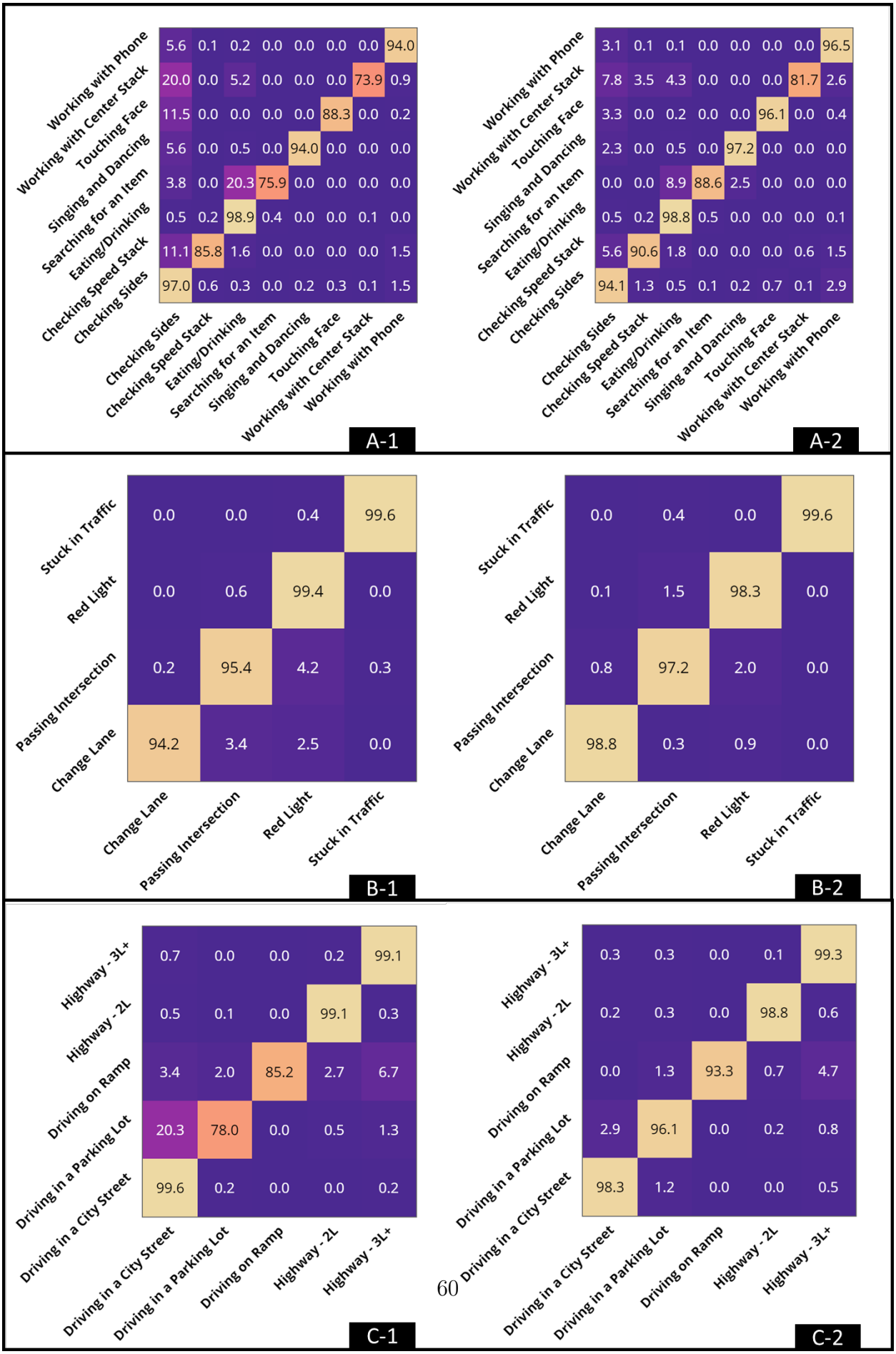
to lower amount of data, the model performs relatively poorly on the Driving in a Parking Lot and Driving on Ramp categories, which is then further enhanced by using the oversampling methods.

Table 4.3: Summary of model performance on different target categories when applied with a 10-fold cross validation. Note the enhancements in performance when generating new samples using the SMOTE method. ©2021 IEEE

	Imbalanced		SMOTE	
Target Category	F1 score	SD	F1 score	SD
Driver Activity	90.99	0.92	94.55	0.77
Outside Events	97.68	0.69	98.27	0.28
Road Type	93.62	0.69	97.86	0.25

Additionally, we assessed the contribution of multimodality to the classification accuracy of the driver's activity model. We emphasize on this attribute of current off-the-shelf smart wearable devices as previous works were mostly focused on laboratory-built devices, lacking multimodal sensor data built into one device. Fig. 4.4 demonstrates the accuracy of the activity recognition model trained on the original data using Random Forest classifier with varying sensor combinations. As shown, overall by adding different modalities, the activity recognition model's F1 Score increases.

To further analyze the effect of each modality on the classification task, we performed a permutation feature importance analysis (78) on the three aforementioned classifiers. For simplicity, this analysis is performed only on a hold-out sample in our dataset. Similar to the previous section, the training set is oversampled through SMOTE. Fig. 4.5 shows the top ten features in each target category of driver activity, outside events, and road type. As shown on Fig. 4.5 - A for classifying the driver's activities, five out of the top ten features are from the accelerometer modality, which draws attention to the importance of the IMU sensor for activity recognition similar to the previous research as mentioned in the background section. However, as shown on Fig 4.5 - B, heart rate, PPG, and light are also contributing to the classification task, pointing out the importance of multimodal sensing. Additionally, for classifying the outside events, we observe the inclusion of gyroscope in the top ten features, which points out the differences among driver's hand rotations for certain outside events (e.g., change lane action). Moreover, note that three out of the top ten features are related to the PPG sensor, which is shown to be correlated with driver's state, such as stress levels (49). Lastly, the PPG sensor is contributing the most to the road type detection task. This is in line with previous research demonstrating the variation in driver's HR when driving in different road and weather conditions (103).


Working with Phone
Working with Center Stack
Touching Face
Singing and Dancing
Searching for an Item
Eating/Drinking
Checking Speed Stack
Checking Sides
5.6 0.1 0.2 0.0 0.0 0.0 0.0 94.0
20.0 0.0 5.2 0.0 0.0 0.0 73.9 0.9
11.5 0.0 0.0 0.0 0.0 88.3 0.0 0.2
5.6 0.0 0.5 0.0 94.0 0.0 0.0 0.0
3.8 0.0 20.3 75.9 0.0 0.0 0.0 0.0
0.5 0.2 98.9 0.4 0.0 0.0 0.1 0.0
11.1 85.8 1.6 0.0 0.0 0.0 0.0 1.5
97.0 0.6 0.3 0.0 0.2 0.3 0.1 1.5
A-1
3.1 0.1 0.1 0.0 0.0 0.0 0.0 96.5
7.8 3.5 4.3 0.0 0.0 0.0 81.7 2.6
3.3 0.0 0.2 0.0 0.0 96.1 0.0 0.4
2.3 0.0 0.5 0.0 97.2 0.0 0.0 0.0
0.0 0.0 8.9 88.6 2.5 0.0 0.0 0.0
0.5 0.2 98.8 0.5 0.0 0.0 0.0 0.1
5.6 90.6 1.8 0.0 0.0 0.0 0.6 1.5
94.1 1.3 0.5 0.1 0.2 0.7 0.1 2.9
A-2
Stuck in Traffic
Red Light
Passing Intersection
Change Lane
0.0 0.0 0.4 99.6
0.0 0.6 99.4 0.0
0.2 95.4 4.2 0.3
94.2 3.4 2.5 0.0
B-1
0.0 0.4 0.0 99.6
0.1 1.5 98.3 0.0
0.8 97.2 2.0 0.0
98.8 0.3 0.9 0.0
B-2
Highway - 3L+
Highway - 2L
Driving on Ramp
Driving in a Parking Lot
Driving in a City Street
0.7 0.0 0.0 0.2 99.1
0.5 0.1 0.0 99.1 0.3
3.4 2.0 85.2 2.7 6.7
20.3 78.0 0.0 0.5 1.3
99.6 0.2 0.0 0.0 0.2
C-1
0.3 0.3 0.0 0.1 99.3
0.2 0.3 0.0 98.8 0.6
0.0 1.3 93.3 0.7 4.7
2.9 96.1 0.0 0.2 0.8
98.3 1.2 0.0 0.0 0.5
C-2

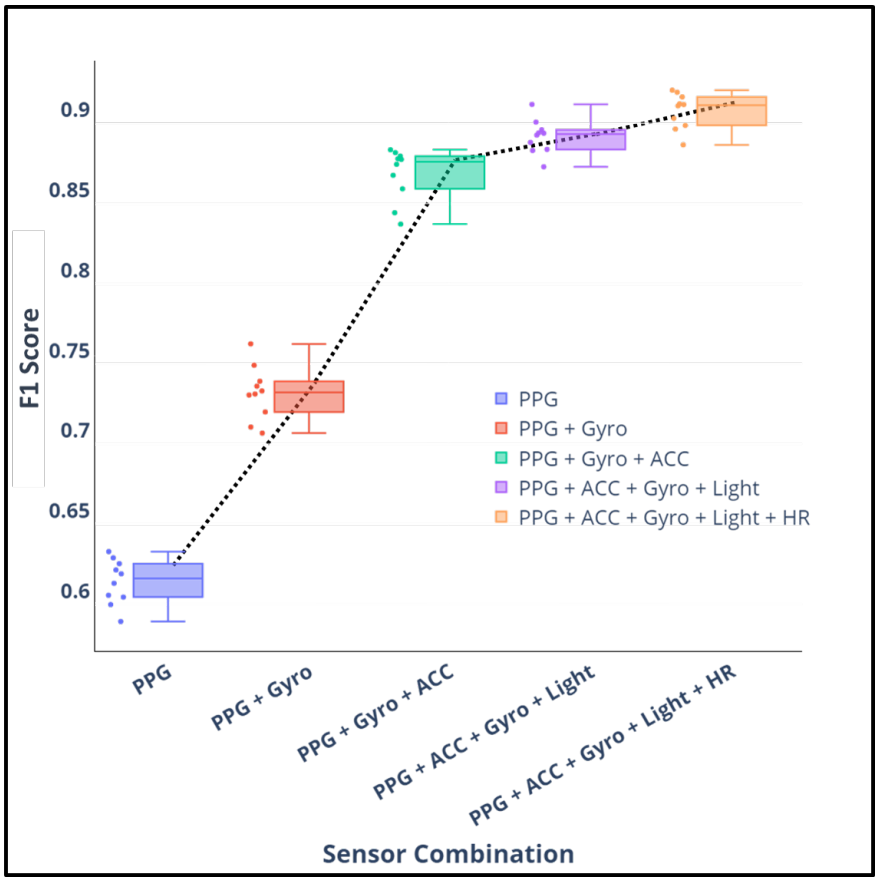


Figure 4.4: The comparison of adding each sensor to the classification scheme in a 10 fold cross validation scheme for driver activity classification model. Note the increase in F1-Score as new modalities provide additional information to the classifier. ©2021 IEEE

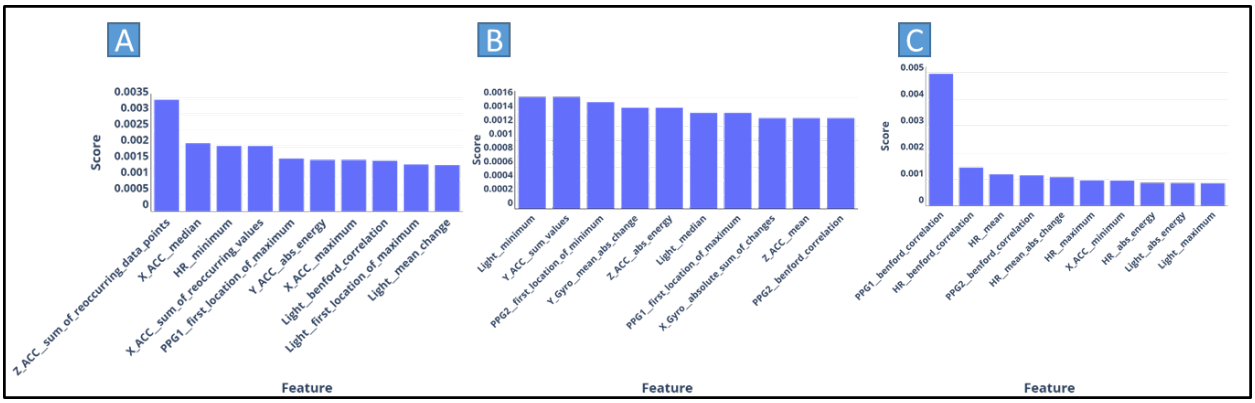


Figure 4.5: Feature importance analysis on the three targeted categories of (A) driver's activities, (B) outside events , and (C) road types. Note that the contribution of different modalities varies across the different targeted predictions. ©2021 IEEE

4.1.5 Discussion

In this study, we have classified real-world driving instances such as driver's activities, outside events, and outside road attributes using data extracted from commercially available smartwatches. Smartwatch multimodal data has the potential to reveal different aspects of driving behavior and activities such as movements (e.g., through IMU sensors), underlying state of the driver such as stress (e.g., through HR sensor), and environmental attributes (e.g., through noise and light sensors). Note that with the current developments in ubiquitous computing devices, multiple other sensors such as skin temperature, breathing patterns, skin conductance, and even electrocardiogram (ECG) are starting to be built into these devices, which can further push the boundaries of driver state recognition. While one sensing modality (e.g., Light) might not be able to provide all the information to detect certain activities, other sensing modalities such as IMU may still provide crucial missing information to aid the classifier (see Fig. 4.4). As a result, by having access to different modalities, we can identify different activities and events with higher accuracy. This has also been shown in our previous work (103) that certain environmental events might leave a driver's visual and behavioral patterns unchanged, while affecting his/her underlying state, which can be measured through HR. This highlights the importance of multimodal sensing when performing classification in real-world scenarios.

Our system can be integrated with cameras and other vision systems (e.g., eye trackers) to provide more accurate detection of driver's activities and behaviors. Previous studies have developed deep-learning based methods for classifying driver's activities, and behaviors, as well as outdoor events using video cameras when the lighting conditions are suitable (76; 60; 37; 48). The integration of ubiquitous computing in such systems through sensor fusion techniques can help in edge cases of vision system applications (e.g., low illumination condition) and in cases that the video streams are not available due to different reasons (e.g., driver not allowing for video recording, lack of visibility due to lighting). For example, by integrating smartwatches into vehicles, a user can request between two levels of privacy. In a high privacy case, driver's activities, and behaviors are monitored through the smartwatch. In the other case, the user can choose to provide more data streams, in which the driver's video can be fused with the smart wearable to detect different activities, behaviors, and environmental attributes. In the case of low privacy mode, depending on the light level measured through an illuminance sensor (e.g., smartwatches often provide this sensor), the system can perform using both camera and smartwatch or by only relying on the smartwatch. All of these decisions and analysis can also be performed using edge computing devices that are already being used for autonomous vehicle development and data collection purposes such as NVidia Jetson family, which have onboard GPU and other computing units for real-

time analysis. Additionally, it should be considered that wearable data is much smaller than videos and requires lesser computational resources as compared to bigger video-based deep learning based models.

Our work can be improved from multiple aspects. First, by labeling more driving data, we can enhance the accuracy of our method for driver activity classification. Although we are using manual annotation for providing ground truths, we can optimize our labeling speed by using more automated approaches such as leveraging change point detection methods (103). By having more labeled data, we can then use other classification methods that require a higher amount of data (e.g., deep neural networks). Additionally, we will build and implement a system that can perform the modeling scheme provided here either on the watch, using an embedded system (e.g., Nvidia Jetson) in the vehicle, or on the cloud, to optimize our system for real-time applications (e.g., battery considerations). This can also help us make our system more human-centered by incorporating user feedback. Lastly, we will continue to test newer wearable devices to increase our number of sensor modalities (e.g., adding skin temperature), which can further increase the accuracy of our system.

4.2 Project 2: Driver State, Behavior, and Environmental Attributes Classification Using Unsupervised Techniques

4.2.1 Introduction

People have different preferences in their choice for driving styles (77). For instance, a recent study mentioned individual variability across participants when choosing between conservative and risky driving styles (77). Additionally, different studies have shown that personalization of autonomy has a positive effect on the user's perceived trust in autonomous vehicles (AV) (69; 95; 111; 108). This implies that AV's acceptance relies on correctly understanding people's preferences in driving styles and act accordingly. One method to identify individual responses to different driving behaviors is using historical driving data (25). For instance, driver-related physiological data (e.g., facial expressions, heart rate (HR), and skin temperature) can show how a driver reacts within different driving styles (109; 47; 70).

Analysis of historical driving data, which is often retrieved through naturalistic driving studies (NDS), also comes with problems such as the difficulty (time and cost) in analyzing the massive amount of collected information (58). NDS is often conducted in a longitudinal fashion to help detect behaviors while different environmental noise and challenges exist in the data (103), which drastically increases the amount of data that needs to be analyzed for detecting specific driver behaviors, actions, and responses. One method to address this issue is to apply unsupervised learning on both driving behaviors as well as the driver's state.

Through unsupervised learning, different patterns in driver's state and behaviors can be detected (58). These patterns, when analyzed together, may reveal important information regarding the driver's responses in each driving behavior.

Previous NDS data often lack information related to driver's physiological responses and cognitive metrics (19; 103). With the current advancements in wearable technology, it is now viable to detect driver's psychophysiological states through monitoring their heart rate (HR), skin temperature, skin conductance, arm movement, and other physiological metrics. In line with these improvements, recent studies have used driver's physiological responses such as changes in HR to detect driver's stress levels correlated with the increase in HR values (103; 110). Additionally, with the improvements in computer vision applications, it is now viable to use raw videos for estimating driver's breathing patterns, HR, gaze directions, and emotional features through facial landmarks (79). As a result of these improvements in technology, recent NDS are starting to implement novel sensors and technologies to collect such data (33; 110; 81; 103; 36). Utilizing these data streams and by applying unsupervised machine learning techniques, we can understand a driver's response within each driving behavior.

In this paper, we propose a methodology to classify driver's reactions to different vehicular maneuvers using unsupervised learning in a fully naturalistic environment. We use two case studies to analyze different driving patterns (e.g., harsh braking) coupled with driver state patterns (e.g., high stress level) extracted from a naturalistic driving dataset in an unsupervised fashion. Using the detected unsupervised patterns and by having access to multimodal human sensing data, we then analyze the driver's reactions under each driving pattern. Analyzing the fraction of each state pattern (i.e., HR and gaze pattern) within each driving behavior depicts that most of the drivers of our study are calmer in a more conservative driving style with route selection that can include more segments of highway driving.

4.2.2 Background Study

Previous studies have pointed out the utility of unsupervised learning methods in finding behavioral patterns in naturalistic driving data (58). However, past studies have not focused on applying the same methods on driver physiological data (e.g., HR). Below we first elaborate on the relationship between driver sensing measures (i.e., HR and gaze patterns) and driver's state (i.e., stress level, and workload). We then outline past progress in unsupervised modeling techniques for driver state and behavior detection.

4.2.2.1 Driver State Detection

Human bio-signals have been used extensively in multiple fields such as psychology, health sciences, and engineering for retrieving a deep understanding of a human's psychophysiological states (21; 94; 44; 75; 51; 71; 85). Psychophysiology refers to psychological states such as emotional responses (e.g., anger, frustration, and happiness), cognitive load, and distraction, which can be measured through changes in human physiological responses (e.g., HR, skin temperature, and skin conductance) (63; 57). In human-centered research, psychophysiological measures such as driver's HR (94; 98; 100), gaze patterns (8; 103), skin conductance (75), and brain signals (74; 31) were all used for retrieving driver's stress level, and cognitive load.

One popular definition of stress is the state of a human when the demand of a situation is perceived to be more than the available internal resources (34). Different bio-signals can then be used to detect stressful instances. Human cardiovascular measures have been extensively used in literature for detecting stress levels. In naturalistic environments and mainly driving research, cardiovascular activity can be retrieved through using either of Electrocardiography (ECG) or photoplethysmogram (PPG) technologies. ECG measures heart electrical activity through the usage of contact electrodes. PPG, on the other hand, records the same activity through measuring blood volume in the vein using infrared technology (63; 103). The cardiovascular measures can then be used to estimate features of the beat-to-beat signal of the heart such as HR itself, and root mean squared of successive intervals (RMSSD). Studies have demonstrated that higher stress levels are correlated with an increase in HR and a decrease in RMSSD (103; 49; 72).

Driver's cognitive workload is defined as cognitive resources that are taken from the driver by any activity other than the driving itself (29). In this definition, the cognitive load consist of mind wandering and the load imposed on the driver by secondary tasks. Secondary tasks are the ones that require attention but are not required for the task of driving, such as working with a phone (29). Human bio-signals such as driver's gaze measures, cardiovascular measures, and brain signals have been used for cognitive load estimation in both controlled environments and in naturalistic settings (63). Stationary and transition gaze entropy are two of the main eye gaze metrics that are commonly used for cognitive load estimation (54; 32; 90; 91). In information theory, the uncertainty associated with a choice is referred to as the information entropy (91) in which the more the uncertainty, the more the entropy and the more randomness in the system. There are two ways to calculate the entropy in gaze analysis. The first is generally calculated through Shannon's equation (86). In gaze analysis, Shannon's entropy shows the overall predictability of fixation locations in a sequence of gaze patterns, which is a measure of gaze dispersion (86) and is referred to as the Stationary Gaze

Entropy (SGE). Specifically, for a set of fixation locations in a sequence of eye movements, fixation locations can be assigned to spatial bins of p_i and calculate the SGE as:

$$SGE = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{4.1}$$

Different studies used SGE to infer a human's state in different conditions. For instance, SGE has been shown to be correlated with task difficulty, complexity, and experience with the task (91). Additionally, studies have used SGE in driving research to infer driver's state such as workload (91), drowsiness (90), and being under the influence of alcohol (89). For instance, (90) used driver's eye SGE to predict lane change events in sleep-deprived drivers, where an increase in SGE was associated with a higher probability of lane change events. While SGE was shown to be correlated with different human states, a recent review suggests that inferences based on changes in SGE can be very task-specific. For instance, if we know a specific task requiers a higher SGE and we observe that the participant is having a lower SGE, this may imply the participant could be disengaged from the task (91). This highlights the importance of the second measure of gaze entropy, referred to as conditional entropy, which is task-independent.

Conditional entropy considers the dependency between back-to-back fixation points in a temporal fashion. This results in the Gaze Transition Entropy (GTE). GTE is a measure of predictability of the next fixation location given the current location. For a sequence of transitions between different spatial bins of i and j, with a probability of p_{ij}, the GTE is calculated as:

$$GTE = -\sum_{i=1}^{n} p_i \sum_{j=1}^{n} p_{ij} \log_2 p_{ij} \tag{4.2}$$

GTE was shown to be correlated with multiple aspects of the human's state in both driving and non-driving research. In general, higher task demand, higher scene complexity, and higher levels of cognitive load were shown to increase the GTE (91; 32). Additionally, higher levels of GTE when performing the same task with the same scene complexity can be associated with higher stress levels (91).

4.2.2.2 Unsupervised Modeling of Driver's state and Behavior

Previous studies provided significant insight into analyzing driver's behaviors in naturalistic conditions through unsupervised methods. Bando et al. (11) have used multimodal LDA to infer driving topics using a combination of image sequences, annotated tags and driving behavioral data. In their study, authors first used a double articulation analyzer (DAA) for driving data segmentation. Then by applying LDA on the multimodal data, they were

able to achieve a dimensionality reduction of 5% on the raw data. This then led to an increase in classification accuracy achieved by a baseline that was trained using a support vector machine. In another study, the authors proposed a modification to DAA to not only perform segmentation but also to predict the duration of each segment (96). Based on the assumption that driving data has a two-layer hierarchical structure, the authors proposed a double-layer articulation structure model which uses a hierarchical Dirichlet process hidden semi-Markov model to predict the duration of each segment. Li et al. (61) used the density-based spatial clustering method to cluster physiological data of drivers into one of the normal, event, and noise clusters. Based on the data collected from three drivers and through an on-road controlled study, authors were able to achieve a recall rate of 75 percent in clustering the three categories.

Bender et al. (13) proposed an unsupervised method to provide high-level clusters for time series data streams from naturalistic driving behavior. In their study, the authors used a Bayesian multivariate linear model to segment the driving data. Then by using simplicial mixture models, they find unsupervised patterns associated with different driving behaviors such as acceleration and braking. This method was applied in an online fashion through an on-road controlled study. The data for the study was collected from a vehicle equipped with multiple telemetry sensors driving a 13 min route. Wang et al. (107) proposed a method to build a library of human's basic driving motion primitives that can be used for human-like actions. In their paper, the authors have used a probabilistic inference based on iterative expectation-maximization (EM) algorithm on driving data collected from one driver. Their probabilistic method achieves a more meaningful segmentation when compared with the classical EM-GMM approach by merging back-to-back clusters together, which is closer to the real-world driving situation. Lastly, (58) proposed a framework to automatically provide a description for driver's behavioral data retrieved through telemetry sensors inside the vehicle. In their method, they applied Bayesian multivariate linear regression to segment the driving data. Through using three different clustering methods of Gaussian mixture model LDA (GMM-LDA), Gaussian Wishart LDA (GW-LDA), and Multimodal LDA (mLDA), they found out that GMM-LDA provides the most useful description generation for naturalistic driving behavior data.

Although previous studies have provided significant insights into the application of unsupervised learning in driver states and behavior detection, most of them did not couple the two together to understand the driver's state in each behavioral pattern. This is extremely important as combining driver's states with behavioral patterns can provide a deeper description for each driving segment. For instance, while the unsupervised category for a driving segment might be "high acceleration", it is important to know whether the driver was car-

rying a high work load or not in that segment. For the future AV, such information might help with better prediction of driver's state in each driving scenario. Additionally, most of the past studies in driver state detection are performed in experimental conditions either in a driving simulator or in an on-road controlled study where the conditions are different from a fully natural environment. Such shortcomings mostly existed due to not having access to naturalistic human sensing data. This is now partly achieved through novel wearable technology devices that are implemented day after day in driving studies (see (103; 110). In this paper, by performing an unsupervised analysis on both driver state data collected through conventional wearable devices and driver's behavioral data collected through vehicle's kinematic data, we find driver's state (i.e., stress levels and workload) in each driving pattern.

4.2.3 Methodology

In this section the framework for retrieving state and behavior patterns is described. We then apply our framework to a fully naturalistic driving data collected through one of our previous studies (103). We first perform a data exploration using our framework on a driver's behaviors and states extracted through a 2 hour and 10 minutes long trip of a vehicle equipped with multiple sensors collecting contextual information including both driver and environmental sensing modules (i.e., case study I). We then apply our method to a larger pool of data collected from 12 participants (i.e., case study II). Below we first outline the methodology followed by the dataset details, the selected parameters.

In our framework, kinematic Sensors are the ones that record movements and acceleration in different directions such as an inertial measurement unit (IMU), which records acceleration and angular velocity in 3 different directions of X, Y, and Z. These information are used as driving behavior data such as vehicle's lateral and forward acceleration. On the other hand, human sensing modules are sensors specifically used for human related data such as smart watches and in-cabin and outdoor facing cameras. These data are used to detect drivers state such as stress level and workload.

The goal of this framework is to apply unsupervised learning methods on the data from **kinematic sensor** readings and **human sensing modules**, to detect **driving-related** and **driver state-related** patterns automatically. These patterns are then compared with known **driving behaviors** (e.g., harsh acceleration) and **driver states** (e.g., high stress level) to provide descriptions for each detected pattern. We then statistically compare different driving behaviors based on the proportion of each driver state within them.

Our framework consist of six main sections. As a summary, through a formerly proposed NDS framework, multiple kinematic and human sensing data-streams are collected (Fig. 4.6 - A). Driver's stress level and work load information are retrieved from human sensing data

(Fig. 4.6 - B). Similarly, driving behavior data is used to retrieve driving segments (Fig. 4.6 - B). Then through an unsupervised learning method, a driving behavior and a driver state pattern is generated for each driving segment (Fig. 4.6 - C and D). Lastly, within each detected behavioral pattern we assess the likelihood occurrence of each state patterns (Fig. 4.6 - E).

4.2.3.1 Details of Case Study Datasets

In our study we used two different datasets with different sensors and different resolutions to test the method. For case study I, we use the data collected from a 190 kilometer (km) driving scenario between cities of Charlottesville and Washington DC. This data is available through an online repository (106). Additionally, all of the code for our analysis is available in a GitHub repository (97). Fig. 4.7 shows the map of the trip. Based on the framework suggested in our previous work (103), the vehicle has a dash camera which collects both in-cabin and outside environment as well as vehicle's speed.

The videos collected from the in-cabin camera is fed into the OpenFace software (10) to retrieve driver gaze patterns at 30 frame per second resolution. OpenFace can perform facial landmark detection and gaze direction analysis on one or multiple people within a frame. OpenFace uses Conditional Neural Fields (CLNF) (9) for facial landmark detection. As an open source off-the-shelf software, it has achieved the least error in landmark detection, gaze estimation and head pose estimation benchmarks. Specifically, it has achieved an error of 9.96 % on the MPIIGAZE dataset (112).

Additionally, the driver is provided with a smartwatch that collects HR data, hand movement (through IMU sensor), and GPS locations. Also, we used an IMU sensor in the vehicle that collects vehicle's kinematic data including acceleration, linear acceleration as well as rotational velocity in 3 different directions. The detail of data collection and setup is provided in (103).

In case study II, we use driving data collected from 12 participants through our previous study, HARMONY (103). From each participant, we randomly chose a continuous highway trip that lasted approximately 90 minutes. In HARMONY, we did not have access to IMU sensors inside the vehicle. Instead, we calculated vehicle's forward acceleration through applying a gradient on the speed data retrieved from GPS on the camera. Although this GPS information was sampled at a much lower frequency (i.e., 1 HZ), it provides significant insights into differences across participants.

4.2.3.2 Driver State Feature Extraction

We use driver's raw HR directly for driver stress inference as previous studies showed the positive correlation between driver's HR and stress levels (see section 4.2.2.1, and Fig. 4.6

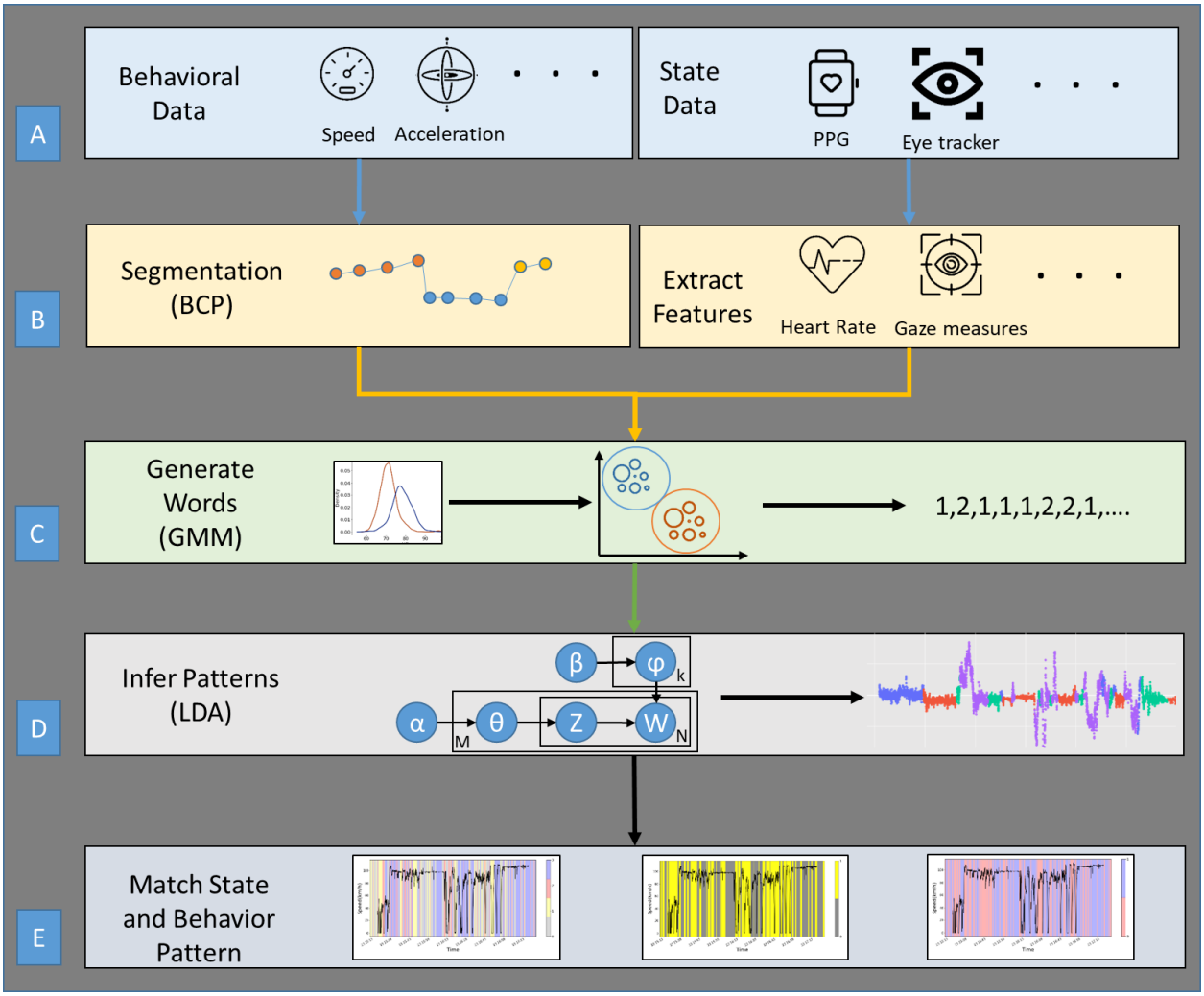


Figure 4.6: An overview of the framework of analysis consisting of six main sections. At first we retrieve features from driver's state data that align with stress level and work load (A and B). Additionally, we perform segmentation on behavioral data that provides specific segments in driving (B). We then perform unsupervised analysis through GMM-LDA method to find different patterns in driver's behavior and state. We first apply GMM to find word-like objects in the continuous inputs (C). We then apply the LDA on the sequence of word like objects in each segment to find different patterns in state and behavior (D). Lastly, we assess the presence of each state pattern within in driving behavioral pattern to find the more suitable driving behavior pattern to the driver.

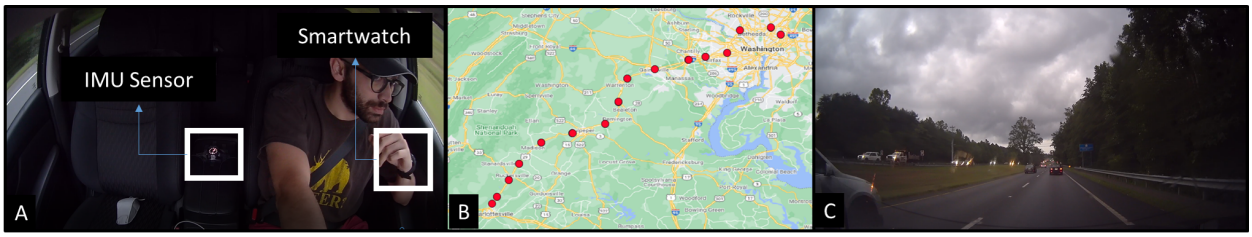


Figure 4.7: An overview of the case study dataset. The vehicle is equipped with a camera, smart watch, and an IMU sensor (A, and C). The data is collected from a 120 mile driving scenario between cities of Charlottesville and Washington DC, with a duration of 2 hour and 10 minutes (B).

- B). Moreover, our previous study in driving environments showed that stressful events on the road might change the distribution of driver's HR momentarily, and increase the HR for a short period of time (103).

As described in section 4.2.2.1, driver's gaze patterns can be used for workload estimation (Fig. 4.6 - B). For this purpose, from the driver's gaze data, we retrieve the GTE. For this task, we first construct a 2D space of the range of the gaze angles (53; 104). The 2D space is divided into equally distanced areas of interest (AOI) which can be a $4 * 4$ grid. The GTE is then calculated based on a rolling window of size 240s on the driving data (53). As a summary, for a sequence of AOIs, we find the transition matrix between AOIs by assuming a first-order Markov process for the gaze sequences (53). To this end, a transition matrix is retrieved with p_{ij} being the probability of switching between AOIs i and j (in S), with a stationary probability of π_i. Then the GTE is calculated based on Shannon's entropy as:

$$\hat{H}_t = -\sum_{i \in S} \pi_i \sum_{j \in S} p_{ij} \log_2 p_{ij} \tag{4.3}$$

4.2.3.3 Segmentation

Each driving scenario can be imagined as a combination of multiple segments. These segments are either separated using road objects (e.g., intersections and stop signs) or road users (e.g., a slow lead vehicle). In order to analyze the driving behavior data, it is first required to find different segments of interest (58). This is mostly because behaviors vary across segments, in which a person might be comfortable with different levels of acceleration/deceleration in a city segment, while the same acceleration/deceleration level may not feel suitable for a highway segment. We propose using Bayesian Change Point detection for driving behavior data segmentation in which different driving segments can be separated using change points in the distribution of vehicle's kinematic data (e.g., sudden acceleration)

(Fig. 4.6 - B). In this regard we first provide a brief overview on Bayesian Change Point detection method. A vector of multiple kinematic signals is then used as the input to the Bayesian Change Point detector.

4.2.3.3.1 Bayesian Change Point Detection (BCP)

Previous studies have pointed out the utility of BCP in detecting changes in the underlying distribution of data in different fields such as health (64), epidemiology (46), ecology (18), and transportation engineering (103). More specifically in transportation engineering, (103) used BCP to detect changes in driver's HR in naturalistic scenarios, which might be indicative of presence of stressors on the road. In another study Maye et al. (66) also used BCP to find different segments in vehicle's telemetry data. To perform BCP, we use the Bayesian change point model provided in Barry and Hartigan's book (12). In summary this model assumes that the mean of the input (e.g., acceleration of the vehicle) within different segments remains constant. Below we formally define BCP for a univariate case. The reader is also referred to (12; 66; 30) for more information and details on this method.

Assuming we have n input data point (e.g., vehicle's acceleration data points) $\{a_1, ..., a_n\}$, we will use a_{ij} to refer to the accelerometer data between indices i and j. Let $\rho = (U_1, ..., U_n)$ indicate a segment of the time series inside the non overlapping acceleration segments. A boolean array of change points is then used to to denote the segments. At each time step, if U_i takes a value 1, we have a new acceleration segment; else we remain in the same segment.

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

$$f(\rho|X) \propto f(X|\rho)f(\rho) \tag{4.4}$$

The Prior cohesion density can be retrieved as follows: Let p denote the probability of finding a new change point at each data point. We assume this probability to be the same at each data point. If we assume that there are b segments, the prior cohesion density can be written as

$$f(\rho|p) = p^{b-1}(1-p)^{n-b} \tag{4.5}$$

The joint density of observations and parameters given ρ is a product of densities of different segments in ρ. Let us consider a single segment. If we assume that the data in this block is generated by a Gaussian process with mean θ and variance σ^2, let the prior density of θ be a Gaussian with mean μ_0 and variance σ_0^2

$$
\begin{aligned}
f(X_{ij}, \theta) &= \Pi f(X_k|\theta) f(\theta) \\
f(X_{ij}) &= \int \Pi f(X_k|\theta) f(\theta) d\theta
\end{aligned}
\tag{4.6}
$$

We can then simplify the above integral to:

$$
f(X_{ij}) = (\frac{1}{2\pi\sigma^2})^{(j-i)/2} (\frac{\sigma^2}{\sigma_0^2 + \sigma^2})^{1/2} exp(V_{ij}) \tag{4.7}
$$

Where

$$
V_{ij} = -\frac{\sum_{l=i+1}^{l=j} (X_l - \hat{X}_{ij})^2}{2\sigma^2} - \frac{(j-i)(\hat{X}_{ij} - \mu_0)^2}{2(\sigma^2 + \sigma_0^2)} \tag{4.8}
$$

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:

$$
\begin{aligned}
f(\mu_0) = 1, -\infty \leq \mu_0 \leq \infty \\
f(p) = 1/p_0, 0 \leq p \leq p_0 \\
f(\sigma^2) = 1/\sigma^2, 0 \leq \sigma^2 \leq \infty \\
f(w) = 1/w_0, 0 \leq w \leq w_0
\end{aligned}
\tag{4.9}
$$

$$
f(X|\rho, \mu_0, w) = \int_0^\infty 1/\sigma^2 \prod_{ij \epsilon P} f(X_{ij}) d\sigma^2 \tag{4.10}
$$

After integrating out μ_0 and w, This can be simplified to the indefinite integral below. The full derivation is provided in (12).

$$
f(X|\rho) \propto \int_0^{w_0} \frac{w^{(b-1)/2}}{(W + Bw)^{(n-1)/2}} dw, \tag{4.11}
$$

where

$$
\begin{aligned}
\hat{X} = \sum_{i=1}^{n} X_i/n, \; B = \sum_{ij \epsilon P} (j-i)(\hat{X}_{ij} - \hat{X})^2, \\
W = \sum_{ij \epsilon P} \sum_{l=i+1}^{l=j} (X_l - \hat{X}_{ij})^2
\end{aligned}
\tag{4.12}
$$

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

$$f(\rho) \propto \int_0^{p_0} p^{b-1}(1-p)^{n-b}dp \tag{4.13}$$

To calculate the posterior distribution over segments, Markov Chain Monte Carlo (MCMC) is used (38). A Markov chain is then defined 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 situations of with and without the change point in location i.

$$\begin{aligned} \frac{p_i}{1-p_i} &= \frac{p(U_i = 1|X, U_j, j \neq i)}{p(U_i = 0|X, U_j, j \neq i)} \\ &= \frac{\int_0^{p_0} p^b(1-p)^{n-b-1}dp}{\int_0^{p_0} p^{b-1}(1-p)^{n-b}dp} x \frac{\int_0^{w_0} \frac{w^{b/2}}{(W_1+B_1 w)^{(n-1)/2}}dw}{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0+B_0 w)^{(n-1)/2}}dw} \end{aligned} \tag{4.14}$$

Finally, this is simplified to a probabilistic model with the following two parameters p_0 and w_0.

We use the package *bcp* (30) in R programming language to implement our change point analysis. For the case study I, we use vehicle's kinematics as the main segmentation input. To this end, we use the input vector X for segmentation algorithm as combination of forward acceleration (a), lateral acceleration (l), and lateral rotational speed (ω_Z):

$$X = \{a, l, \omega_Z\}$$

The vector X is then fed into BCP. The important point with respect to BCP is that it does not require to know the number of segments. This helps with longitudinal data where the actual number of driving segments is unknown.

In order to illustrate the results of segmentation we use the vehicle's speed data. Note that the speed data was not used as the input into the segmentation as it is collected at 1 Hz, which is much lower than the IMU sensor (i.e., 100 Hz) and have not been used as the input to the segmentation. We use the speed data mostly due to the fact that speed can naturally show different segments of driving such as slowing down due to a lead vehicle, stopping for an intersection, and turning (Fig. 4.8). This helps with a better visual inspection of our results.

For case study II, as we did not have access to an IMU sensor, we use vehicle's speed data, as well the acceleration, which is retrieved through applying a derivative on the speed data. The segmentation is performed separately for each participant.

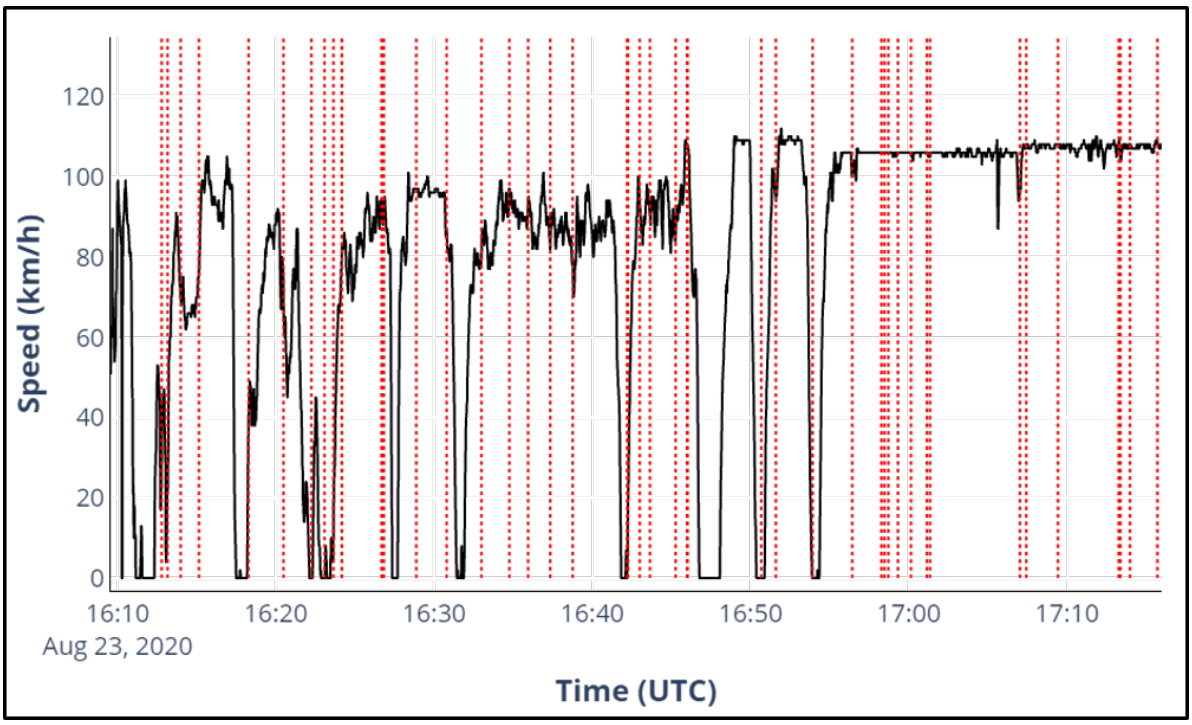


Figure 4.8: Different segments of driving detected by the BCP applied on the kinematics data shown with dashed vertical lines.

4.2.3.4 Pattern Inference

In order to perform pattern inference we use a combination of Gaussian Mixture Modeling (GMM) and Latent Dirichlet Allocation (LDA), referred to as GMM-LDA (3; 2; 80; 58). GMM-LDA for topic inference application of driving behavior data has shown significant advantage as compared to other unsupervised clustering methods such as GW-LDA and M-LDA in the past (3; 58). LDA is a text analytics probabilistic hierarchical Bayesian model, which is used to infer topics from a set of words (15). When applying it to driving data, the goal is to provide a main driving pattern and state pattern for each segment of driving. To do so, we first need to generate 50 word-like objects (i.e., discretize the data) from the continuous raw vehicle telemetry, gaze, and HR data using GMM (Fig. 4.6 - C). GMM-LDA method then uses LDA on top of GMM objects to produce topics for each segment of driving, which can be considered a driving pattern (Fig. 4.6 - D). Below we formally outline this method.

4.2.3.4.1 GMM-LDA

In LDA, each word is modeled as a finite random mixtures over an underlying set of k topics and each topic is represented as a distribution over a set of words (15). A word (w) is

defined as a basic unit of discrete data which is drawn from a set of vocabulary. In this way, a document is then a sequence of N words $N = \{w_1, w_2, ..., w_n\}$ and a corpus is a set a of M documents (15). The details of LDA method is provided in appendix (section 4.2.8.1).

Historically LDA is designed for inferring topics from a set of documents and is based on text type data, which is by nature a discrete type data. When considering continuous data sources (e.g., vehicle's acceleration), multiple methods were proposed to overcome this issue. A more common approach is to apply GMM clustering method on the continuous data to generate word-like objects (3; 2; 80; 58). GMM is a clustering method that identifies a mixture of multidimensional Gaussian probability distributions that can best describe the input vector (82). For a formal description of GMM please review appendix (section 4.2.8.2).

We first apply GMM algorithm to obtain word-like objects (i.e., descretizing data) from the multimodal vehicle kinematic data, which includes forward acceleration (a), lateral acceleration (l), and lateral rotational speed (ω_Z) (Fig. 4.6 - C). These inputs are the same inputs that were used in segmentation section (4.2.3.3). Based on the previous literature we chose 50 as the number of words (58). Also, note that using Bayesian Information Criterion (BIC) for defining the number of components, we observed very small enhancements (< 0.0001) in the BIC value after increasing the number of words above 50.

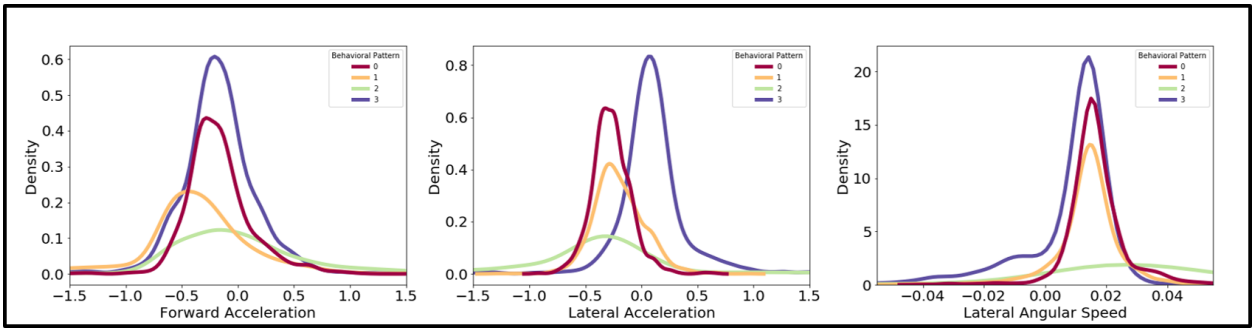


Figure 4.9: Distribution of different kinematic sensor readings within different driving patterns. Note that differences across patterns in the kinematic sensor distributions, for instance, pattern 3 has a more positive lateral acceleration than the other 3 patterns. Also note that for some patterns the difference is only visible through one sensor such as the comparison between patterns 0 and 3 lateral acceleration.

Based on the generated GMM distribution, we then apply LDA to find the driving pattern that best describes each segment in the driving scenario (Fig. 4.6 - E). For LDA, we have used the default values of Scikit Learn for the α and β which are equal to the reverse of number of topics. When using LDA, we need to define the number of topics prior to analysis. Driving scenarios often consist of different patterns (58). For instance, braking behavior includes a limited number of patterns which can be high negative acceleration (i.e., harsh brake), or

a low negative acceleration (i.e., normal brake). Thus it is likely that there are more than one driving pattern present in an individual's driving data (58). While previous studies suggested the number of patterns to be 4 we also tested different numbers ranging from 2-6 both quantitatively and qualitatively (13; 58). First, through using the perplexity metric of an LDA model (113; 52), we find the location that there is a change point in the rate of change in perplexity for different numbers of clusters. This is similar to finding an elbow in the perplexity for different number of clusters. This results in 4 clusters. In addition to quantitative analysis, we observed that with increasing the number of topics in LDA (patterns in driving) more than 4, some of the topics become very similar from the visual inspection of the distributions so it becomes impractical to have them as separate patterns. For instance, moving from 4 to 5 topics, two of the generated topics will be similar in their mean and standard deviation of acceleration, and speed. Additionally, reducing the topics from 4 to 3 misses the difference in the braking patterns. These observations, confirms the quantitative analysis.

For the case study II, we also performed segmentation using vehicle's speed and acceleration. Then for each segment we followed the GMM-LDA method and applied it to the driving behavior as well as driver state data to find patterns of behaviors and state in each segment. Below we first present the results for case study I, which helps better illustrate how the method is applied. We then discuss the results of applying the same method to a larger pool of participants.

4.2.4 Results

In this section we present the results of applying the GMM-LDA method with Bayesian Change Point segmenting on two separate case studies. As described in section 4.2.3.1, the first case study is from one participant with high resolution kinematic data, while the second case study is from 12 participants vehicles that had low resolution kinematic data.

4.2.4.1 Results - Case Study I

In this section, we showcase how GMM-LDA method can be applied on driver behavior and state data to find patterns in driver's behavior (e.g., harsh braking) and states (e.g., high stress level). We first discuss the patterns detected in vehicle kinematic data. We then move to driver's state data (e.g., HR). Fig. 4.9 shows the distribution of kinematic data across the four different detected driving patterns (refer to section 4.2.3.4 for details on pattern recognition). As shown on this figure, each behavior pattern has its own distinct distribution. For instance, when comparing patterns 3 and 2, we observe the difference in the location of peak in the forward acceleration as well as lateral acceleration. However, comparing across patterns 0 and 3, we only observe a major difference when comparing the

distribution of their lateral acceleration.

Additionally, Table 4.5 provides the detailed statistics of each driving pattern. The first step is to confirm that the driving patterns detected by GMM-LDA are as a result of different underlying distributions and are statistically different. The Kolmogorov-Smirnov test between different distributions is performed (65). The pairwise comparison of all distributions were significant at $p < 0.05$ level and is shown on Table 4.4. Below we elaborate in detail on each driving pattern detected in the kinematic data and how they relate to certain predefined driving behaviors.

Table 4.4: Details of statistical tests on different patterns detected through unsupervised modeling

Source	Comparison	KS Test Statistic	p-value	Significant at 0.05?
Forward Acceleration	0&1	0.3369	<0.0001	Y
	0&2	0.2467	<0.0001	Y
	0&3	0.0622	0.0003	Y
	1&2	0.3447	<0.0001	Y
	1&3	0.3021	<0.0001	Y
	2&3	0.2133	<0.0001	Y
Lateral Acceleration	0&1	0.1670	<0.0001	Y
	0&2	0.2761	<0.0001	Y
	0&3	0.7096	<0.0001	Y
	1&2	0.2883	<0.0001	Y
	1&3	0.5568	<0.0001	Y
	2&3	0.6054	<0.0001	Y
Angular Velocity	0&1	0.1201	0.0003	Y
	0&2	0.4365	<0.0001	Y
	0&3	0.2642	<0.0001	Y
	1&2	0.4520	<0.0001	Y
	1&3	0.1488	<0.0001	Y
	2&3	0.5152	<0.0001	Y

Considering the differences in sensor readings across the driving patterns, we find dominant differences that can be associated to certain well-known driving behaviors. Driving patterns 0 and 1 both exhibit high negative forward acceleration (deceleration) values (-0.1667 and -0.4180) respectively. These two driving patterns are mostly related to braking behavior. The more negative value of mean forward acceleration in driving pattern 1 suggest a harsher braking pattern. Note in Table 4.5 the mean speed of the two driving patterns are similar at 76.27 and 76.61 km/h respectively. This provides evidence for the two patterns to

be similar in nature, as both are related to braking. We refer to pattern 0 as normal braking behavior, and pattern 1 as harsh braking behavior. Driving pattern 2 has a relatively flat forward acceleration distribution including a broad range of forward accelerations with a mean value 0.032 which is close to zero. Due to exhibiting an average higher rotational velocity (0.027), and a broader lateral acceleration distribution, this driving pattern might be associated to normal road curvatures. We refer to this driving pattern as road curvature driving behavior. Driving pattern 3 has the highest mean speed value (89 km/h), with a positive forward acceleration (0.095). This driving pattern also exhibits one large peak at a positive lateral angular speed at 0.0087 which is very close to zero indicating straight driving. This pattern can be associated to normal free flow driving in the highway, and we refer to it as the highway free flow driving behavior. A sample of such behaviors are also shown on Fig. 4.10, where snapshots of different behaviors (e.g., free flow highway driving) are shown from the dataset.

Table 4.5: Detailed statistics of forward acceleration, lateral acceleration, angular velocity, and speed of each recognized behavior. Note that different patterns resemble well-known driving behaviors.

Driving Behavior Data	Statistical Index	Driving Pattern 0 (Normal Braking)	Driving Pattern 1 (Harsh Braking)	Driving Pattern 2 (Road Curvature Driving)	Driving Pattern 3 (Free Flow Driving)
Forward Acceleration	Mean	-0.16677	-0.41807	0.032059	-0.17298
	Standard Deviation	0.343867	0.617844	0.677998	0.397301
	25th percentile	-0.34826	-0.61339	-0.41371	-0.3527
	Median	-0.20286	-0.40094	-0.08027	-0.18616
	75th percentile	-0.02847	-0.13217	0.303769	0.0067
Lateral Acceleration	Mean	-0.26936	-0.20002	-0.35933	0.095032
	Standard Deviation	0.205908	0.263469	0.774748	0.347671
	25th percentile	-0.38046	-0.36103	-0.66054	-0.0605
	Median	-0.27365	-0.21968	-0.345	0.077288
	75th percentile	-0.15625	-0.04575	-0.08588	0.221336
Rotational velocity	Mean	0.016956	0.01432	0.027152	0.008777
	Standard Deviation	0.012567	0.012698	0.073349	0.018666
	25th percentile	0.011648	0.00877	0.012364	0.004099
	Median	0.015752	0.014653	0.02795	0.012516
	75th percentile	0.021457	0.020216	0.047836	0.01734
Speed	Mean	76.27409	76.61246	79.16773	89.84004
	Standard Deviation	36.47230	29.24584	27.87107	27.15401
	25th percentile	65	53	67	90
	Median	97	88	89	98
	75th percentile	98	97	99	106

Additionally, the sequence of different detected patterns in kinematic sensors can provide insights into driver's behavior. We discuss the findings when analyzing the sequence of behaviors through a case study and a transition matrix of behavior sequences. First through a case study, Fig. 4.11 depicts the sequence of driving behaviors through parts of the driving scenario. As visually observed, in locations that speed varies sporadically, the diver behavior also switches more often (e.g., 15:43 until 15:46 shown with a blue dashed box on Fig. 4.11).

Figure 4.10: Samples of the detected driving patterns. Patterns 0 and 1 are associated with normal and harsh braking behavior, while pattern 2 is associated to curve driving behavior, and pattern 3 is related to highway free flow driving behavior.

To quantify these sequences, as shown in Table 4.6, the probability of transition between the different driving behaviors as recognized by our method. Each element of the table at location i, and j shows the $p_{i,j}$ transition probability between the elements i and j. As shown on the table, when in normal braking behavior, the chances of switching to behaviors of road curvature driving behavior and free flow driving is higher than harsh braking behavior. However, when being at harsh braking behavior, the driver is most likely to switch to normal braking than other two behaviors. Looking at the free flow driving behavior, the driver is most likely to continue with the high speed highway driving ($p_{3,3} = 0.55$), and the probability of switching to a harsh braking behavior ($p_{3,1} = 0.13$) is equal to the normal braking behavior, while being less than switching to the curve driving behavior. A sample of a probable sequence can be to start from free flow highway driving, switch to road curvature driving and attempt a harsh brake. For a better illustration, such a sample sequence is shown with black dashed box on Fig. 4.11.

The GMM-LDA method was then applied to the driver's gaze and HR data. We considered two patterns for these two modalities relating to normal and abnormal (high) for HR, and low and high for gaze entropy, respectively. Fig. 4.12 shows the different driver state patterns detected through driver's gaze entropy and HR data. We first confirmed that the distribution of normal/abnormal HR and low/high gaze are statistically significant. Similar to previous sections this is performed through Kolmogorov-Smirnov test (65) and the results of comparison between different distribution where significant at $p < 0.0001$ level. Table 4.7 shows the significance of the tests across the different distributions.

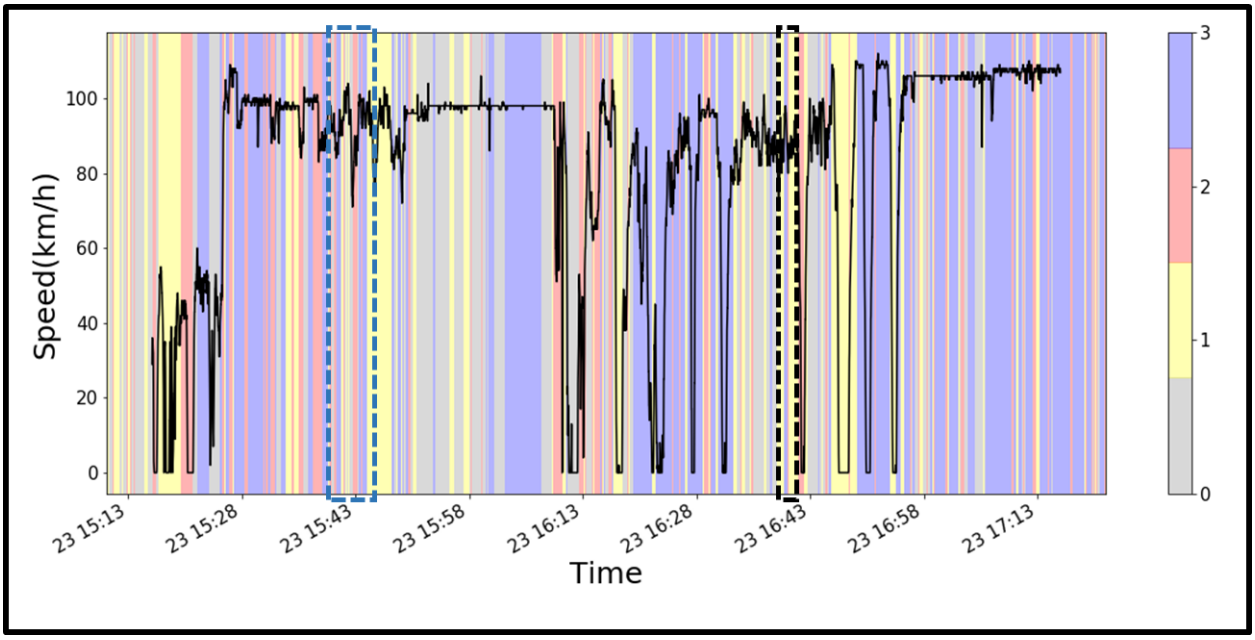


Figure 4.11: The change in driving patterns throughout the different segments of driving. Each segment is associated with one of the four driving patterns as described previously. Note that when speed changes more sporadically, the behavior also changes more often (e.g., blue dashed box). Also, the black dashed box shows a sequence of sample behavior starting from free flow highway driving, switching to road curvature driving, and finishing with a harsh brake.

Table 4.6: Transition matrix between different driving behaviors for case study I dataset.

Driving Behavior	**Normal Braking**	**Harsh Braking**	**Road Curvature Driving**	**Free Flow Driving**
Normal Braking	0.28	0.21	0.26	0.26
Harsh Braking	0.27	0.27	0.21	0.25
Road Curvature Driving	0.25	0.27	0.26	0.21
Free Flow Driving	0.13	0.13	0.19	0.55

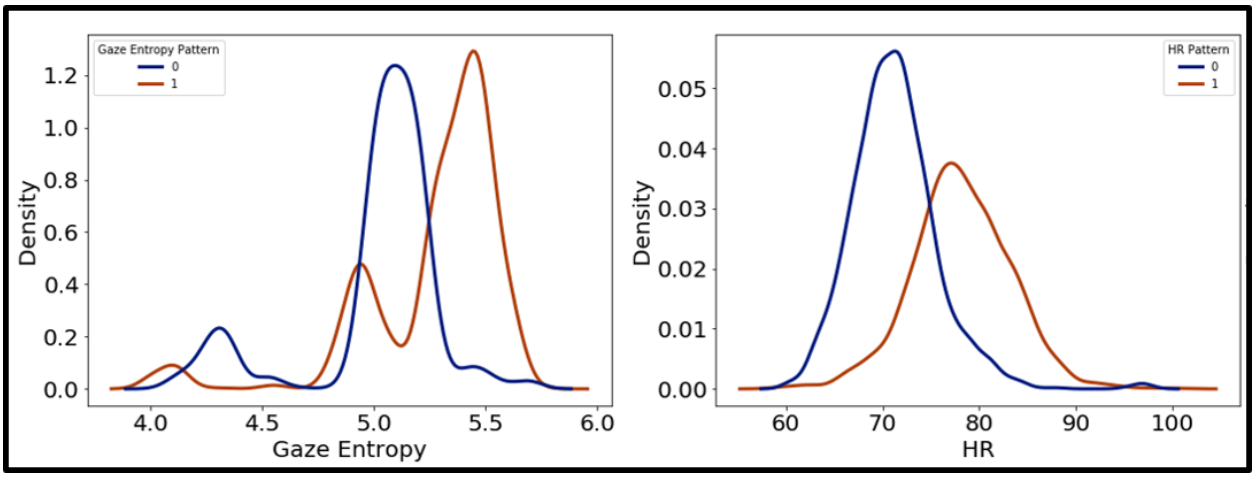


Figure 4.12: Distribution of different driver state patterns detected through HR and gaze entropy. Note the differences across the distributions pointing to different states for the driver

Table 4.7: Details of statistical tests across different distributions of HR and gaze data

Driver State Data	Comparison	KS Test Statistic	p-value	Significant at 0.05?
HR	Abnormal & Normal HR	0.5999	<0.0001	Y
Gaze	Low & High Gaze Entropy	0.6688	<0.0001	Y

Driver state pattern 0 in HR is related to normal HR values, which has a mean value of 71.23 bpm with an standard deviation of 4.47 bpm. We refer to this driver state pattern as the normal HR or calm state. On the other hand, driver state pattern 1 is related to high HR with a mean of 78 bpm and a standard deviation of 5.13 bpm, which we refer to as abnormal HR depicting a situation of having high stress driver state. As shown on Fig. 4.12, density of the abnormal HR is lower, indicating the less amount of time spent in this state pattern throughout all observations. This is also in line with our previous study (103) showing that driver's HR through stressors on the road increases from it's baseline values for a short period of time and moves back to it's baseline.

Fig. 4.13 shows the fraction of normal and abnormal HR in the recognized driving patterns such as normal brake, and harsh brake. As shown, the driver state pattern with abnormal HR happens more often than the normal HR in the road curvature driving behavior. A Chi-Square test (67) also shows that the ratio between the counts of each state is significantly different than being equal (i.e., ratio of 1) with a $\chi^2 = 43.87$ and a $p-value < 0.0001$. This figure also shows that the driver has more abnormal HR states during harsh brake compared to normal brakes, in which a Kruskal-Wallis test (55) confirms the figure with a $statistic = 404.57$ and a $p-value < 0.0001$. Considering that normal HR indicates a calmer

Table 4.8: Detailed statistics of HR, and gaze for each recognized state pattern

Driver State Data	Statistical Index	State Pattern 0 (Normal HR/ Low Gaze Entropy)	State Pattern 1 (Abnormal HR/ High Gaze Entropy)
HR	Mean	71.23	78.13
	Standard Deviation	4.47	5.13
	25th percentile	68.4	75
	Median	71	78
	75th percentile	73.7	81.4
Gaze	Mean	5.01	5.27
	Standard Deviation	0.30	0.31
	25th percentile	4.99	5.12
	Median	5.08	5.36
	75th percentile	5.17	5.47

driving state, this implies the driver is calmer during normal braking behavior compared to harsh braking behaviors. Additionally, the driver had a higher level of normal HR as compared to abnormal HR in free flow driving behavior, in which we confirmed this through a Chi-Square test with a $\chi^2 = 558.6$ and a $p - value < 0.0001$.

Driver state pattern 0 in gaze with a mean of 5.01, is related to samples of gaze with lower GTE. On the other hand, driver state pattern 1 has a mean value of 5.27. As mentioned in section 4.2.2, higher GTE in pattern 1 might be associated to higher task demand, higher scene complexity, and work load (Table 4.8).

Fig. 4.14 shows the fraction of low and high GTE in the recognized driving patterns such as normal brake, and harsh brake. As shown, in the free flow driving, low GTE pattern happens more often than high GTE which might indicate lower work load for this driver during free flow driving. A Chi-Square test (67) shows the counts of each low and high GTE within free flow driving is significantly different than equal probability (i.e., ratio of 1) with a $\chi^2 = 587.26$ and a $p - value < 0.0001$. Also, the ratio of high to low GTE in normal braking is higher than that of harsh braking, which implies that the probability of having a higher workload in normal braking is more than harsh braking. This is confirmed with A Kruskal-Wallis test (55) indicating a $statistic = 88.31$, and a $p - value < 0.0001$. This can be due to the fact that in normal braking the period of time that the driver is in the process of braking is higher, thus driver's GTE and the associated workload are more likely to be higher.

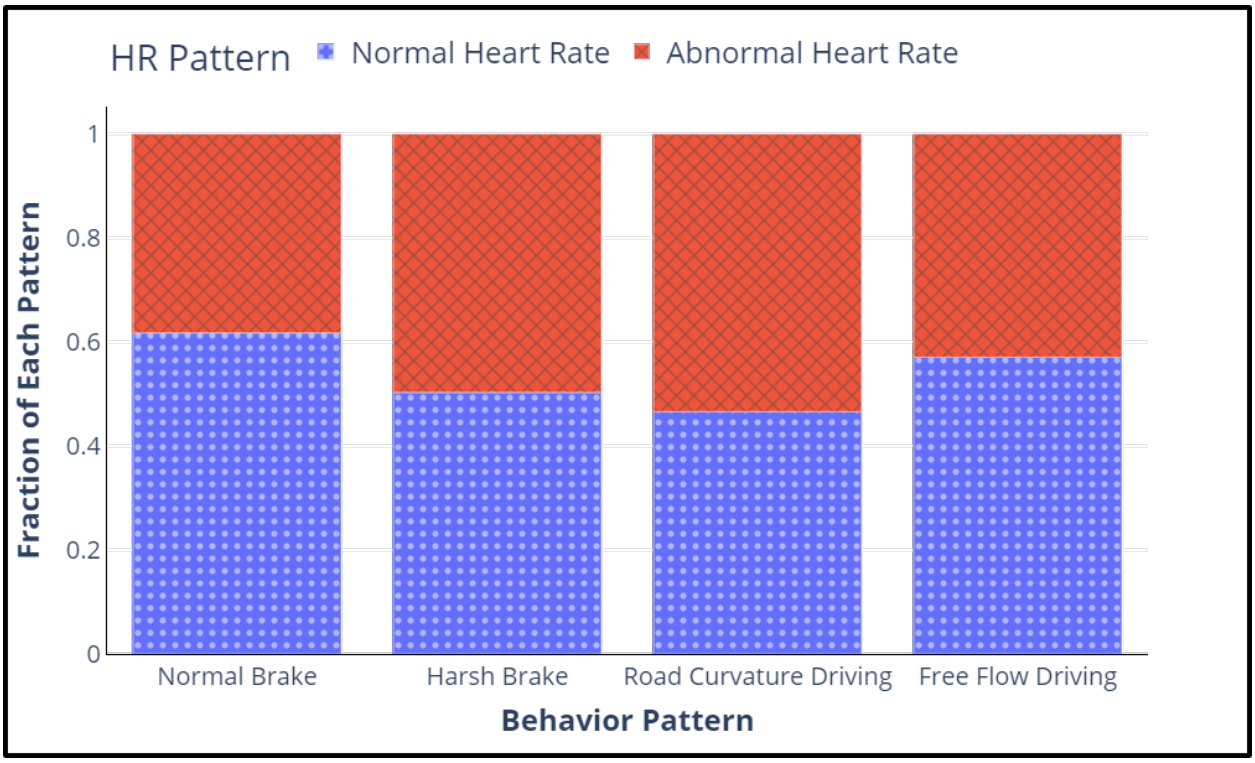


Figure 4.13: The co-occurence of driving clusters with clusters of HR

4.2.4.2 Results - Case Study II

Similar to the case study I, we performed segmentation, clustering and pattern inference on the case study II dataset. Note that case study II includes kinematic data from vehicle's speed and acceleration from 12 participants. Fig. 4.15 depicts the distribution of speed and accelerometer data across the different detected driving patterns. Similar to case study I, each of the different detected patterns exhibit distinct characteristics. For instance, comparing across patterns 0, and 1, we observe the difference in the location of the peak in the distribution. In order to confirm that the patterns are statistically different, we have applied the Kolmogorov-Smirnov test (65). The results of the pairwise tests across patterns are depicted in Table 4.9 and are all statistically significant.

Considering the differences in vehicle's speed and acceleration across the driving patterns, we find distinct differences associated to certain well-known driving behaviors, such as differences in mean and standard deviation as depicted in Table 4.10. Similar to case study I, driving pattern 0 is related to normal brakes, with a negative acceleration (-0.006). Additionally, driving pattern 1 is related to harsh brakes with the highest negative acceleration and the lowest mean speed as compared to all the other patterns. The high value of the standard deviation of speed (42.01) in the driving pattern 1 confirms that this behavior

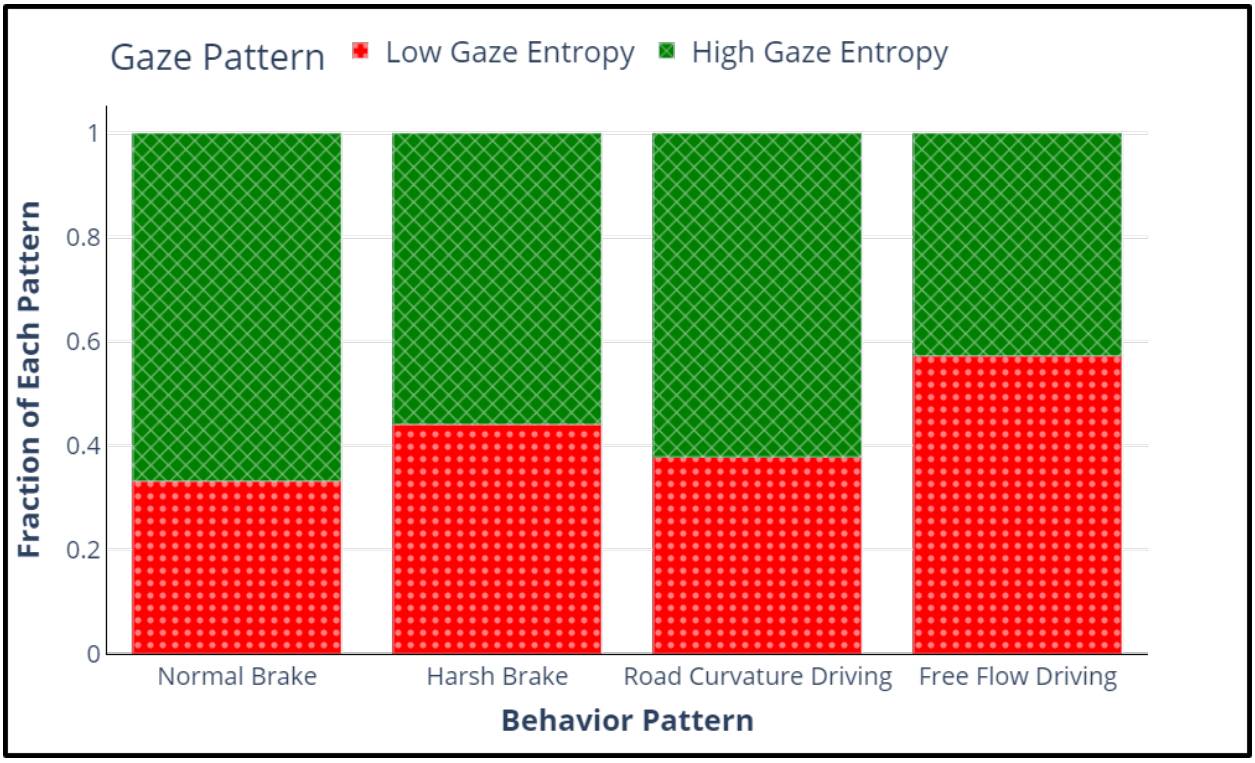


Figure 4.14: The co-occurence of driving clusters with clusters of gaze entropy

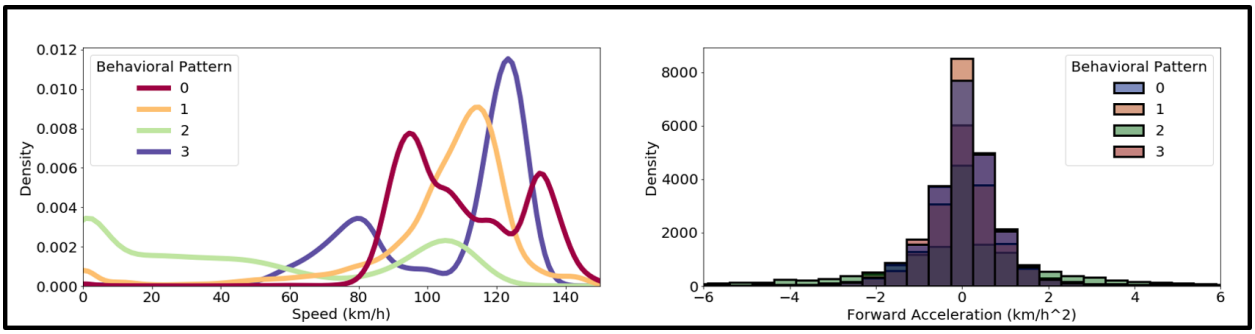


Figure 4.15: The distribution of forward acceleration and speed across different patterns retrieved unsupervised from 12 participants. Note that because we retrieved accelerometer data from the low resolution speed, the distributions are less smooth as compared to case study I. Thus, we are visualizing the differences across driving patterns with a histogram instead of a kernel density estimation (KDE).

Table 4.9: Details of statistical tests performed on the detected driving patterns. The tests shows that the distributions of speed and accelerometer data are significantly different across the detected patterns.

Source	Comparison	KS Test Statistic	p-value	Significant at 0.05?
Forward Acceleration	0&1	0.04	<0.0001	Y
	0&2	0.08	<0.0001	Y
	0&3	0.08	<0.0001	Y
	1&2	0.08	0.0003	Y
	1&3	0.09	<0.0001	Y
	2&3	0.02	<0.0001	Y
Speed	0&1	0.7	<0.0001	Y
	0&2	0.9	<0.0001	Y
	0&3	0.9	<0.0001	Y
	1&2	0.9	<0.0001	Y
	1&3	0.8	<0.0001	Y
	2&3	0.5	<0.0001	Y

can happen at different speeds depending on the environment, and it can quickly change the speed from high to lower values. Driving pattern 2 is related to accelerating behavior, which has the highest positive acceleration (0.041). Similar to case study 1, driving pattern 3 exhibits a very low negative acceleration (-0.001) with the highest mean speed (110.368) and lowest standard deviation (18.78). This in fact, depicts the behavior through free flow driving where the driver keeps a constant speed while not accelerating/decelerating. Also, note the difference between driving pattern 0 and 3. Driving pattern 0 exhibits a higher standard deviation in speed relative to pattern 1 (23.4 vs. 18.74) with a lower mean speed value (107.195 vs. 110.368), respectively. These differences align with the difference between braking and highway driving behaviors. Please note that in case study II, we did not have access to lateral kinematic data thus we could not confirm the curve driving behavior across participants.

Similar to case study I, we analyze the sequence of behaviors through a transition matrix. Table 4.11 shows the probability of transitioning between different driving behaviors. As depicted, when drivers are in free flow driving, there is a much higher probability to switch to a normal brake as compared to a harsh brake (0.16 vs. 0.02). This might imply that drivers are more likely to continue a more gentle driving style if they are in free flow driving which has an acceleration very close to zero. However, once being in an accelerating behavior, the driver is more likely to switch to a harsh brake as compared to a normal brake (0.15 vs. 0.13). Among the 12 participants, this might imply that an accelerating behavior continues

Table 4.10: Detailed statistics of forward acceleration and speed in each recognized driving pattern. Note that each pattern exhibits certain characteristics that resembles a well-known driving behavior.

Behavioral Attribute	Statistical Index	Driving Pattern 0 (Normal Braking)	Driving Pattern 1 (Harsh Braking)	Driving Pattern 2 (Accelerating)	Driving Pattern 3 (Free Flow Driving)
Forward Acceleration (km/h^2)	Mean	-0.006	-0.045	0.041	-0.001
	Standard Deviation	0.805	1.259	1.291	0.523
	25th percentile	-0.118	-0.297	-0.259	-0.162
	Median	0.0001	-0.0006	0.0038	-0.0006
	75th percentile	0.108	0.233	0.353	0.126
Speed (km/h)	Mean	107.195	47.872	103.252	110.368
	Standard Deviation	23.404	42.014	25.964	18.784
	25th percentile	84	6	99	96
	Median	119	38	110	107
	75th percentile	124	96	117	128

Table 4.11: Transition Matrix between different driving behaviors detected through the case study II dataset

Driving Behavior	**Harsh Brake**	**Free Flow Driving**	**Normal Brake**	**Accelerating**
Harsh Brake	0.75	0.1	0.12	0.12
Free Flow Driving	0.02	0.65	0.16	0.16
Normal Brake	0.15	0.14	0.60	0.11
Accelerating	0.15	0.13	0.08	0.64

with a harsh decelerating behavior, which can be considered a more risky driving style with high negative/positive values of acceleration.

For state data, we first normalized each participant's HR and gaze entropy values to have a 0 mean and standard deviation of 1. This is performed due to the fact that the abnormal and normal HR as well as GTE values across people can be quite different. For instance, a person's normal HR might be 65 bpm while for another person, it can be 75 bpm. For a better illustration, distribution of HR data is also shown on Fig. 4.16 where different participants' HR distribution can be very different on the mean and standard deviation, implying different normal HR values across different people.

We then applied the GMM-LDA method to both HR and gaze data separately, for case study II. Fig. 4.17 depicts different driver state patterns retrieved through both HR and gaze data. Similar to case study I, HR has two patterns of normal versus abnormal. Additionally, gaze entropy has two patterns of low versus high. The patterns recognized are significantly

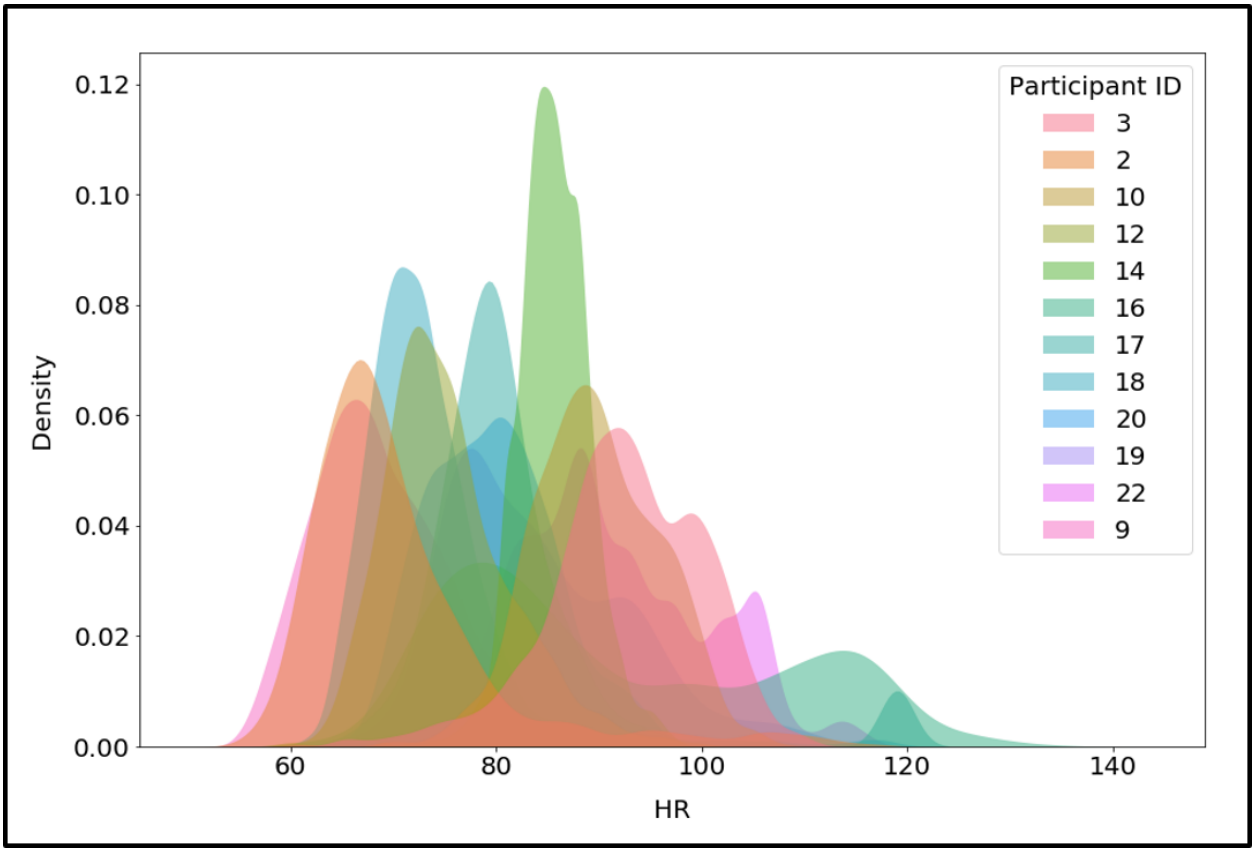


Figure 4.16: The distribution of HR for 12 participants. Note the differences across the distributions pointing to different baselines for different participants.

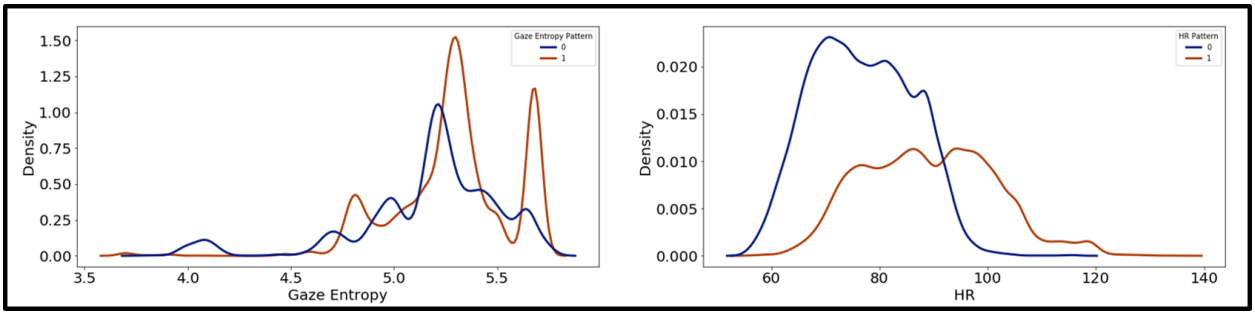


Figure 4.17: The distribution of gaze entropy and HR for 12 participants. Note the differences across the distributions pointing to different states for the drivers.

Table 4.12: Detailed of statistical tests across different patterns of HR and gaze

Driver State Data	Comparison	KS Test Statistic	p-value	Significant at 0.05?
HR	0&1	0.8	<0.0001	Y
Gaze	0&1	0.9	<0.0001	Y

different as tested through the Kolmogorov-Smirnov test (65). The results of the tests are depicted on Table 4.12.

The two detected patterns in drivers' HR can be translated into a calm and stressful state for the drivers. When considering their interaction with the driving behaviors, we observe that, harsh brakes include a higher amount of abnormal patterns in HR in our dataset as compared to normal brakes (Fig. 4.18). This is then supported through a Kruskal-Wallis test (55) with a $statistic = 219.68$, and a $p - value < 0.00001$. Additionally, the free flow driving with close to zero acceleration has the higher amount of normal HR, when compared to accelerating behavior. This is confirmed through a Kruskal-Wallis test (55) with a $statistic = 71.79$, and a $p - value < 0.00001$.

With respect to drivers' gaze patterns, we observe that free flow driving has the lowest amount of high GTE values when compared to the accelerating behavior ($statistic = 363.390$, and a $p - value < 0.00001$). This may point to the lowest cognitive load in free flow driving with close to zero acceleration as compared to the accelerating behavior with positive high acceleration (Fig. 4.19). Additionally, we observe that similar to case study I, normal braking behavior has a higher GTE fractions as compared to the harsh braking behavior ($statistic = 158.969$, and a $p - value < 0.0001$).

Lastly, we aggregate the driving patterns with significantly higher than zero acceleration (both positive and negative) into one group titled as aggressive driving style. Additionally, we consider the free flow driving pattern as a conservative driving style where accelerations are close to zero. We then compare the fraction of abnormal HR and high GTE in each of

Table 4.13: Detailed statistics of HR and gaze for each recognized driver state pattern. Note that the values are normalized

Driver State Data	Statistical Index	State Pattern 0 (Normal HR/ Low Gaze Entropy)	State Pattern 1 (Abnormal HR/ High Gaze Entropy)
HR	Mean	76.78	89.15
	Standard Deviation	9.26	12.02
	25th percentile	69.45	79.76
	Median	76.32	88.92
	75th percentile	84	98
Gaze	Mean	5.17	5.28
	Standard Deviation	0.35	0.28
	25th percentile	5.01	5.14
	Median	5.21	5.29
	75th percentile	5.39	5.44

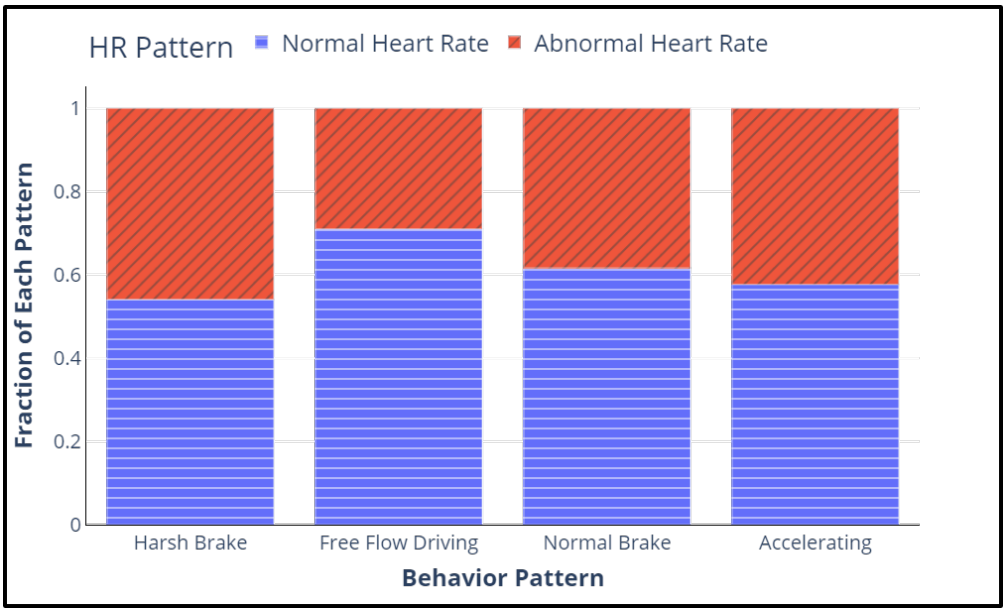


Figure 4.18: The fraction of each HR pattern within in each driving pattern from case study II. Note that free flow driving has the highest fraction of normal HR as compared to other driving behaviors, implying a calmer state for drivers.

Figure 4.19: The fraction of each gaze pattern within each driving pattern from case study II. Note that free flow driving has the highest fraction of low gaze entropy depicting lower work load.

these two categories of driving styles. Fig. 4.20 shows the fraction of driving segments with abnormal HR as well as high GTE together with the driving styles. As shown on the graph, almost all of the drivers are less likely to have high HR as well as high GTE in conservative monotone driving as compared to aggressive high acceleration driving. We also observe that the difference between fraction of High HR and GTE between aggressive and conservative driving varies across participants leading to different slopes for the lines connecting them. Additionally, among our participants, we observe that two of them has a higher fraction of abnormal HR and GTE in conservative driving (shown with red and blue color on Fig. 4.20). The implications of these differences will be discussed later.

4.2.5 Discussion

The fast-paced improvement in the development and testing of AVs has been one of the important advancements of the automobile industry over the recent years. To enhance the human-AV collaboration, realistic human behavior models are needed where a driver's comfort and behaviors can be modeled with a multidimensional approach. This is because human behavior is dynamic and can be affected by different internal dynamics (e.g., emotion and cognition), as well as external situational factors (e.g., traffic density or roadway conditions). The proposed framework in this paper couples both internal and external factors, namely context, to understand the pattern in driver's state in each driving behavior through unsu-

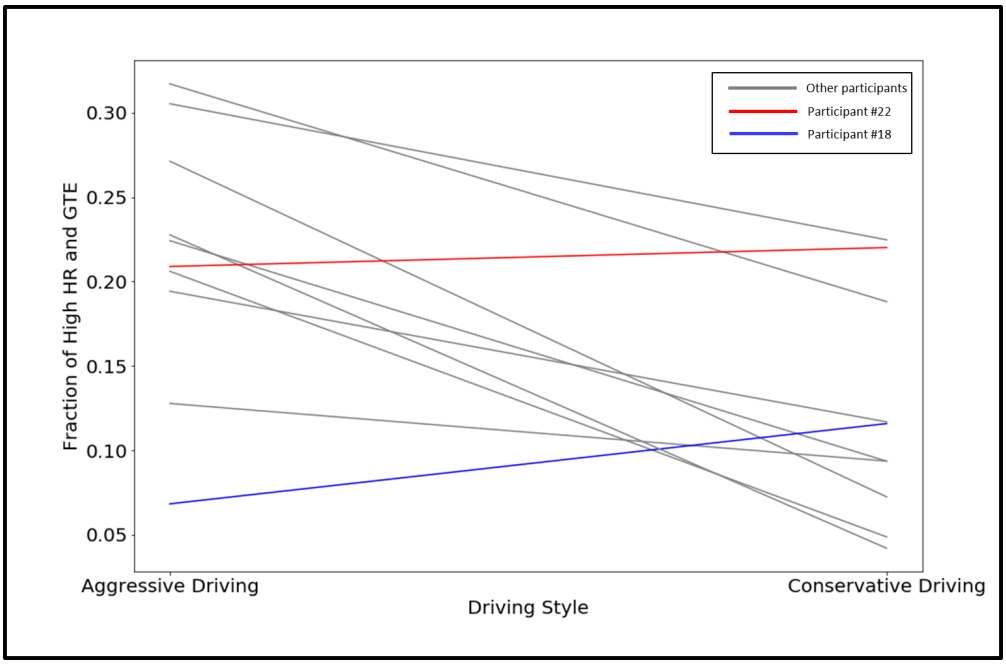


Figure 4.20: The fraction of each driver states patterns of high GTE and abnormal HR in each of conservative and aggressive driving styles. A conservative driving style tend to keep participants HR at a calmer state while using less cognitive resources. Two of the participants do not follow the general trend shown with red and blue.

pervised modeling techniques. While supervised learning has been an impactful approach over the past decade of driving research, the massive amount of data collected through past NDS points towards more efficient, fast-paced methods that can understand behaviors with no or minimum manual annotation. While in the presented case study I, we have used very high-resolution data from one participant for method illustration, case study II confirms that low-resolution data when coupled with unsupervised modeling techniques, can help detect certain states and behaviors across participants. As current vehicles are equipped with multiple sensors such as GPS, our case study II suggests that coupling a conventional wearable device with sensors that are already in vehicles can provide significant insight into the driver's state in each driving behavior.

Driver's state patterns indicated the level to which a driver might be affected by different driving behaviors. More specifically, in our first case study, using high precision data, we observed that the driver had a higher fraction of normal HR in free-flow driving on the highway, implying a calmer state for the driver. This has then been confirmed in our case study II where we analyzed the data from 12 participants. Free flow driving not only had the lowest fraction of high gaze entropy, implying a lower workload but also was accompanied by the least fraction of abnormal HR among our participants in case study II. These findings are in line with previous studies indicating less reported subjective emotional states (26), as well as calmer states (99) in highway driving. Additionally, our results suggest that even the same action can have different effects on the driver when performed with different styles. For instance, we observe that harsh brakes are more likely to be accompanied by abnormal HR values implying higher stress levels as compared to normal brakes.

Our method can be leveraged to understand the effects of different driving styles on drivers' states. Multiple studies have pointed out the importance of driving style selection for drivers' trust in AVs (28). For instance, a recent study points out that a "defensive" driving style as compared to an aggressive driving style is perceived to be more trustworthy (28). In case study II, our results suggest that a more conservative driving style that is accompanied by close to zero acceleration values is less probable to be accompanied by lower workload and stress levels. Although our sample size might not generalize to the larger pool of drivers, it shows the importance of choosing driving styles according to the drivers' psychophysiological states given specific road environments (e.g., highway versus city driving). For instance, our results show that free flow driving is a better fit for the drivers of our study in keeping their HR at a normal pattern. Choosing highways over crowded city streets can help driving with the free flow speed, which then helps to keep drivers calmer with a lower workload. These differences in stress levels and workload have strong implications for humanizing automated services, including both the AV and routing services. Taking our

results into account, different driving behaviors can be chosen based on how to they might affect a driver based on the unsupervised modeling of their historical driving data within the safety boundary of each behavior. Additionally, services such as route selection can take these factors into consideration for providing a human-centered service that considers how a driver might feel through driving on each route based on potential driving behaviors that can occur on the route (e.g., amount of highway driving, curve driving, etc.).

When comparing conservative versus aggressive driving style, we observe different levels to which participants are affected by each style, which might be indicative of individual differences when being in the two different driving styles. Additionally, among our participants, we have seen two participants that had higher abnormal HR and GTE during conservative driving style. This was in contrast with all the other 10 participants. After careful examination of the videos and interviewing the participants, we realized that one of the two drivers uses loud music in highways which might be indicative of the high HR values in conservative highway driving scenarios (shown with red on Fig. 4.20), while, the second driver often uses cruise control in highway driving (shown with blue). Although more data is required to address the effect of loud music in naturalistic driving, part of our future work will be focused on using the data provided in HARMONY (103), to assess the effect of music in naturalistic scenarios. This preliminary finding paves the way for understanding the naturalistic effect of music on driving behaviors and driver's states.

Additionally, while we did not have access to the cruise control usage data by the vehicle, we hypothesize that the abnormal HR and GTE state for the participant shown with blue on Fig. 4.20 could be due to interaction with the semi-automated cruise control system. Current cruise control systems are generally used to keep a constant speed during highway free flow driving. This might imply that although automated driving assisting systems (ADAS) can help the drivers for delivering a safe driving task, they might be the cause of more stress and might result in changes in drivers' state that can only be detected through using psychophysiological metrics. Changes in drivers' behaviors and states, while engaged in ADAS systems were mentioned previously in other naturalistic driving study researches ((27; 50)). It is possible that when the cruise control system was active, which is often on highways and at higher speeds, the driver had more stress as they had to supervise the automation. Alternatively it could also be the case that participant engaged in other activities or started mind-wandering, which increased the fraction of high HR and GTE. A very recent study focusing on semi-automated vehicles drivers is also pointing towards more eyes-off-the-road during using automated systems (68). More research with taking psychophysiology into account is required to understand the real effect of shared-autonomy on the driver. This also has strong implications for human sensing modules inside vehicles

that rely on computer vision. Such modules often only consider the driver's facial expressions and gaze patterns while leaving behind the driver's physiological responses such as HR, skin temperature, etc. While computer vision has shown to be a very strong method for detecting a driver's state, coupling it with physiological metrics can provide a deeper understanding of the driver's state.

Although we have not focused specifically on the sequence of behaviors, using Tables 4.6 and 4.11 we observe important insights regarding the transition probabilities across different behaviors. This helps with analyzing driving styles in a deeper fashion. From Table 4.6, we first observe that a harsh braking behavior is often followed by a normal braking behavior, while a normal braking behavior is more likely to continue in normal braking. This might imply that while performing normal braking, the driver is more confident regarding the context such as the surrounding road environment, traffic density, or other vehicles, thus the chances of switching to a harsh brake is lower than other two behavior. We also observe that the braking behavior that succeeds free flow driving is divided between the two types of braking behaviors. This might imply that depending on the context driver chooses between two different behavioral patterns. More information from the context can provide insight into the reasoning behind the choice between the two. Information such as the presence of other vehicles, the speed of other vehicles, and their distance can be predictive of such choice. This will be analyzed in our future work. Additionally, from Table 4.11 we observe that when being in accelerating driving behavior, the probability of switching to a hard brake is the highest among other behaviors. This indicates that aggressive, less monotone driving through accelerating behavior is more likely to result in a harsh braking/deceleration behavior, which can be inferred as continuing the less monotone aggressive driving. On the other hand, being in a free flow driving behavior is much less likely to switch to a harsh brake ($p = 0.02$). Such results could indicate the persistence of both aggressive or conservative behavior throughout the time, which can be used for driving style prediction. One part of the future work will be focused on this area to understand and predict the sequence of behaviors in different contexts.

Lastly, the findings of our approach to driver state modeling when applied in real-time, has the potential to guide an autonomous vehicle to better predict human behavior and take actions that are fit to driver's physiology. Note that a vehicle does not need to know the meaning of each unsupervised pattern. For our study, we translated driving patterns into driving behaviors, which helps with illustrating each pattern and justifying the method. However, a vehicle only requires to know the characteristics of each planned driving pattern, compare it with the detected patterns in the historical data, and the probability of abnormal HR within different planned behaviors. This can then help the vehicle choose among different

actions. For instance, when choosing between a harsh and normal brake, the vehicle can estimate the acceleration in each action, find the most similar driving pattern to each action based on acceleration characteristics and model fit, and choose the one that is more suitable to drivers' HR and gaze patterns. Additionally, using our approach in driver state modeling, we can better predict future driving decisions by building behavioral models of past driver states and incorporating theoretical findings in psychology such as memory based judgment and decision making (87; 88).

4.2.6 Limitations

Our work has a number of limitations, which will be addressed in the future work of this research. We will increase the number of participants as well as the duration of collected data to understand individual differences in how various driving styles may affect each driver. This can help us cluster different drivers based on how they are affected by the driving behaviors. We will perform deeper feature extraction on both driver's HR as well as gaze variation to retrieve a deeper understanding of the driver's state in each driving pattern. A deeper feature extraction together with a higher amount of data might reveal differences in drivers' states within different behavior in varying roadway designs. Our current dataset is focused on highways. Part of the future work should be focused on performing the same on other urban environments such as city streets and comparing the results with highway driving.

We will perform the same analysis with a higher number of sensors and modalities in both the human sensing and vehicle sensing part of our study. This can include modalities such as driver's other bio-signals (e.g., skin temperature) and modalities retrieved from the computer vision techniques (e.g., distance to other vehicles). The added modalities may reveal extended details regarding human psychophysiology within different driving pattern. These information can enhance our knowledge in understanding other important factors associated to changes in drivers' states with at a deeper level.

It should be noted that we only focused on using change point detection for segmenting applied on the kinematic driving data. It is indeed the case that other contextual information should be coupled with vehicle kinematic to provide a more fine-grained segmentation. For instance, computer vision can help with detecting intersections, lead vehicles, and other road objects which can be helpful in defining segments. Information such as time headway, distance to the lead vehicle and other behavioral metrics can help with both segmentation and pattern inference to become more fine-grained.

We will focus deeper on the sequence of driving behaviors to be able to predict behaviors using time series prediction methods. Currently, the sequence presented in this paper may suffer from road objects rather than drivers' choice in certain situations. For instance, when moving from free flow driving to curved driving, it is indeed the case that the curve on

the road dictates the driving behavior. In our future work, we will expand on fine grained behaviors within each of the curve driving and free flow driving to understand different levels of behaviors that are not dictated by the environment. Nevertheless, it might be the case that some behavioral sequences are drivers' choice which are related to the driving context. Future work should expand this section to a greater detail.

Lastly, we will consider other unsupervised methods that can help understand behaviors in a more fine-grained manner, such as the application of deep temporal clustering methods. It is important to note that most of the behavioral patterns in our study are generalized in which future work should couple a higher number of modalities together with a bigger dataset to find fine grained patterns. Such patterns can then be used for immediate path planning of future autonomous vehicles.

4.2.7 Conclusion

In this paper, we propose a method to understand driver's reactions to different driving maneuvers in real-world driving scenarios by using unsupervised methods. Using our approach, different driving behaviors can be ranked based on how they affect a user's well-being, thus producing actions that are more aligned with human's state, such as stress levels and cognitive loads. Our method uses a combination of Bayesian Change Point detection together with GMM-LDA methods applied on both driving and human sensing data to mine such associations between driver's state and behavior.

4.2.8 Appendix

4.2.8.1 LDA Details

LDA is based on the following generative assumptions (15; 58):

$$N \sim Poisson(\xi) \tag{4.15}$$

$$\theta_m \sim Dirichlet(\alpha) \tag{4.16}$$

$$\varphi_k \sim Dirichlet(\beta) \tag{4.17}$$

which θ_m is the topic distribution of document m in M and φ_k is the distribution of each word w in topic k in K. Note that α and β are the prior distributions of θ and φ respectively. In this way, the topic of the nth word in document m is $Z_{m,n}$, which has a Multinomial distribution on θ_m. Additionally, the nth word in document m has a Multinomial distribution over $\varphi_{z_{m,n}}$.

4.2.8.2 GMM Details

Formally, a GMM model is a weighted summation of M different components as:

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g(x|\mu_i, \Sigma_i) \tag{4.18}$$

In the equation above, x is the multidimensional vector and w_i is the identified different Gaussian distributions that sum up to 1, and are described as:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right) \tag{4.19}$$

in which μ_i is the mean vector and Σ_i is the covariance matrix. The parameters of different Gaussian distributions can be retrieved through the Expectation Maximization algorithm which is described in detail in (82).

Bibliography

[1] Abdic, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B. and Schuller, B. (2016). Driver frustration detection from audio and video in the wild, *Springer*, Vol. 9904, Springer, p. 237.

[2] Agamennoni, G., Ward, J. R., Worrall, S. and Nebot, E. M. (2014). Bayesian model-based sequence segmentation for inferring primitives in driving-behavioral data, *17th International Conference on Information Fusion (FUSION)*, IEEE, pp. 1–8.

[3] Agamennoni, G., Worrall, S., Ward, J. R. and Neboty, E. M. (2014). Automated extraction of driver behaviour primitives using bayesian agglomerative sequence segmentation, *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1449–1455.

[4] Alharbi, R., Stump, T., Vafaie, N., Pfammatter, A., Spring, B. and Alshurafa, N. (2018). I can't be myself: effects of wearable cameras on the capture of authentic behavior in the wild, *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* **2**(3): 1–40.

[5] Alrefaie, M. T., Summerskill, S. and Jackon, T. W. (2019). In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles, *Accident Analysis & Prevention* **131**: 180–190.

[6] Alsaid, A., Lee, J. D. and Price, M. (2020). Moving into the loop: An investigation of drivers' steering behavior in highly automated vehicles, *Human factors* **62**(4): 671–683.

[7] Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L. and Amirat, Y. (2015). Physical human activity recognition using wearable sensors, *Sensors* **15**(12): 31314–31338.

[8] Baee, S., Pakdamanian, E., Kim, I., Feng, L., Ordonez, V. and Barnes, L. (2019). Medirl: Predicting the visual attention of drivers via maximum entropy deep inverse reinforcement learning, *arXiv preprint arXiv:1912.07773* .

[9] Baltrusaitis, T., Robinson, P. and Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild, *Proceedings of the IEEE international conference on computer vision workshops*, pp. 354–361.

[10] Baltrusaitis, T., Zadeh, A., Lim, Y. C. and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, pp. 59–66.

[11] Bando, T., Takenaka, K., Nagasaka, S. and Taniguchi, T. (2013). Automatic drive annotation via multimodal latent topic model, *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 2744–2749.

[12] Barry, D. and Hartigan, J. A. (1993). A bayesian analysis for change point problems, *Journal of the American Statistical Association* **88**(421): 309–319.

[13] Bender, A., Agamennoni, G., Ward, J. R., Worrall, S. and Nebot, E. M. (2015). An unsupervised approach for inferring driver behavior from naturalistic driving data, *IEEE transactions on intelligent transportation systems* **16**(6): 3325–3336.

[14] Bi, C., Huang, J., Xing, G., Jiang, L., Liu, X. and Chen, M. (2019). Safewatch: a wearable hand motion tracking system for improving driving safety, *ACM Transactions on Cyber-Physical Systems* **4**(1): 1–21.

[15] Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation, *the Journal of machine Learning research* **3**: 993–1022.

[16] Boukhechba, M. (n.d.). Swear - apps on google play.
URL: *https://play.google.com/store/apps/details?id=uva.swear*

[17] Boukhechba, M. and Barnes, L. E. (2020). Swear: Sensing using wearables. generalized human crowdsensing on smartwatches, *2019 IEEE 11th International Conference on Applied Human Factors and Ergonomics. IEEE.*

[18] Calder, W. J. and Shuman, B. (2017). Extensive wildfires, climate change, and an abrupt state change in subalpine ribbon forests, colorado, *Ecology* **98**(10): 2585–2600.

[19] Carsten, O., Kircher, K. and Jamson, S. (2013). Vehicle-based studies of driving in the real world: The hard truth?, *Accident Analysis & Prevention* **58**: 162–174.

[20] Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique, *Journal of artificial intelligence research* **16**: 321–357.

[21] Cheng, T., Migliaccio, G. C., Teizer, J. and Gatti, U. C. (2013). Data fusion of real-time location sensing and physiological status monitoring for ergonomics analysis of construction workers, *Journal of Computing in Civil engineering* **27**(3): 320–335.

[22] Choi, M., Koo, G., Seo, M. and Kim, S. W. (2017). Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness, *IEEE Transactions on Instrumentation and Measurement* **67**(3): 634–645.

[23] Christ, M., Braun, N., Neuffer, J. and Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package), *Neurocomputing* **307**: 72–77.

[24] Comma.ai (n.d.). Open pilot.
URL: *https://comma.ai/*

[25] Dhahir, B. and Hassan, Y. (2019). Modeling speed and comfort threshold on horizontal curves of rural two-lane highways using naturalistic driving data, *Journal of Transportation Engineering, Part A: Systems* **145**(6): 04019025.

[26] Dittrich, M. (2021). Why drivers feel the way they do: An on-the-road study using self-reports and geo-tagging, *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 116–125.

[27] Dunn, N., Dingus, T. and Soccolich, S. (2019). Understanding the impact of technology: Do advanced driver assistance and semi-automated vehicle systems lead to improper driving behavior?

[28] Ekman, F., Johansson, M., Bligård, L.-O., Karlsson, M. and Strömberg, H. (2019). Exploring automated vehicle driving styles as a source of trust information, *Transportation research part F: traffic psychology and behaviour* **65**: 268–279.

[29] Engström, J., Markkula, G., Victor, T. and Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis, *Human factors* **59**(5): 734–764.

[30] Erdman, C. and Emerson, J. W. (2007). bcp: an r package for performing a bayesian analysis of change point problems, *Journal of Statistical Software* **23**(1): 1–13.

[31] Ergan, S., Radwan, A., Zou, Z., Tseng, H.-a. and Han, X. (2019). Quantifying human experience in architectural spaces with integrated virtual reality and body sensor networks, *Journal of Computing in Civil Engineering* **33**(2): 04018062.

[32] Fabio, R. A., Incorpora, C., Errante, A., Mohammadhasni, N., Caprì, T., Carrozza, C., De Santis, S. and Falzone, A. (2015). The influence of cognitive load and amount of stimuli on entropy through eye tracking measures., *EAPCogSci*.

[33] Farah, H., Koutsopoulos, H. N., Saifuzzaman, M., Kölbl, R., Fuchs, S. and Bankosegger, D. (2012). Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior, *Transportation research part C: emerging technologies* **21**(1): 42–56.

[34] Francis, A. L. (2018). The embodied theory of stress: A constructionist perspective on the experience of stress, *Review of General Psychology* **22**(4): 398–405.

[35] Fridman, L. (2018). Human-centered autonomous vehicle systems: Principles of effective shared autonomy, *arXiv preprint arXiv:1810.01835* .

[36] Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Patsekin, A., Kindelsberger, J., Ding, L. et al. (2019). Mit advanced vehicle technology study: Large-scale naturalistic driving study of driver behavior and interaction with automation, *IEEE Access* **7**: 102021–102038.

[37] Fridman, L., Reimer, B., Mehler, B. and Freeman, W. T. (2018). Cognitive load estimation in the wild, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, p. 652.

[38] Gilks, W. R. (2005). Markov chain monte carlo, *Encyclopedia of biostatistics* **4**.

[39] Goel, B., Dey, A. K., Bharti, P., Ahmed, K. B. and Chellappan, S. (2018). Detecting distracted driving using a wrist-worn wearable, *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, pp. 233–238.

[40] Hancock, P. A., Nourbakhsh, I. and Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles, *Proceedings of the National Academy of Sciences* **116**(16): 7684–7691.

[41] Hernandez, N., Lundström, J., Favela, J., McChesney, I. and Arnrich, B. (2020). Literature review on transfer learning for human activity recognition using mobile and wearable devices with environmental technology, *SN Computer Science* **1**(2): 66.

[42] Huang, H., Chen, H. and Lin, S. (2019). Magtrack: Enabling safe driving monitoring with wearable magnetics, *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 326–339.

[43] Hussain, Z., Sheng, Q. Z. and Zhang, W. E. (2020). A review and categorization of techniques on device-free human activity recognition, *Journal of Network and Computer Applications* p. 102738.

[44] Jebelli, H., Hwang, S. and Lee, S. (2018). Eeg signal-processing framework to obtain high-quality brain waves from an off-the-shelf wearable eeg device, *Journal of Computing in Civil Engineering* **32**(1): 04017070.

[45] Jegham, I., Khalifa, A. B., Alouani, I. and Mahjoub, M. A. (2020). Vision-based human action recognition: An overview and real world challenges, *Forensic Science International: Digital Investigation* **32**: 200901.

[46] Kass-Hout, T. A., Xu, Z., McMurray, P., Park, S., Buckeridge, D. L., Brownstein, J. S., Finelli, L. and Groseclose, S. L. (2012). Application of change point analysis to daily influenza-like illness emergency department visits, *Journal of the American Medical Informatics Association* **19**(6): 1075–1081.

[47] Katsis, C., Goletsis, Y., Rigas, G. and Fotiadis, D. (2011). A wearable system for the affective monitoring of car racing drivers during simulated conditions, *Transportation research part C: emerging technologies* **19**(3): 541–551.

[48] Khurana, R. and Goel, M. (2020). Eyes on the road: Detecting phone usage by drivers using on-device cameras, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–11.

[49] Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. and Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature, *Psychiatry investigation* **15**(3): 235.

[50] Kim, H., Song, M. and Doerzaph, Z. (2021). Is driving automation used as intended? real-world use of partially automated driving systems and their safety consequences, *Transportation Research Record* p. 03611981211027150.

[51] Kim, J., Yadav, M., Chaspari, T. and Ahn, C. R. (2020). Environmental distress and physiological signals: Examination of the saliency detection method, *Journal of Computing in Civil Engineering* **34**(6): 04020046.

[52] Krasnov, F. and Sen, A. (2019). The number of topics optimization: Clustering approach, *Machine Learning and Knowledge Extraction* **1**(1): 416–426.

[53] Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A. and Villalobos, N. (2015). Gaze transition entropy, *ACM Transactions on Applied Perception (TAP)* **13**(1): 1–20.

[54] Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C. and Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze, *PloS one* **13**(9): e0203629.

[55] Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis, *Journal of the American statistical Association* **47**(260): 583–621.

[56] Kundinger, T., Yalavarthi, P. K., Riener, A., Wintersberger, P. and Schartmüller, C. (2020). Feasibility of smart wearables for driver drowsiness detection and its potential among different age groups, *International Journal of Pervasive Computing and Communications* .

[57] Lee, G., Choi, B., Jebelli, H., Ahn, C. R. and Lee, S. (2020). Wearable biosensor and collective sensing–based approach for detecting older adults' environmental barriers, *Journal of Computing in Civil Engineering* **34**(2): 04020002.

[58] Li, G., Chen, Y., Cao, D., Qu, X., Cheng, B. and Li, K. (2021). Extraction of descriptive driving patterns from driving data using unsupervised algorithms, *Mechanical Systems and Signal Processing* **156**: 107589.

[59] Li, L., Xie, Z., Xu, X., Liang, Y. and Horrey, W. (2019). Recognition of manual driving distraction through deep-learning and wearable sensing.

[60] Li, L., Zhong, B., Hutmacher Jr, C., Liang, Y., Horrey, W. J. and Xu, X. (2020). Detection of driver manual distraction via image-based hand and ear recognition, *Accident Analysis & Prevention* **137**: 105432.

[61] Li, N., Misu, T. and Miranda, A. (2016). Driver behavior event detection for manual annotation by clustering of the driver physiological signals, *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2583–2588.

[62] Liu, L., Karatas, C., Li, H., Tan, S., Gruteser, M., Yang, J., Chen, Y. and Martin, R. P. (2015). Toward detection of unsafe driving with wearables, *Proceedings of the 2015 workshop on Wearable Systems and Applications*, pp. 27–32.

[63] Lohani, M., Payne, B. R. and Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving, *Frontiers in human neuroscience* **13**: 57.

[64] Malladi, R., Kalamangalam, G. P. and Aazhang, B. (2013). Online bayesian change point detection algorithms for segmentation of epileptic activity, *2013 Asilomar Conference on Signals, Systems and Computers*, IEEE, pp. 1833–1837.

[65] Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit, *Journal of the American statistical Association* **46**(253): 68–78.

[66] Maye, J., Triebel, R., Spinello, L. and Siegwart, R. (2011). Bayesian on-line learning of driving behaviors, *2011 IEEE International Conference on Robotics and Automation*, IEEE, pp. 4341–4346.

[67] McHugh, M. L. (2013). The chi-square test of independence, *Biochemia medica* **23**(2): 143–149.

[68] Morando, A., Gershon, P., Mehler, B. and Reimer, B. (2021). A model for naturalistic glance behavior around tesla autopilot disengagements, *Accident Analysis & Prevention* **161**: 106348.

[69] Motamedi, S., Wang, P., Zhang, T. and Chan, C.-Y. (2020). Acceptance of full driving automation: personally owned and shared-use concepts, *Human factors* **62**(2): 288–309.

[70] Nacpil, E. J. C., Wang, Z. and Nakano, K. (2021). Application of physiological sensors for personalization in semi-autonomous driving: A review, *IEEE Sensors Journal* .

[71] Napoli, N., Adams, S., Harrivel, A. R., Stephens, C., Kennedy, K., Paliwal, M. and Scherer, W. (2020). Exploring cognitive states: Temporal methods for detecting and characterizing physiological fingerprints, *AIAA Scitech 2020 Forum*, p. 1193.

[72] Napoli, N. J., Demas, M. W., Mendu, S., Stephens, C. L., Kennedy, K. D., Harrivel, A. R., Bailey, R. E. and Barnes, L. E. (2018). Uncertainty in heart rate complexity metrics caused by r-peak perturbations, *Computers in biology and medicine* **103**: 198–207.

[73] Ohn-Bar, E., Martin, S., Tawari, A. and Trivedi, M. M. (2014). Head, eye, and hand patterns for driver activity recognition, *2014 22nd International Conference on Pattern Recognition*, IEEE, pp. 660–665.

[74] Pakdamanian, E., Namaky, N., Sheng, S., Kim, I., Coan, J. A. and Feng, L. (2020). Toward minimum startle after take-over request: A preliminary study of physiological data, *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 27–29.

[75] Pakdamanian, E., Sheng, S., Baee, S., Heo, S., Kraus, S. and Feng, L. (2020). Deeptake: Prediction of driver takeover behavior using multimodal data, *arXiv preprint arXiv:2012.15441* .

[76] Pan, C., Cao, H., Zhang, W., Song, X. and Li, M. (n.d.). Driver activity recognition using spatial-temporal graph convolutional lstm networks with attention mechanism, *IET Intelligent Transport Systems* .

[77] Park, S. Y., Moore, D. J. and Sirkin, D. (2020). What a driver wants: User preferences in semi-autonomous vehicle decision-making, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

[78] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* **12**: 2825–2830.

[79] Perepelkina, O., Artemyev, M., Churikova, M. and Grinenko, M. (2020). Hearttrack: Convolutional neural network for remote video-based heart rate monitoring, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 288–289.

[80] Prabhudesai, K. S., Mainsah, B. O., Collins, L. M. and Throckmorton, C. S. (2018). Augmented latent dirichlet allocation (lda) topic model with gaussian mixture topics, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 2451–2455.

[81] Qiu, Y., Misu, T. and Busso, C. (2019). Analysis of the relationship between physiological signals and vehicle maneuvers during a naturalistic driving study, *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, IEEE, pp. 3230–3235.

[82] Reynolds, D. A. (2009). Gaussian mixture models., *Encyclopedia of biometrics* **741**: 659–663.

[83] Roitberg, A., Ma, C., Haurilet, M. and Stiefelhagen, R. (2020). Open set driver activity recognition, *2020 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, pp. 1048–1053.

[84] Saeed, A., Trajanovski, S., van Keulen, M. and van Erp, J. (2017). Deep physiological arousal detection in a driving simulator using wearable sensors, *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp. 486–493.

[85] Sakib, M. N., Chaspari, T. and Behzadan, A. H. (2021). Physiological data models to understand the effectiveness of drone operation training in immersive virtual reality, *Journal of Computing in Civil Engineering* **35**(1): 04020053.

[86] Shannon, C. E. (1948). A mathematical theory of communication, *The Bell system technical journal* **27**(3): 379–423.

[87] Sharif, M. A. and Oppenheimer, D. M. (2016). The effect of relative encoding on memory-based judgments, *Psychological science* **27**(8): 1136–1145.

[88] Sharif, M. A. and Oppenheimer, D. M. (2021). The effect of categories on relative encoding biases in memory-based judgments, *Organizational Behavior and Human Decision Processes* **162**: 1–8.

[89] Shiferaw, B. A., Crewther, D. P. and Downey, L. A. (2019). Gaze entropy measures detect alcohol-induced driver impairment, *Drug and alcohol dependence* **204**: 107519.

[90] Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M., Berlowitz, D. J., Swann, P. and Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers, *Scientific reports* **8**(1): 1–10.

[91] Shiferaw, B., Downey, L. and Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency, *Neuroscience & Biobehavioral Reviews* **96**: 353–366.

[92] Sousa Lima, W., Souto, E., El-Khatib, K., Jalali, R. and Gama, J. (2019). Human activity recognition using inertial sensors in a smartphone: An overview, *Sensors* **19**(14): 3213.

[93] Stark, L., Stanhaus, A. and Anthony, D. L. (2020). "i don't want someone to watch me while i'm working": Gendered views of facial recognition technology in workplace surveillance, *Journal of the Association for Information Science and Technology* **71**(9): 1074–1088.

[94] Sugie, R., Arakawa, T. and Kozuka, K. (2016). Detection of fatigue in long-distance driving by heart rate variability, *ICIC Express Letters* **10**(7): 1553–1559.

[95] Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., Li, H. N. and Wang, Q. (2020). Exploring personalised autonomous vehicles to influence user trust, *Cognitive Computation* **12**(6): 1170–1186.

[96] Taniguchi, T., Nagasaka, S., Hitomi, K., Takenaka, K. and Bando, T. (2014). Unsupervised hierarchical modeling of driving behavior and prediction of contextual changing points, *IEEE Transactions on Intelligent Transportation Systems* **16**(4): 1746–1760.

[97] Tavakoli, A. (2020). Code for multimodal driver state modeling through unsupervised learning.
URL: *https://github.com/arashtavakoli/Multimodal-Driver-State-Modeling-Through-Unsupervised-Learning*

[98] Tavakoli, A., Balali, V. and Heydarian, A. (2019). A multimodal approach for monitoring driving behavior and emotions, *Technical report*, Transportation Research Board, Washington DC, United States.

[99] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2020). Personalized driver state profiles: A naturalistic data-driven study,, *11th International Conference on Applied Human Factors and Ergonomics*.

[100] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2021). Leveraging ubiquitous computing for empathetic routing: A naturalistic data-driven approach, *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[101] Tavakoli, A., Kumar, S., Boukhechba, M. and Heydarian, A. (2021). Driver state and behavior detection through smart wearables, *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE.
URL: *https://doi.org/10.1109/iv48863.2021.9575431*

[102] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021a). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* pp. 1–1.

[103] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021b). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* **9**: 23956–23978.

[104] Tran, M., Sen, T., Haut, K., Ali, M. R. and Hoque, M. E. (2020). Are you really looking at me? a feature-extraction framework for estimating interpersonal eye gaze from conventional video, *IEEE Transactions on Affective Computing* .

[105] Twomey, N., Diethe, T., Fafoutis, X., Elsts, A., McConville, R., Flach, P. and Craddock, I. (2018). A comprehensive study of activity recognition using accelerometers, *Informatics*, Vol. 5, Multidisciplinary Digital Publishing Institute, p. 27.

[106] UVABRAINLAB (2020). Harmony case study.
URL: *https://osf.io/zextd/*

[107] Wang, B., Gong, J., Zhang, R. and Chen, H. (2018). Learning to segment and represent motion primitives from driving data for motion planning applications, *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1408–1414.

[108] Xing, Y., Lv, C., Cao, D. and Hang, P. (2021). Toward human-vehicle collaboration: Review and perspectives on human-centered collaborative automated driving, *Transportation Research Part C: Emerging Technologies* **128**: 103199.

[109] Yamakoshi, T., Rolfe, P., Yamakoshi, Y. and Hirose, H. (2009). A novel physiological index for driver's activation state derived from simulated monotonous driving studies, *Transportation research part C: emerging technologies* **17**(1): 69–80.

[110] Zepf, S., Dittrich, M., Hernandez, J. and Schmitt, A. (2019). Towards empathetic car interfaces: Emotional triggers while driving, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[111] Zhang, T., Zeng, W., Zhang, Y., Tao, D., Li, G. and Qu, X. (2021). What drives people to use automated vehicles? a meta-analytic review, *Accident Analysis & Prevention* **159**: 106270.

[112] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A. (2015). Appearance-based gaze estimation in the wild, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4511–4520.

[113] Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y. and Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling, *BMC bioinformatics*, Vol. 16, Springer, pp. 1–10.

Chapter 5

How Does the Driving Context Affect Driver's State?

This chapter discusses methods in understanding changes in drivers' states when interacting with the complicated dynamic driving environment. In this regard, three projects are discussed where the first project focuses on momentarily changes in driver's HR as a proxy for stress level and its correlation with environmental changes. First, through a manual annotation scheme, we show that abrupt changes in HR are correlated with specific environmental attributes (e.g., presence of intersection). These environmental attributes were previously shown to be correlated with higher stress levels in literature. In the second project, we find the differences across environmental attributes as to how much they affect drivers' HR and how this varies across individuals. In this project, automatic environmental attribute detection will be performed through Computer Vision, and linear mixed effect models will be leveraged to find individual differences. In the third project latent variable state space approach is leveraged to detect drivers state variation (i.e., stress and workload) under the perturbations of traffic density and task demands with a structural approach.

5.1 Project 1: Driver's Heart Rate as a Feedback to the Environmental Events - A Case Study

5.1.1 Introduction

[1] One of the key elements of human-vehicle interaction is the prediction and analysis of drivers' states and behaviors, such as stress, anxiety, and negative emotions. Detecting and mitigating drivers' stress level and negative emotions on the road are of high importance for decreasing accident rates as well as providing a human-centered experience in driving (18; 6). Additionally, research shows that driver-state changes are associated with certain

[1] ©2021 IEEE. Reprinted, with permission, from (111)

environmental events. For instance, recent research suggests that driver stress level and anxiety are correlated with specific events, such as the presence of passengers, big vehicles, pedestrians, and intersections (120; 22; 15; 110).

Integrating the above information into the human-vehicle interaction development can help mitigate the adverse effects of driving on the users and provide a human-centeric driving experience. However, methods of data collection other than subjective self-reports are required to understand the real-time reaction of drivers to these environmental attributes (6). In this regard, studies have already shown the utility of drivers' psychophysiology in capturing drivers' reactions to environmental attributes.

Literature suggests that stress is the process in which the demand of a certain situation is perceived to be more than the available resources (33). The perceived demand can be defined based on the overall situation, including the previous experiences, internal body sensations, and the external stimuli (33). The experienced stress may be accompanied by changes in physiology, such as increases in HR. Thus, it is possible that abrupt changes in drivers' HR, are preceded by the presence of stress-inducing road objects and events (e.g., passing through an intersection or presence of a big vehicle) (22).

This project is mainly focused on understanding the interaction between environmental attributes and changes in drivers' heart rate (HR). The collected data in HARMONY is employed to show the utility of driver's HR as feedback to environmental events. This is first done through a visual inspection of videos that are accompanied by abrupt increases in drivers' HR. The abrupt increases are detected through Bayesian Change Point Detection (BCP). Fusing the information from the in-cabin and outside videos with the HR changepoint locations from 15 participants results in understanding the reasons behind abrupt increases in HR. After manual inspection, this chapter proposes using object detection techniques to detect the categories of stressors on the road in a larger dataset. Based on applying YOLOV5 object detection algorithm and by taking advantage of linear mixed effect models, this project will focus on understanding the individual differences across participants in their responses to environmental attributes. The preliminary results of this project are published in (111; 106). This chapter specifically answers the following research questions:

Research Question 1: Is a driver's heart rate as retrieved through conventional wearable devices indicative of external contextual changes?

Research Question 2: What are the differences across people in the change in their HR with respect to such environmental stressors?

5.1.2 Background

Human biomarkers were used in the literature for detecting stress levels (17). In this regard, human HR has received special attention as HR can be collected using conventional

wearable devices and have the potential to be used in the wild. Wearable devices often use photoplethysmogram (PPG) technology, which is based on using infrared to detect the blood volume pressure in the veins on the wrist (108). This is then used to estimate the HR and a set of heart rate variability (HRV) features. Using conventional wearable devices, studies showed that an increase in HR is correlated with an increase in the levels of certain states such as anxiety, stress, and anger (55; 111).

Within the driving research area, studies have provided significant insights into the causation behind drivers' stress and negative emotions. These studies were mostly conducted in an on-road controlled fashion or within a driving simulator framework. In this regard, (120) monitored 33 drivers during a 50 minutes naturalistic on-road controlled study through a combination of cameras, sensors, and self-reports through voice. Through analyzing 531 self-reports, the authors found that the four main types of emotion triggers include traffic & driving task, environment, HCI & navigation, and vehicle and equipment. Their study mentions different detailed categories such as traffic lights, road design, the behavior of others, weather, and building and sites as reasons within the aforementioned larger categories for negative emotions of drivers. Another study performed by (111) identified that different characteristics of the scene might be associated with abrupt increases in drivers' HR. In their study, authors found lead vehicles, intersections, being followed by another vehicle, and performing secondary tasks to be the most significant factors associated with abrupt changes in drivers' HR. Later a study by (15) attempted to classify drivers' self-reported stress levels using the objects in the scene. In their study, by applying computer vision methods such as convolutional neural networks on the outside videos as input, authors achieved an accuracy of 72 % in the prediction of drivers' self-reported stress in an on-road controlled study. Their study also points out that objects such as traffic signs, cars, pedestrians, big vehicles, and riders are associated with medium to high subjective stress levels. Another recent study performed by (22) tracked the emotions of 34 drivers through self-reports on an on-road controlled study and found out that intersections are the hotspots for emotional triggers. Additionally, they found out that highways are associated with the least stress level as compared to other urban environments. This study also points out that the behavior of other road users, traffic lights, and navigation have a higher fraction of negative emotions as compared to the category of entertainment, which had a more positive emotion.

Research has shown the association between changes in drivers' physiological signals, and environmental attributes. A recent study through a naturalistic driving framework found out that drivers' have a lower heart rate when driving in clear versus rainy, highway versus city, and with a passenger versus alone conditions (107). Another study found out that drivers' physiological responses are correlated with vehicle's kinematic, which is governed by

driver's behavior (74). In their study, authors conducted a naturalistic study spanning over 383 days and from 16 drivers. They concluded that drivers' HR decreases when the number of people in the car increases. Another study performed by (109) analyzed drivers' HR and gaze entropy within different driving patterns. By using unsupervised modeling techniques, the authors found out that drivers exhibit a higher fraction of the pattern of normal HR during highway free-flow driving with higher speed. Additionally, they found out that a more conservative driving style was accompanied by smaller fractions of abnormal HR and gaze entropy patterns as compared to a risky driving style. Lastly, (62) proposed a framework to detect drivers' emotions using a combination of vehicle telemetry as well as the outdoor visual scene. Based on the facial expressions retrieved from the Affectiva software applied on 675 hours of driving data, authors were able to detect the emotions in a user-dependent and user-independent fashion with 70 % and 60 % accuracy, respectively.

5.1.3 Hypothesis

The above literature taken together suggests certain objects on the road are correlated with higher subjective stress levels and negative emotions. Additionally, higher stress levels, anxiety, and negative emotions are also correlated with increases in HR relative to the resting condition. This allows us to hypothesize that:

- **Hypothesis:** Certain on-road objects (e.g., big vehicles) and infrastructures (e.g., intersections) might be correlated with abrupt increases in drivers' stress levels. In this regard, abrupt increases in drivers' HR, which is indicative of drivers' stress level, might be accompanied with the presence of these on-road stressors.

5.1.4 Methodology

In order to detect abrupt increases in HR, we propose using Bayesian Change Point (BCP) method. After finding the moments that have abrupt changes in HR, we can search in the driving videos at the location of change points in HR to find the specific environmental attribute that can be related to the increase in HR. 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.

5.1.4.1 Bayesian Change point Analysis

[2] 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 (9) Bayesian change point model for this analysis. This model assumes there is an unknown partition ρ of the data in the contiguous regime,

[2] ©2021 IEEE. Reprinted, with permission, from (111)

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 $\{X_1, ..., X_n\}$. We will use X_{ij} to refer to the observations between indices i and j. Let $\rho = (U_1, ..., U_n)$ 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

$$f(\rho|X) \propto f(X|\rho)f(\rho) \tag{5.1}$$

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

$$f(\rho|p) = p^{b-1}(1-p)^{n-b} \tag{5.2}$$

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

$$\begin{aligned} f(X_{ij}, \theta) &= \Pi f(X_k|\theta)f(\theta) \\ f(X_{ij}) &= \int \Pi f(X_k|\theta)f(\theta)d\theta \end{aligned} \tag{5.3}$$

The above integral can be simplified to the expression below

$$f(X_{ij}) = (\frac{1}{2\pi\sigma^2})^{(j-i)/2}(\frac{\sigma^2}{\sigma_0^2+\sigma^2})^{1/2}exp(V_{ij}) \tag{5.4}$$

Where

$$V_{ij} = -\frac{\sum_{l=i+1}^{l=j}(X_l - \hat{X}_{ij})^2}{2\sigma^2} - \frac{(j-i)(\hat{X}_{ij}-\mu_0)^2}{2(\sigma^2+\sigma_0^2)} \tag{5.5}$$

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:

$$\begin{aligned} f(\mu_0) = 1, -\infty \leq \mu_0 \leq \infty \\ f(p) = 1/p_0, 0 \leq p \leq p_0 \\ f(\sigma^2) = 1/\sigma^2, 0 \leq \sigma^2 \leq \infty \\ f(w) = 1/w_0, 0 \leq w \leq w_0 \end{aligned} \tag{5.6}$$

$$f(X|\rho, \mu_0, w) = \int_0^\infty 1/\sigma^2 \prod_{ij \epsilon P} f(X_{ij}) d\sigma^2 \tag{5.7}$$

After integrating out μ_0 and w, This can be simplified to the indefinite integral below. We refer the readers to (9) for the full derivation.

$$f(X|\rho) \propto \int_0^{w_0} \frac{w^{(b-1)/2}}{(W + Bw)^{(n-1)/2}} dw, \tag{5.8}$$

where

$$\begin{aligned} \hat{X} = \sum_{i=1}^{n} X_i/n, \; B = \sum_{ij \epsilon P} (j-i)(\hat{X}_{ij} - \hat{X})^2, \\ W = \sum_{ij \epsilon P} \sum_{l=i+1}^{l=j} (X_l - \hat{X}_{ij})^2 \end{aligned} \tag{5.9}$$

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

$$f(\rho) \propto \int_0^{p_0} p^{b-1}(1-p)^{n-b} dp \tag{5.10}$$

To calculate the posterior distribution over partitions, we use Markov Chain Monte Carlo (MCMC) (37). 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.

$$\begin{aligned}\frac{p_i}{1-p_i} &= \frac{p(U_i=1|X,U_j,j\neq i)}{p(U_i=0|X,U_j,j\neq i)} \\ &= \frac{\int_0^{p_0} p^b(1-p)^{n-b-1}dp}{\int_0^{p_0} p^{b-1}(1-p)^{n-b}dp} x \frac{\int_0^{w_0} \frac{w^{b/2}}{(W_1+B_1w)^{(n-1)/2}}dw}{\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0+B_0w)^{(n-1)/2}}dw}\end{aligned} \tag{5.11}$$

This ultimately leaves us with a probabilistic model with the following two parameters p_0 and w_0. We use the package *bcp* (27) in R programming language to implement our change point analysis.

5.1.5 Case Study

In order to contextualize the change points in HR and variation in driving environment, we first use a case study from the HARMONY dataset. In this regard, we use approximately 2 hours of driving data randomly drawn from 10 participants that have been collected through the HARMONY dataset. The BCP is applied to the participants HR, and videos were used manually as reasoning. After finding the location of change points in drivers' HR, we use manual annotations from the videos to find out possible reasons for the abrupt increase in HR. The annotations are based on the annotation scheme detailed out in section 3.4.7.

5.1.5.1 Preliminary Results

[3] We first focus on one participant to illustrate our results. Fig. 5.1 depicts the participant #9 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. 5.1 - 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. 5.1. 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 then performed the same analysis on the 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

[3] ©2021 IEEE. Reprinted, with permission, from (111)

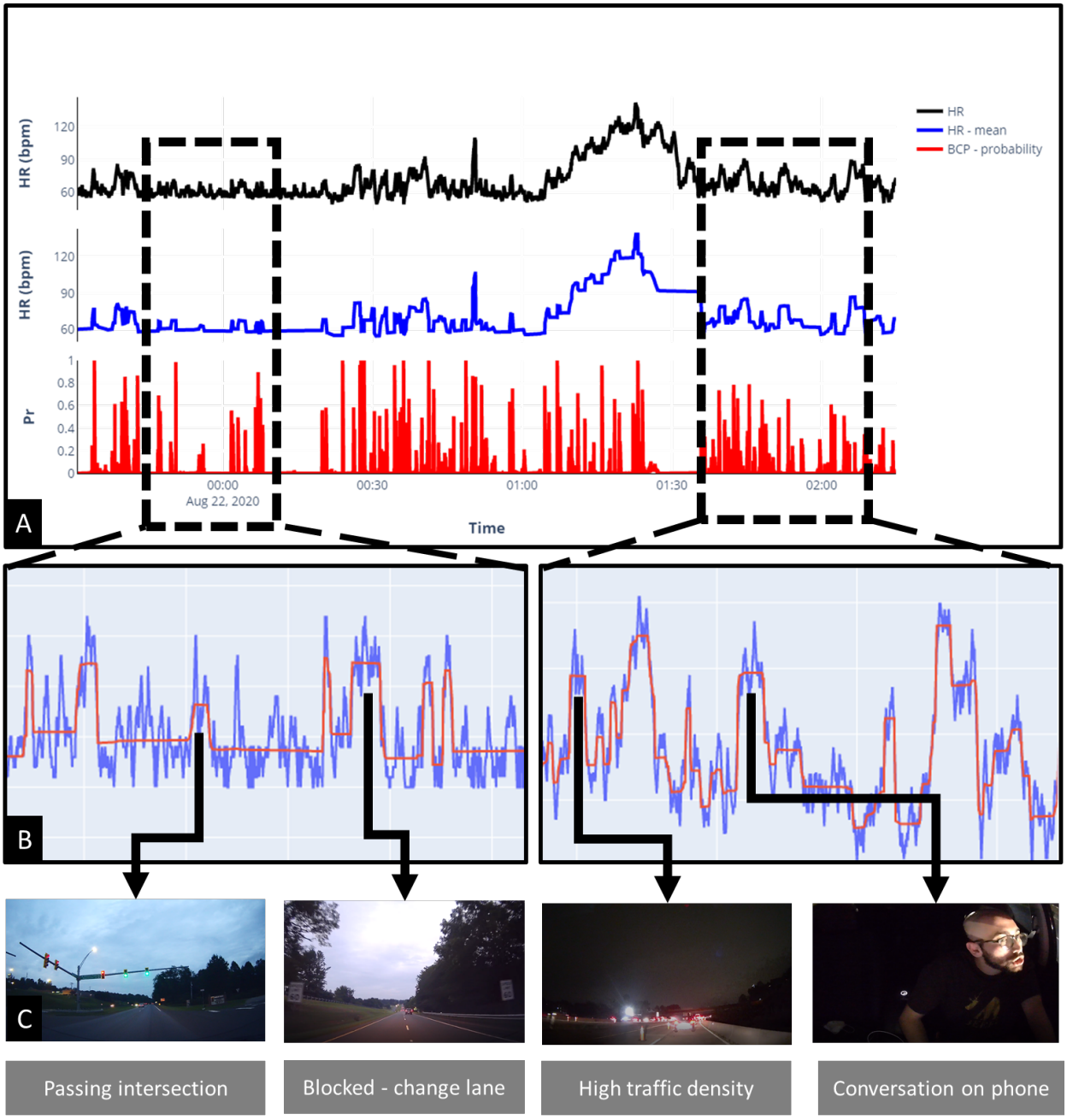


Figure 5.1: Participant's HR values together with the overlay of mean values calculated by BCP (A) the time series of participant's HR (black) together with average HR value between each two consecutive change points (blue) and the probability of detecting a change points (red) (B) the visualization 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. ©2021 IEEE

Table 5.1: Different driving events detected by the HR change point detection. ©2021 IEEE

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

simultaneously with certain categories of in-cabin or outside events. Specifically, the following categories of events are identified that 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

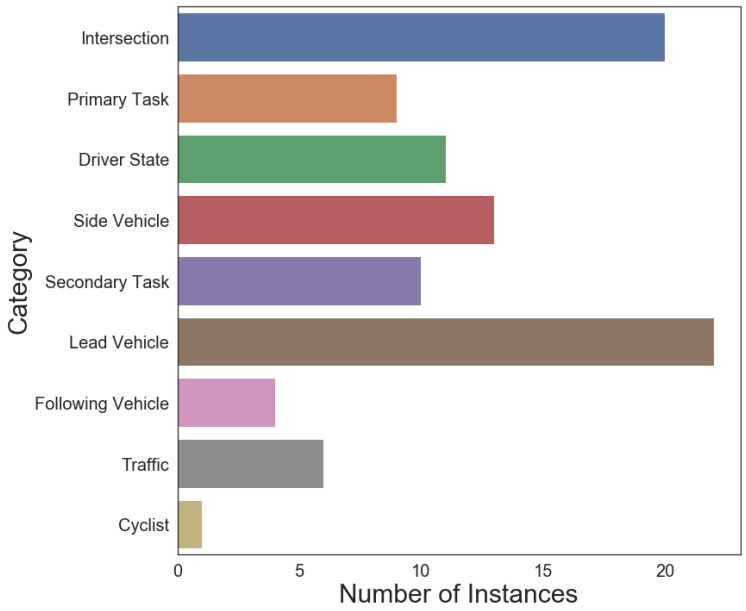


Figure 5.2: Count in each category of events happening in all 10 participants' data. Note that the two top occurring categories are intersection and lead vehicle. It's important to note that current autonomous systems have a one size fits all approach for situations when these parameters are present (e.g., a common car-following distance) which can result in different states for their users. ©2021 IEEE

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.

- 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. 5.2 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, (120) 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 (67). 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 in-

terest 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.

5.1.6 Discussion on HR Utility as a Feedback

[4] 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 (84). 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 (84). 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

[4] ©2021 IEEE. Reprinted, with permission, from (111)

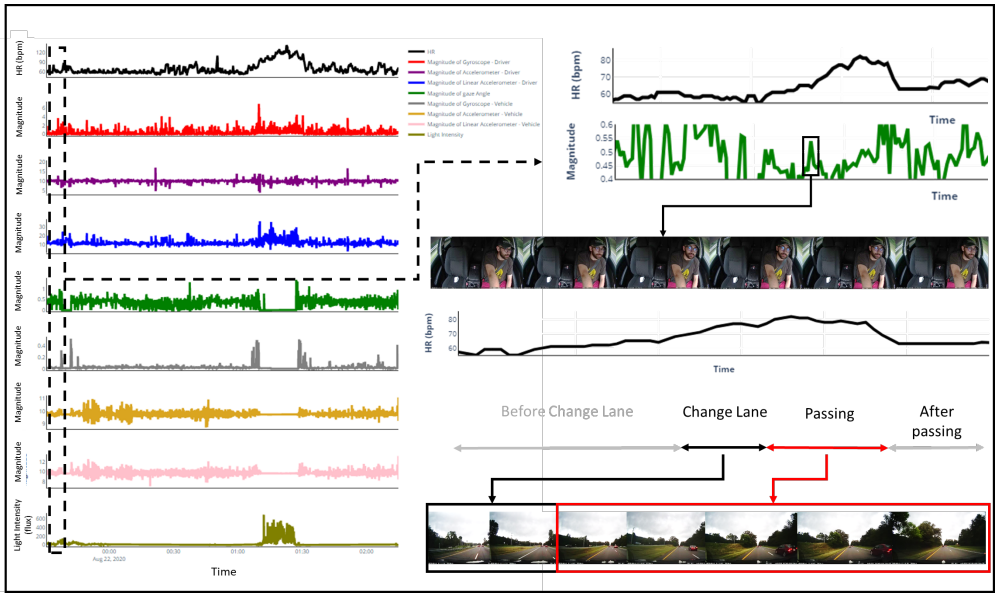


Figure 5.3: Different modalities of data in a lane change action (left). Looking deeper into the physiological data provides insight on how events like lane change can result in a prolonged increase in HR, even post removal of the initial stimulus. (right). It's important to note that it's hard to get such psychophysiological insights based purely on modalities like vision. The video for this event is provided as a demo for the paper. ©2021 IEEE

dataset. Fig. 5.3 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 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.

5.2 Project 2: Understanding the Differences Across Environmental Attributes on How They Affect the Driver's State

5.2.1 Introduction

Designing human-centered vehicles require understanding how different factors in and out of the vehicle can affect a user's state, such as stress level, workload, and anxiety (15). This is mostly due to the fact that drivers' decision-making and resulting behaviors are affected by their emotional and cognitive states, which ultimately have severe impacts on driving

safety. For instance, recent studies through naturalistic environments suggest that negative emotions, higher stress levels, and cognitive load may increase the probability of driving accidents (21). While multiple studies are targeting emotion detection in-cabin, not that many are focused on understanding which part of the environment (in-cabin and on the road) might be associated with changes in driver's states such as inducing negative emotions and higher cognitive load (120). In line with addressing this gap, recent research has started to analyze the driving environment together with subjective self-reports of driver's stress to find possible correlations between environmental attributes and driver's stress level. For instance, recent studies found that certain road objects such as bigger vehicles (e.g., trucks), road users (e.g., cyclists), and infrastructural elements (e.g., intersections), as well as in-cabin situations (e.g., working with the center stack), are highly associated with higher subjective stress levels (120; 22; 15). In addition to road objects, some research studies showed that car-following distance and behavior could also affect drivers' psychological states (79). For instance, (121) showed that within a simulated platooning scenario, drivers' mental stress measured through biosignals increased as the distance to the lead vehicle decreased. Another very recent study performed in a driving simulator found out that drivers' workload was higher during a shorter time headway (88).

Understanding the aforementioned reasons behind drivers' state including emotions, workload, and stress induction, can help mitigate them in driving by choosing less stressful routes (108), personalizing car-following distance (79), or by providing interventions (e.g., listening to music) (30). While previous studies provided significant evidence on the correlation between drivers' subjective measures of stress and environmental attributes, it is still difficult to apply these findings in real vehicles as most of these studies are centered on subjective measurements. Recent developments in ubiquitous computing devices such as smartwatches are facilitating their applications in detecting unhealthy states of the users such as abrupt changes in anxiety, stress level, and experiencing negative emotions at both individual and community level (116; 111). Using ubiquitous computing devices, studies have found strong correlations between human psychophysiological measures (e.g., HR and skin conductance) and stress level and work load, and more specifically in driving, studies show that increase in human HR might be correlated with stressful experiences (80; 18; 64; 54; 111).

In addition to smartwatches, advancements in computer vision techniques have made it viable to detect certain objects fully automatically without any manual annotation. For example, recent developments in end-to-end object detection algorithms (e.g., see MASK RCNN (2)) together with high-quality datasets (e.g., see COCO (61)) have made it possible to detect many road objects such as signs, road users, and infrastructural elements. Even in the case of not having access to off-the-shelf models, current computer vision frameworks

are making it easier compared to the past to build newer models for different road objects. Coupling smartwatches with features extracted from videos can help with finding possible associations between the presence of road objects and drivers' stress changes, objectively. This information can then be leveraged to enhance the driving experience and mitigate possible faulty decisions as a result of being under stress, ultimately decreasing driving accident rates. In the context of automated driving, these methods can help with take-over control of the vehicle in a much more efficient and faster manner.

In this paper, we take an explorational approach to understanding the relationship between changes in drivers' stress levels and emotions in real-world driving context by using multimodal naturalistic driving data, which includes drivers' psychophysiological measures, and behavioral metrics (vehicle speed) as well as driving environment videos. Based on a naturalistic driving dataset, namely HARMONY (111), we first retrieve drivers' facial expressions as well as abrupt increases in their HR, which might be indicative of increases in stress level (111). We detect abrupt increases by using a change point detector based on Barry and Hartigan's method (9). We analyze the driving scene retrieved from the video recordings by (1) detecting road objects and (2) estimating the relative distance to the lead vehicle. By analyzing the co-occurrence of the abrupt increases in drivers' HR and the presence of lead vehicles, we find that different road objects might be associated with varying levels of increases in drivers' HR, indicating different stress levels as well as different fractions of negative facial emotions. Our results indicate that larger vehicles on the road, such as trucks and buses, might be associated with the highest amount of increase in drivers' HR as well as negative emotions. Additionally, our findings show that shorter distances to the lead vehicle in naturalistic driving, as well as the higher standard deviation in the car-following distance, might be associated with a higher number of abrupt increases in drivers' HR, showing a possible increase in stress level. Moreover, our findings indicate more positive emotions, more facial engagement, and a lower number of abrupt changes in HR at a higher speed of driving. This research lays the ground for designing human-centered vehicles, urban environments, and services that can understand the level to which each road object and environment might affect drivers' and passengers' well-being.

5.2.2 Background

Studies in the past have provided significant information on the interplay of drivers' unhealthy states (e.g., emotions, stress level, anxiety, and cognitive load) and the environmental attributes. These studies have shown that environmental attributes such as in-cabin situation, road type, road users, weather, and in-cabin condition can well affect how a driver feels and can result in affecting their driving performance and behaviors both in automated and manual driving (120; 22; 15). As this paper is mostly centered on drivers' negative

emotions and stress levels among the other drivers' states, we review these two concepts in great detail as follows.

Understanding emotion and its applications have been one of the main topics of psychology, philosophy, neuroscience, artificial intelligence and computer-human interaction. If we go back in the history, the study of emotion has a history all the way back to ancient Greece (93). Merriam-Webster dictionary defines emotion as "A conscious mental reaction (such as anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body". Psychology literature provides different theories of emotion, such as the categorical and dimensional emotion theories. The categorical, sometimes referred to as the basic emotion theory, posits that there is a specific limited number of emotions that are basic psychological and biological concepts and cannot be divided into more basic ingredients. Although there is not a consensus on the number of the emotions and as to which emotion is a basic emotion, but there seems to be an agreement on the definition of the basic emotion. These emotions are distinct in their recurring fixed patterns of neural and bodily expressed components and physiological and behavioral signatures such as variation in heart rate, and facial muscle movements (47; 114; 25) which is in response to a stimulus. Different psychologists proposed a different number of basic emotions and accounted various emotions as basic. For instance, Izard proposes six basic emotions of happiness, sadness, fear, anger, disgust, and interest, while Ekman proposes the basic emotions to be happiness, sadness, fear, anger, disgust, contempt, and surprise (114).

Dimensional emotion theory proposes that emotions can be represented with numerical values in multiple dimensions. One of the famous dimensional emotion theories is Russell's dimensional emotion model, where emotions are represented by their valence and arousal in a two-dimension format (92). Valence refers to the level of positivity and negativity of emotion, whereas arousal refers to the level of activation in each emotion. In this model, an emotion such as "excited" has relatively high positive valence and high arousal, whereas an emotion such as bored has a negative valence with very low arousal.

Driver stress is defined as the process of facing a situation where the perceived demand, mostly defined based on the previous experiences, internal body sensations, and external stimuli, is higher than the available resources (33). Stress can happen at different time scales where short-term stress is referred to as acute stress, in contrast to long-term stress, which is referred to as chronic stress (33). Multiple studies in driving research have attempted to detect changes in emotion and stress level through measuring human physiological metrics such as facial expressions, cardiac measures, and skin temperature and conductance (17; 36; 55; 107). These studies are mostly based on the assumption that changes in human

physiology follow similar patterns within each specific emotional state. For example, studies show that increases in human HR might be correlated with an increase in stress level and negative emotions (55). Additionally, studies have shown that the movement of facial muscles within each emotion category might follow specific patterns, where computer vision applications can be leveraged to detect emotions from the facial expressions (69). Additionally, in a similar approach, certain patterns can be recognized while experiencing stress and can be recognized through machine learning applications (36).

While multiple studies focused on detecting unhealthy states (e.g., stress level), not that many studies analyzed the reason behind the elicitation of each state (120). Understanding emotion triggers is of high importance as it helps with planning for interventions in driving, which can then help mitigate the effect of negative emotions and stress levels on drivers' performance, decision making, and take-over control. Mesken et al. analyzed three of the drivers' emotions (anxiety, happiness, and anger) by using an instrumented vehicle monitored by an experimenter in the vehicle (73). In their study, the authors monitored 44 drivers in an on-road controlled study through speed, videos, and their HR. The participants were asked to verbally talk about their emotions as they faced any situation in driving. They found out that the emotion with the highest frequency was anxiety which was followed by anger and happiness. They also found out that emotions were related to traffic events. The authors also pointed out that drivers' anger was associated with driving events that might affect their progress, while anxiety was related to driving events affecting safety. Additionally, the authors report an increase in HR associated with anxiety situations (73). Roild et al. analyzed the responses of drivers' regarding the emotions that they experienced through a short survey (91). In their study, the authors asked participants to rate their daily emotions in driving through an online questionnaire. Authors found out that drivers' anger, anxiety, and positive emotions were strongly related to situational factors (91). Their results also point out that higher task demands are correlated with higher negative emotions (91). Later, a study by (120) monitored 33 drivers for a duration of 50 minutes through an on-road controlled study without an experimenter being present in the car. The authors also asked drivers to talk about their emotions as they faced them during the driving scenario. The authors analyzed 531 self-reports of drivers' voice recordings and provided four main categories of emotional triggers. The main categories included traffic & driving task, environment, HCI & navigation, and vehicle and equipment, which involved a few subcategories such as weather, other road users' behavior, and road designs (120).

Another study by (22) performed a similar analysis by monitoring 34 drivers' emotions with a focus on spatiotemporal triggers of emotions within an urban environment and found out that the main hotspot of emotional triggers in driving based on self-reports are intersec-

tions. Additionally, they found out that environmental attributes such as other road users' behaviors and traffic lights had a higher fraction of negative emotions as compared to positive emotions within the self-reported stress. In their study, the authors point out that within the urban environment, highways are associated with the least stress level. While most of these studies were centered on self-reports, another study performed by (111) found out that different characteristics of the road environment might be associated with increases in drivers' HR. In their study, the authors found out that being followed by a vehicle, following a lead vehicle too closely, arriving at an intersection, and performing secondary tasks might be associated with abrupt increases in drivers' HR, possibly showing stress and negative emotions. Lastly, a study by (15) attempted to predict self-reports of stress by solely relying on the snapshots of the visual scene. Based on using a convolutional neural network, the authors were able to predict the self-report associated with each visual scene in their database with an accuracy of 72 %. Authors also point out that certain objects in the visual scene are correlated with higher subjective stress levels, such as traffic lights, bigger vehicles, and the presence of riders. Lastly, it should also be noted that previous studies in this area were mostly conducted through on-road controlled studies as well as using a driving simulator.

Previous studies pointed out two major points regarding drivers' stress level and emotions. Firstly it is evident that drivers' stress level and emotions is affected by the driving environment such as certain objects in the visual scene such as presence of other road users, especially vulnerable road users, intersections and traffic lights, as well as lead vehicles and other traffic signs. Secondly, studies show that higher stress level in general is also correlated with increases in HR. Based on the above literature we hypothesize the following:

The hypotheses for this paper are as follows:

1. From the road object point of view, different environmental attributes might be associated with different levels of changes in a participant's HR. For instance, it can be the case that an intersection might be accompanied by a higher increase in HR as compared to a pedestrian.

2. Variation in the car-following distance might be correlated with stress level, which in turn is accompanied by changes in drivers' physiological responses.

5.2.3 Methodology

This section is divided into multiple subsections describing the dataset (section 5.2.3.1), environmental perturbation detection (section 5.2.3.2), detection of changes in drivers' HR (section 5.2.3.3), detection of drivers' facial emotions (section5.2.3.4), and detection of distance to the lead vehicle (section 5.2.3.5).

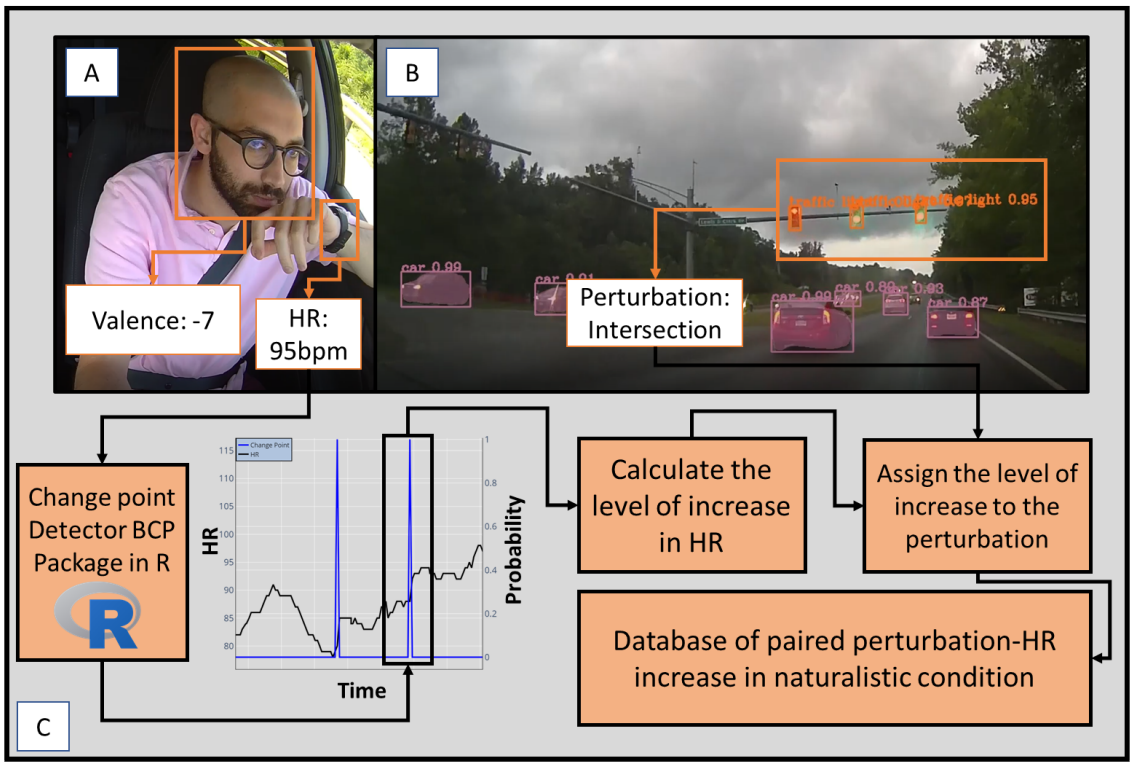


Figure 5.4: A general view of the data from both in-cabin (A) and on road (B), as well as the application of change point detector for finding moments of high stress in drivers' HR (C).

5.2.3.1 Dataset

The dataset for this study is provided by HARMONY, a human-centered multimodal study in the wild (111). This dataset includes driving as well as human sensing data from 22 participants. The dataset is collected in a naturalistic fashion where each participant is provided with a camera and a smartwatch. The participants were asked to drive as they normally would in their daily lives. The camera recorded both in-cabin and outdoor environmental conditions. Additionally, the smartwatch collected the driver's heart rate, hand acceleration, location, and environmental features such as noise and light level in-cabin. To this end, we used the data from 15 participants. A sample of the data is available online at (115). Additionally, figure 5.4 - A and B shows a view of the data from both in-cabin and on-road situation points of view.

5.2.3.2 Perturbation Detection

In order to simplify the process of detection and analysis, this section is mainly focused on seven major environmental perturbations that were previously mentioned in driving research as stressors on the road (120; 22; 15). These categories include the presence of speed limit signs, stop signs, intersections (i.e., traffic lights), big vehicles such as trucks and buses,

riders such as bicycles and motorcycles, and pedestrians. Note that any other category can also be added to the analysis hereafter. However, we only focus on a set of perturbations that were already shown to affect subjective stress levels.

5.2.3.2.1 Detection of Truck, Bus, Motorcycle, Bicycle, Traffic Light, and Pedestrian

In order to detect truck, bus, motorcycle, bicycle, traffic light, and pedestrian, we used an off-the-shelf computer vision algorithm namely MASK RCNN. In this section, we used a pretrained model of the MASK RCNN algorithm (2) that was trained on the Common Objects in Context (COCO) dataset (61). A sample of the detection can be seen on figure 5.4 - C.

5.2.3.2.2 Detection of Stop Signs and Speed Limit

In order to detect stop and speed limit signs within the pool of collected videos, we take advantage of the recent computer vision (CV) applications in sign detection. Current off-the-shelf CV algorithms (e.g., MASK R-CNN (40)) that are trained on large datasets (e.g., COCO (61)) are capable of detecting stop signs. However, contrary to the test set provided by these algorithms, once applying them to real-world videos collected in HARMONY, they often detect any sort of traffic sign as a stop sign, which increases the false positive rate. Additionally, if a stop sign is not facing the driver, it will still be detected, which is not applicable to our case as we are interested in stop signs that might be associated with a change in a driver's state.

To overcome these issues, we first retrained an object detection model on a stop and speed limit sign dataset, which was created by merging three external sub-datasets. The model and code for this section are available through our GitHub (103) . We use transfer learning to retrain a model for detecting stop signs. In this regard, we take advantage of the YOLOV5 model, which is a recent modification of a well-known deep learning object detection architecture, namely YOLO. Below, we provide an overview of YOLOV5.

YOLOV5: You Only Look Once (YOLO) is an object detection algorithm that was introduced by (89). This model reimagines object detection as a regression problem and is inspired by the human visual system. Previous works in object detection generally apply a sliding window on the image of the interest to detect different objects (89). More specifically, methods such as MASK-RCNN (40) first apply a sliding window to detect bounding boxes. Each bounding box is then fed into a classifier for labeling based on the object of interest within the bounding box. This is extremely time-consuming as each part of the model has

to be trained separately. On the other hand, YOLO is based on simultaneously predicting bounding boxes and the probability of each label associated with them. In this regard, YOLO can see the whole image at once and predict for the whole image rather than each individual bounding box, which results in less number of false positives (89). YOLO architecture has 24 convolutional layers followed by two fully connected layers, as shown in detail in (89). YOLO initially was suffering from different limitations, such as struggling with small objects in the visual scene. Different modifications were then added to the base YOLO, which resulted in YOLO versions 2 to 5. For this work, we focus on the most recent version of YOLO, which is YOLOV5 introduced by (52; 28).

Sign Detection Datasets: For this paper, we first trained YOLOV5 on a dataset of stop and speed limit signs. As mentioned previously, the dataset was created by merging three datasets of Laboratory for Intelligent and Safe Automobiles (LISA) (49), Common Objects in Context (COCO) (61), and the Balali Sign Dataset (5). While both the LISA and Balali Sign datasets are only focused on signs (e.g., stop signs, warning, and go left), the COCO dataset includes many road objects such as trucks, sedans, motorcycles, bicycles, and traffic lights. From combining the three datasets, 8042 images were used as training which comprised 4154 stop sign and speed limit images. Within the stop sign and speed limit pool of images, 2420 images were from the LISA dataset (1291 stop signs and 1129 speed limits), and 1734 stop signs from the COCO dataset. Additionally, the training set included a total number of 3888 negatives (none of the stop or speed limit signs) images, which comprised 250 negative images from the Balali Sign Dataset, as well as 2638 negative images from the LISA dataset. Lastly, 721 images were used as a test set, in which 652 images were from the LISA dataset and 69 images were from the COCO dataset.

Utilizing YOLOV5 model, we trained the base YOLOV5 model for 200 epochs with the default hyperparameters. Figure 5.5 shows the mean average precision at 0.5 (mAP@0.5) for the stop sign and speed limit dataset.

Optical Character Recognition (OCR): In order to further enhance the accuracy and to differentiate between speed limit signs, we integrated an Optical Character Recognition (OCR) (81) to detect signs that only have a specific text written on them. Using this method, we separate the signs that are not showing "STOP" or a number such as "25" on them. OCR is referred to the transforming of images into printed text. In order to apply OCR, we have tested the PyTesseract (85), and EasyOCR (24) packages written in Python scripting language. Our initial testing on the two packages showed that EasyOCR is slower but much more accurate in the images retrieved from real videos collected through HARMONY. Thus we continue with the EasyOCR package. Every detection of the speed limit or stop sign is fed into the EasyOCR package. In the case of not detecting any character, it will be

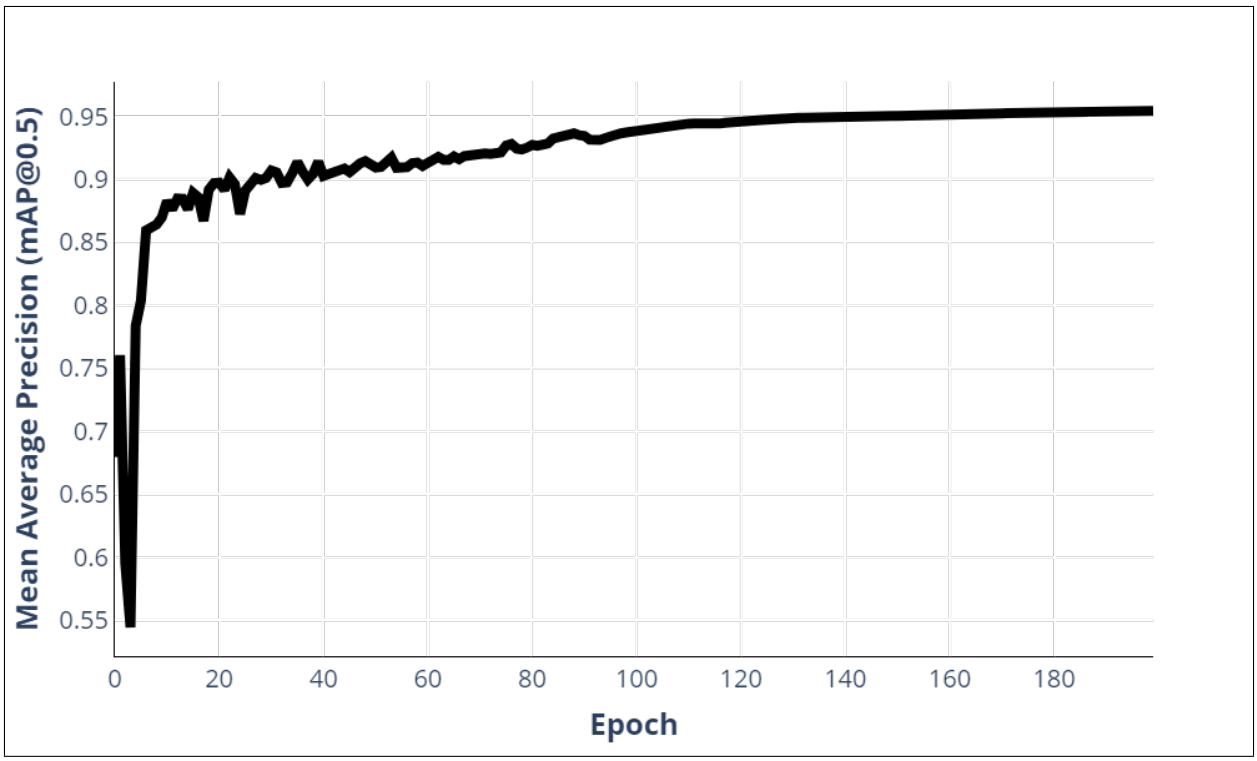


Figure 5.5: The mean average precision of training yolov5 for stop sign and speed limit dataset

automatically removed from the detections. In the case of detecting "Speed Limit", we will then also find the number showing the limit. The whole process above takes between 3-5 minutes for each video, depending on the chosen frame per second.

5.2.3.3 Detecting Abrupt Increases in Drivers' HR

In order to detect the abrupt increases in drivers' HR, which, as mentioned previously, is correlated with increases in drivers' stress level, we take advantage of a change point detector. Due to motion artifacts introduced through different movements of drivers' hands, momentarily peaks can exist in the HR data, which is not of interest to our analysis. We are rather interested in detecting a change in the underlying distribution of the HR data. For this matter, we take advantage of a Bayesian Change Point (BCP) detector. BCP allows for easy quantification of uncertainty and integration of priors. Other studies have also mentioned the utility of BCP in detecting changes in data from different fields such as health (65), driving (111), and behavioral science (60). We leverage Barry and Hartigan's (9) Bayesian change point model for this analysis. This model generally assumes different blocks of HR data within the time series of HR in a way that within each block, the mean is constant. The model then calculates the probability of entering a new block as a change point probability. in order to perform BCP on the HR data, we use the *bcp* package written in R programming

language (27). The BCP is applied to each participant's HR data, and the probability of change at each point is extracted.

After detecting both change points in HR and the presence of certain road objects, we use a window of 10 seconds around each change point in HR to search for the presence of each road object based on the computer vision detection. In the case of the presence of certain road objects, we will search in the next 10 seconds after the detection of the change point for the maximum HR value. Using the HR max values, we define the reaction to the road object as the difference between the HR at the moment of change point and the HR max value. These values are then used to model the individual differences in these environmental perturbations. We use linear mixed effect models to understand the variation in drivers' HR responses around each environmental attribute. Linear mixed effect (LME) models are similar to a simple linear regression with the difference that it accounts for the variability across participants in their responses as random factors while accounting for the effect of fixed factors (each perturbation). In the case of facing a dependent variable that is of a count nature, we use a generalized linear model with a negative binomial process distribution (13). The idea behind LME is described in detail in (14; 32). The analysis above is performed through the lme4 (10) package written in R programming language (86).

5.2.3.4 Detection of Drivers' Facial Emotions

In order to detect drivers' facial emotions, we leverage the Affectiva module on the iMotion software (68; 69). Previous research has shown the utility of this software in detecting facial expressions and their positivity/negativity level, as well as detecting basic emotions and specific facial muscles (59; 1; 72; 90; 104). For this paper we focus on the two measures of "engagement" and "valence". Engagement refers to the level of showing any signs of emotion in the face with a value of 0 (no emotion) to 100 (highest showing of emotion), and valence is a measure of positive or negativity of emotion with a value between -100 (most negative) to 100 (most positive). After performing the analysis with Affectiva, all the frames that did not have any detections were removed from the database. This can be due to the angle of the camera as well as lighting issues, which did not account for a significant portion of the data.

5.2.3.5 Pixelwise Distance to the Lead Vehicle

In order to calculate the pixel-wise distance to the lead vehicle, we have used a combination of lane detection and object detection algorithms. For this task, we used a version of the YOLO algorithm titled, YOLOP: You Only Look Once for Panoptic Driving Perception (117). We applied YOLOP on the outside videos and retrieved lanes, vehicles, and driveable areas. We initially used both lanes and cars, but upon finding YOLOP's car detection to

Figure 5.6: The methodology to calculate the pixelwise distance from the lead vehicle. (A) detection the lead vehicle using a combination of lane tracking and object detection. Enhancing the high number of false positives in detection as the lead vehicle (B) through post-processing (C)

be inaccurate on cars directly in front in our specific videos, we supplemented the detection with external bounding boxes from MASK RCNN (40). Using YOLOP we detect what lane a detected car is in. The modified program assigns lane numbers to every car detection. Lane 0 represents a car directly in front in the same lane, a negative lane number represents cars to the left, and positive lane numbers represent cars to the right.

YOLOP's lane detection creates a pixel mask representing lanes. From the pixel mask, a center is marked, and the amount and location of lanes on either side of the center are found. Then on each car detection, the lane number is assigned to the detection according to the car's relative position to the lane detections. Additionally, a line is drawn from the bottom center of the mask to the center of each car, and the number of times this line intersects a lane is counted. Zero intersections mean the car will be in the center lane, while one intersection would mean the car is in the line directly adjacent to the current lane (figure 5.6 - A).

After detection is performed, post-processing is applied to remove outliers and fill in detections in gaps, as car and lane detection will not always provide accurate results, especially when they examine one frame at a time (figure 5.6 - B). Post-processing also ensures only a single car can have lane number 0 (in front), as only a single car will be visible in front and in the same lane. In some cases, lane detection will not detect lanes at all, and declare all cars visible as the car in front. Post-processing will correct this by assigning the car closest to the center and with the largest bounding box as the car in front while assigning other cars to the side lanes (figure 5.6 - C).

5.2.4 Results

5.2.4.1 Relationship Between Road Objects and Drivers' Psychophysiology

Figure 5.9 shows the distribution of the percentage of increase in drivers' HR after each detected change point. Note that this percentage is calculated based on dividing drivers' maximum HR in the 10-second time window as defined in methodology section by drivers' HR at the location of the detected change point. We have also marked the location of the mean, as well as one standard deviation, two standard deviations, and three standard deviations away from the mean by red, black, blue, and purple lines, respectively. This will help us better visualize the result in the next sections. Additionally, note that due to the imbalanced nature of the dataset (e.g., unequal number of instances with trucks versus pedestrians), we have performed oversampling based on Synthetic Minority Oversampling Technique (SMOTE) (16) to generate new samples for the minority classes. SMOTE generates new data points from convex combinations of nearest neighbors. This will help us better compare the different categories. Additionally, we have grouped different categories of environmental perturbations for a better illustration. The rider group contains cyclists and motorcycle riders, and the bigger vehicle contains trucks and buses.

Figure 5.7 shows the average level of increase in HR for each road object category. As observed, the category of the bigger vehicle has the highest amount of increase which includes trucks and buses. This is then followed by pedestrians, traffic lights (an indicator of intersections), traffic signs, and riders. In order better understand the differences across these groups, we run a Kruskal Wallis test (70) over the different categories of road stressors associated with HR increase. A Kruskal Wallis test is a non-parametric test that assesses the differences across independent samples. This test shows that the categories of stressors are significantly different from each other, with a degree of freedom of 4, a chi-squared value of 34.14, and a p-value of 6.97e-7. We then ran a set of pairwise t-tests that were corrected using the Holm method (43). This is performed due to the fact that multiple comparisons are being made simultaneously. The results of the t-tests are shown in Table 5.2. As can be seen, most of the comparisons produce significant results other than the comparison between bigger vehicles and pedestrians categories.

It is important to note that we focus on the increase in HR rather than the actual value of HR. Figure 5.8 shows the standardized value of HR for each road object category. Note that the HR values are standardized per participant due to the inherent differences across different people's HR baselines. As can be seen visually, the trend is different than the level of increase in HR at each change point location. This might imply that although the HR values themselves might be higher for some perturbations, when a change in HR happens at their presence, the level of increase follows a different pattern for each perturbation. For

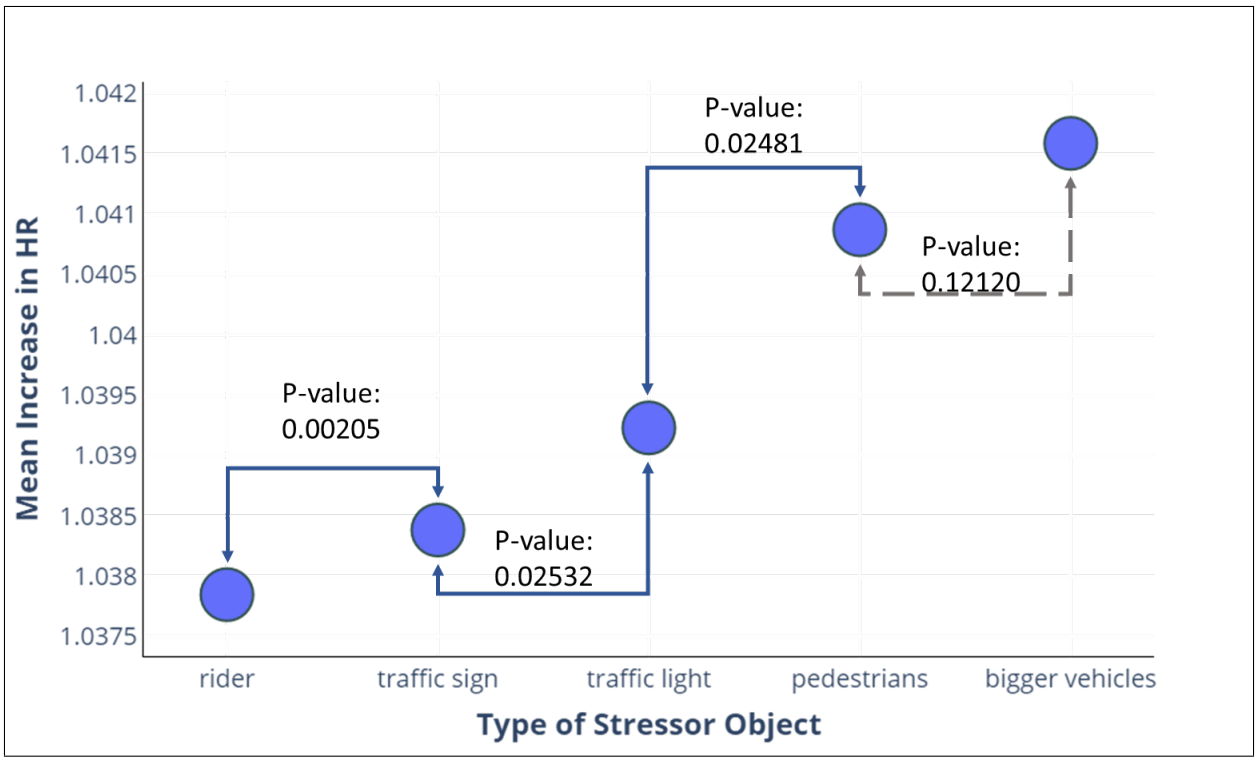


Figure 5.7: The average increase in HR at each change point location associated with each road object category.

instance, while the average HR value is lower for the category of traffic light in comparison to rider, the increase associated to it is more than the rider category.

Using the values in figure 5.9, we define different levels of stress to be low, medium, and high based on the level of increase in HR. More specifically, between μ and $\mu + \sigma$ represents low stress, between $\mu + \sigma$ and $\mu + 2 * \sigma$ represents medium stress, and between $\mu + 2 * \sigma$ and $\mu + 3 * \sigma$ or more represents high stress level. We then perform our analysis based on each of the categories defined.

Figure 5.10 shows the fraction of the presence of each road object within each stress category. On average, As shown, within each stressor object category, the fraction of the presence of each stress level varies, with the category of riders having the lowest fraction of high stress and bus and truck (bigger vehicles) having the highest. Note that although the rider category has the least high-stress level, it is, in fact leading the medium stress level category. In other words, the presence of a rider is most likely to increase the HR only as much as two standard deviations away from an average increase in HR. Additionally, note that almost 30% of all the increases in HR associated with trucks move at least two standard deviations away from the average increase.

We also visualized the individual differences across participants and how different par-

Table 5.2: The comparison between the HR increase associated with different categories of perturbations

	Bigger Vehicles	Pedestrians	Rider	Traffic Light
Pedestrians	0.12120	-	-	-
Rider	<2e-16	<2e-16	-	-
Traffic Light	0.00014	0.02481	1.7e-8	-
Traffic Sign	1.9e-10	1.7e-6	0.00205	0.02532

ticipants might vary in how they perceive different stressors based on their HR increase. As shown in figure 5.11, participant #12 had the highest level of HR increase, while participant #15 had the lowest amount.

Additionally, we observe that within different categories of stressors, participants are also reacting differently. For example, participant #14 has the least increase in HR due to bigger vehicles, while participant #19 has the highest level.

For valence and engagement, we have followed a similar analysis procedure as HR. In this section, we have first assigned categories to valence and engagement based on the mean and standard deviation of valence and engagement. Valence is categorized as negative, neutral, and positive, where the range between $\mu \pm \sigma$ is considered neutral. This is mostly due to the fact that valence values are often close to zero (showing no emotion). For engagement, we considered two categories of neutral and non-neutral engagement. In this fashion, all the values more than $\mu + \sigma$ are considered as non-neutral facial engagement.

Figure 5.13 shows the presence of each stressor category within each engagement level category. Note that similar to HR increases, bigger vehicles have the highest amount of high engagement, and rider is among the lowest categories. It is interesting that the proportion of having high facial engagement is more than 0.5 for the bigger vehicle category, which implies that when facing bigger vehicles, drivers are more likely to show some level of facial expression, while this value is the lowest for the traffic light category. In other words, detecting responses to traffic lights might be more feasible by using HR rather than facial expressions.

Figure 5.14 shows the presence of each stressor category within each valence level category. Similar to both engagement and HR changes, bigger vehicles are among the categories with the highest level of negative valence retrieved from facial expressions. Similar to facial engagement, the traffic light is mostly followed by a neutral facial expression which might indicate that this modality may not be suitable for detecting reactions to the traffic lights.

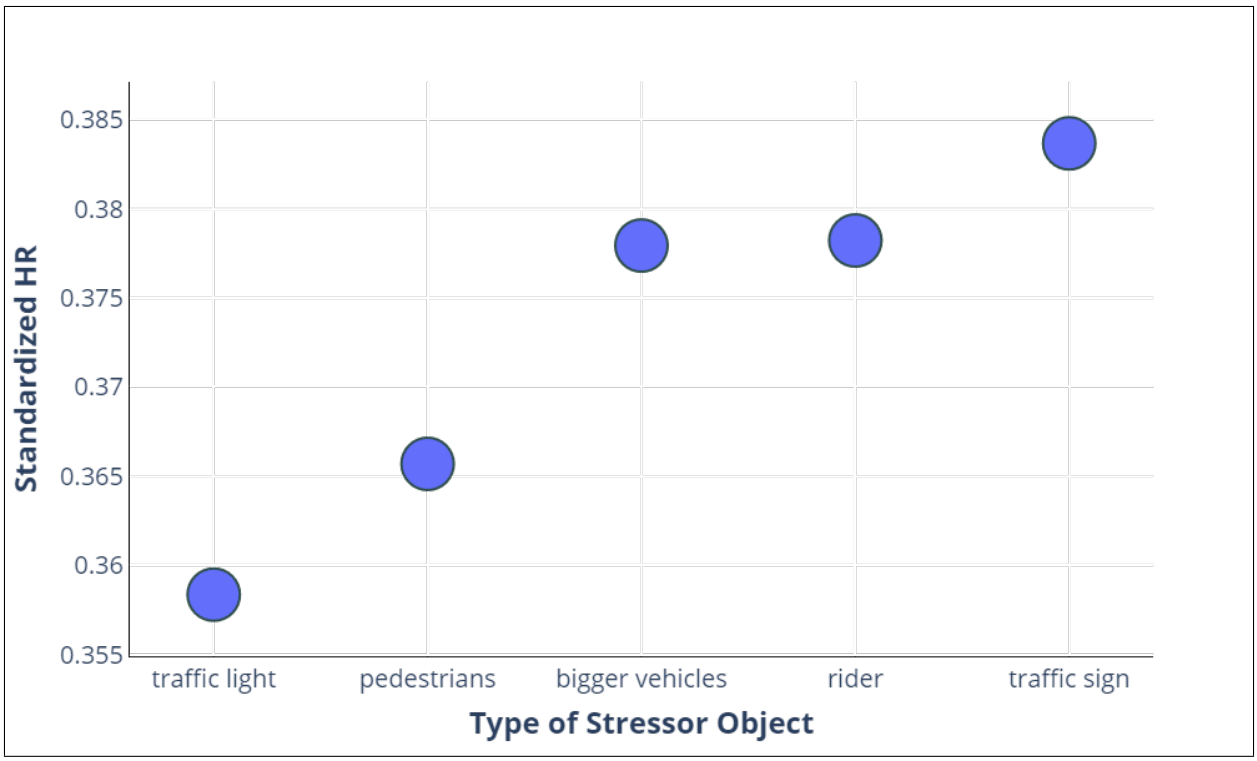


Figure 5.8: The average HR in the presence of each road object category.

5.2.4.2 A Detailed Analysis of HR in the Vicinity of Traffic Signs

In addition to the generic traffic sign detection, we have analyzed drivers' HR in the vicinity of stop signs and speed limit signs from the pool of detected traffic signs. These two traffic signs are categorized as regulatory signs, which might have a different effect on the driver as compared to the other signs. In this section, we are mostly interested in knowing the physiological pattern of drivers around these two regulatory signs. In order to find clusters in drivers' physiological metrics, we perform k-means clustering on the HR and speed signals around these two regulatory signs.

It is important to note that changes in human HR and, in general physiology can happen at different time scales with respect to the detection of certain road objects. For example, we don't exactly know when (e.g., how many seconds prior or after) a person might perceive a road object and react. Thus in order to compare the time series and further cluster them, we need to first bring different instances of the stop sign and speed limit occurrences into the same time frame using Dynamic Time Warping (DTW).

Dynamic Time Warping (DTW) is a famous technique for finding an optimal warping function to transform a time series to another one that might have differences in speed of happening in time (78; 11). In simple terms, DTW attempts to find shape functions that, when applied to a time series, can produce the other time series that may have been

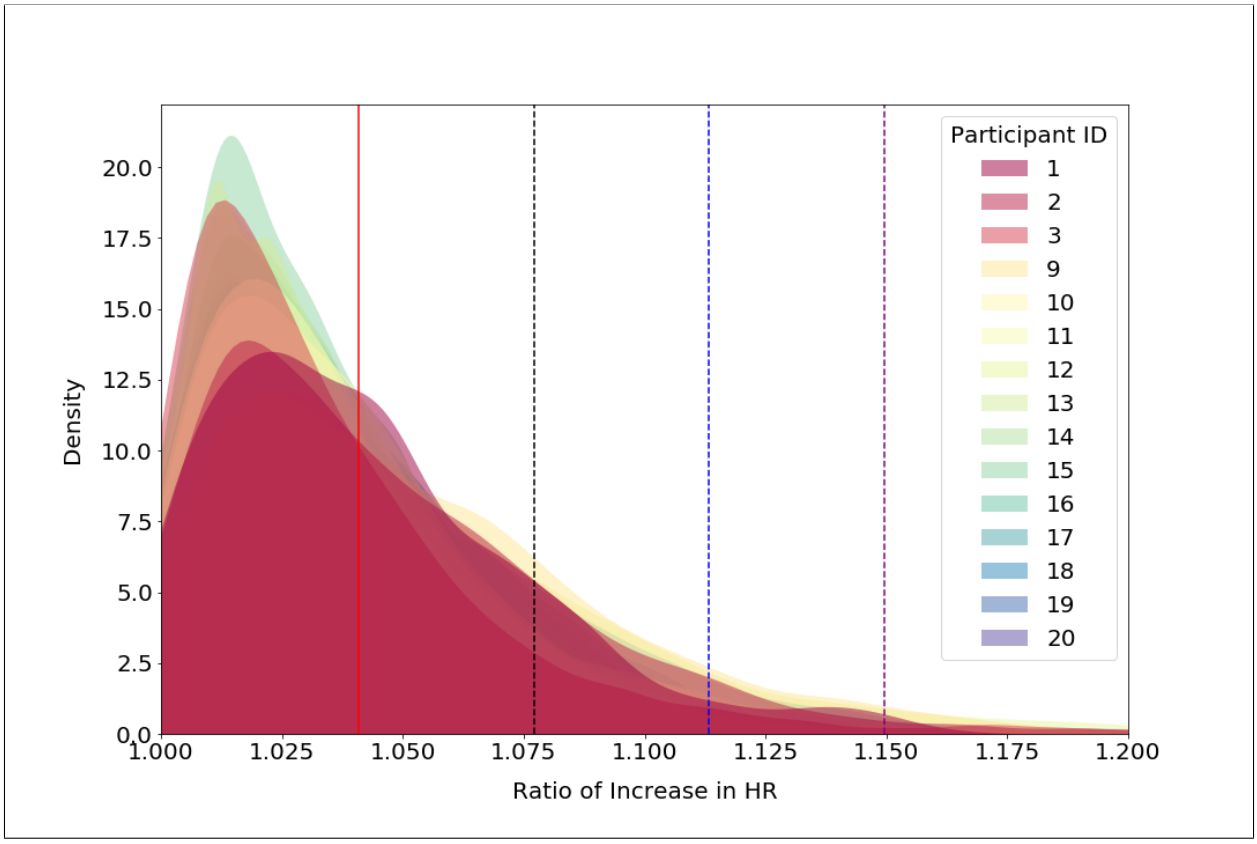


Figure 5.9: The distribution of increase in HR at location of change point in percentage for different participants. Note that the red vertical line show the mean, the black dashed line shows the mean + standard deviation, the blue dashed line shows the mean + two * standard deviation, and the purple dashed line shows the mean plus three * standard deviation.

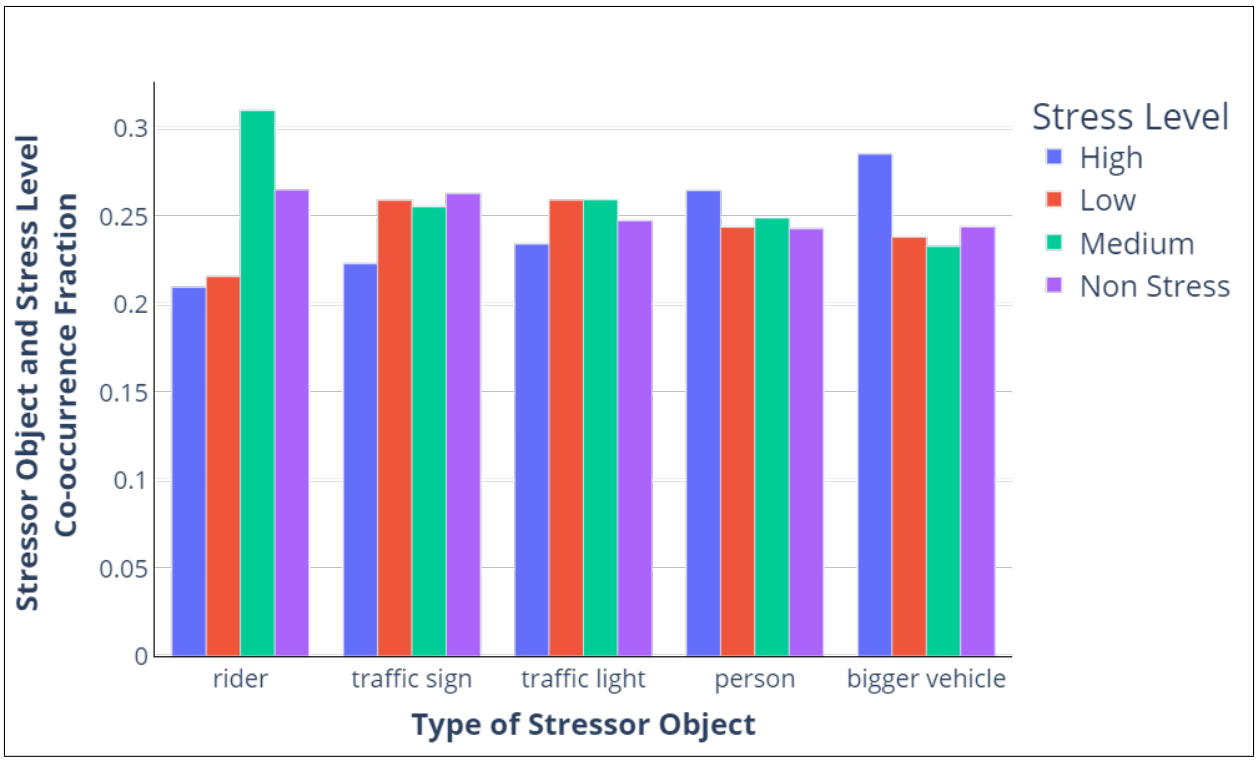


Figure 5.10: The fraction of presence of each stress category within each road object.

recorded in a longer timeframe. This is similar to finding a function that can translate a person walking at a specified speed to another situation of having a different speed but with the same pattern. For example, imagine the HR of participants when reaching a stop sign. Different participants might reach and pass through a stop sign with different durations, thus producing time series with different time frames that all reflect the same situation. We perform DTW using the *tslearn* package programmed in Python (112).

After performing DTW on the time series of HR in the vicinity of stop signs and speed limit signs, we performed k-means clustering. K-means clustering is an unsupervised approach to finding clusters within the data (48). This algorithm which lies under the partition-based clustering methods, performs based on assigning each point to a randomly initialized set of partitions based on their similarity. This procedure is performed until convergence. K-means algorithm is fast and simple to implement, thus attracting many research areas (48).

In order to define the number of clusters needed, we use the silhouette score (98). This score shows the quality of clustering by measuring how close points from different clusters are to each other (98). Silhouette score is a value between -1 and 1, which 1 indicates the most separation between clusters. We assess the Silhouette score for the different number of clusters and choose the number of clusters that produce the highest score, which in both

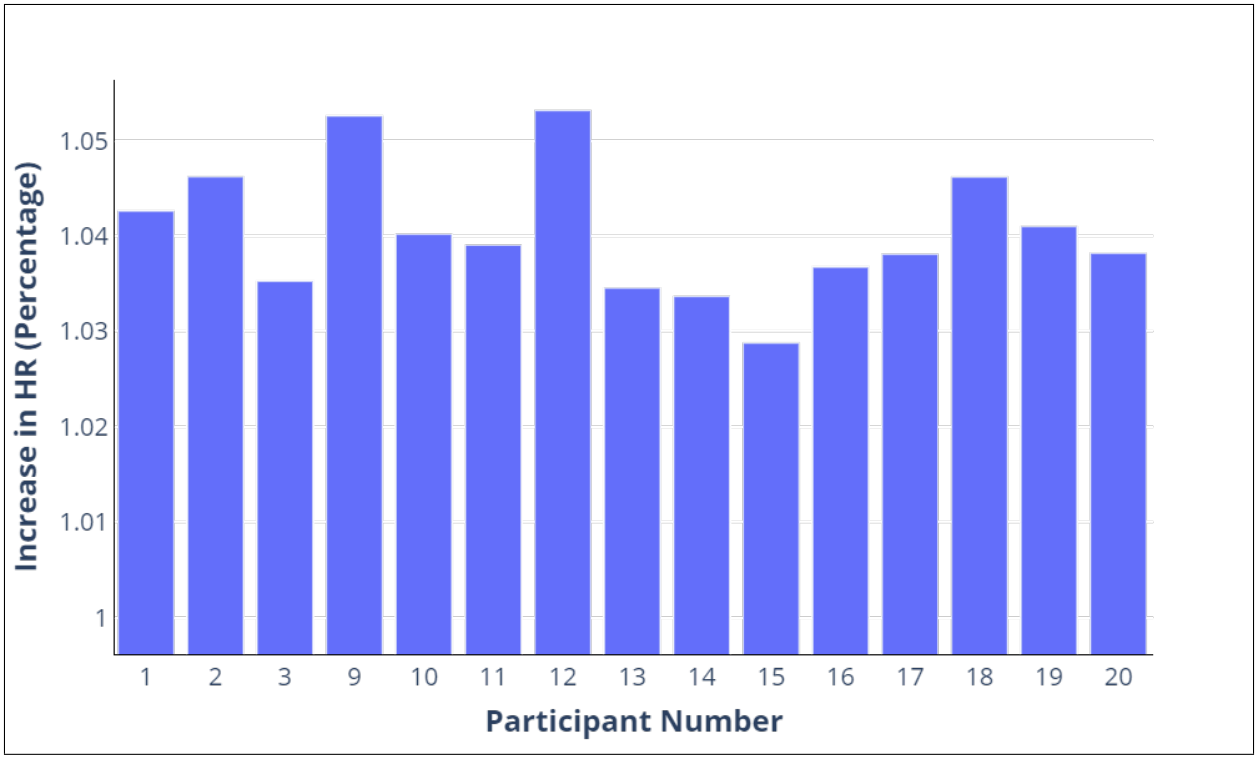


Figure 5.11: A comparison between the differences across participants for the average HR increase

cases of being close to a stop sign or speed limit sign, two clusters have been chosen.

Based on the two clusters detected as the optimal number of clusters, we apply the k-means clustering. Figure 5.16 shows the two clusters detected. Note that x-axis in the middle is the time of reaching a speed limit sign based on computer vision detection (Time = 7.5). While the two patterns around the speed limit signs are very different from each other at the first look, they have certain interesting characteristics. Cluster 1 is related to the cases that the HR is at its normal value prior to reaching the speed limit sign. Note the abrupt increase in HR in the vicinity of the sign. This cluster is also more dominant based on the number of times that it happens visibly from the figure. Cluster 2 is related to the cases where the HR is already higher than normal prior to arriving at the speed limit sign. It is interesting that even in this case, a small increase is observed at the location of speed limit sign detection (time =7.5).

The patterns in the vicinity of stop signs have a similar trend to the speed limit signs with some interesting differences (figure 5.17). First, in cluster 1, the increase in HR happens a few seconds after the detection of stop signs, which can mean a time difference between the effect of these two signs on human physiology. Second, the stop sign trend in cluster 2 has a downward trend after time =12.5 (see figure 5.17 , time $>$ 12.5), whereas this is not

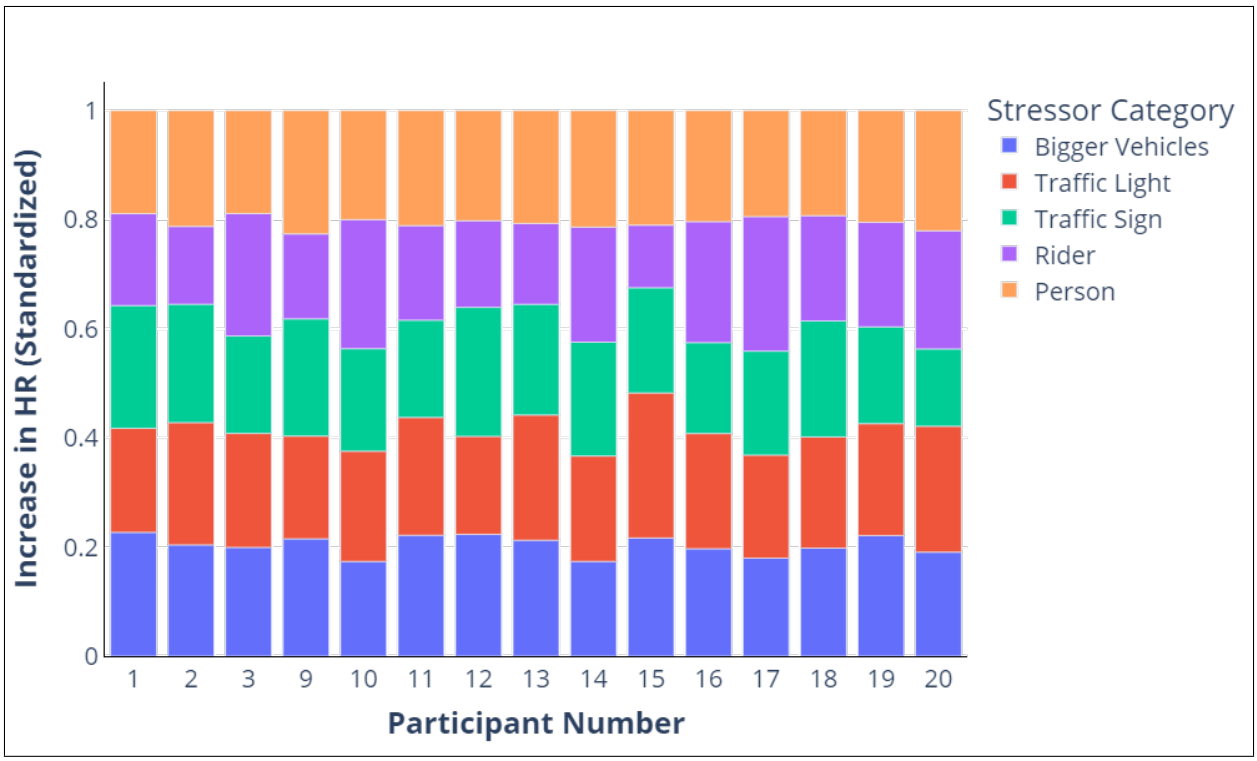


Figure 5.12: A comparison between the differences across participants for various stressors on the road.

the case for cluster 1 in speed limit (figure 5.16 - time $>$ 12.5). This might indicate that the speed limit sign has a more prolonging effect than the stop sign on human physiology.

5.2.4.3 Relationship Between Distance to the Lead Vehicle and HR

Figure 5.18 shows the number of abrupt increases in drivers' HR versus the average distance to the lead vehicle. Visual inspection of the figure suggests that a negative relationship exists between the average distance to the lead vehicle and the count of abrupt increases in drivers' HR. We test this relationship by using a generalized linear mixed-effect model (GLMM) with a negative binomial process, in which the independent variable is the average distance to the lead vehicle and the dependent variable is the count of the number of abrupt increases in HR, while considering the random effect of different participants' baselines with separate intercepts for each participant. We chose this model specifically by comparing it with a generalized model with no random effect. We chose the aforementioned model based on the AIC model comparison. As shown on Table 5.3, the GLMM model has lowest AIC value.

Table 5.4 shows the result of the chosen GLMM model for the mean distance to lead vehicle data. The results show that participants' HR had a higher number of abrupt changes as the mean distance to the lead vehicle decreased.

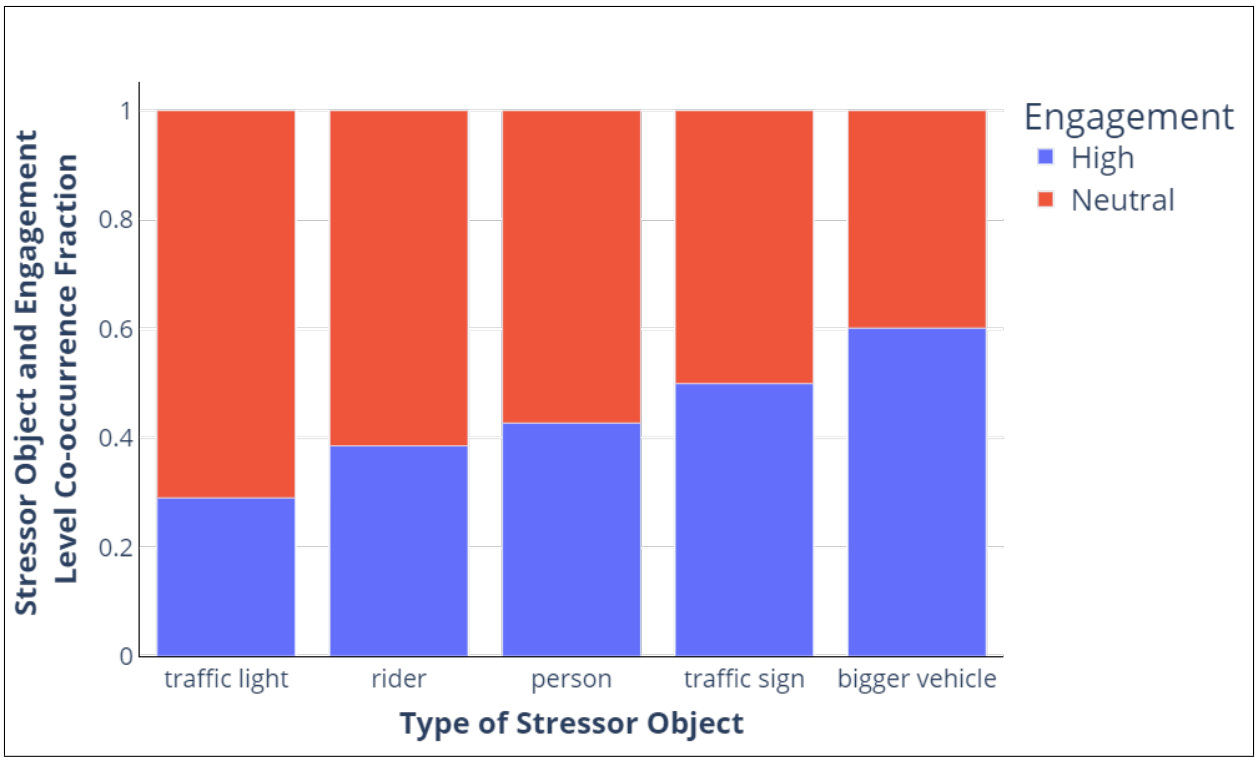


Figure 5.13: The presence of each road object in each category of engagement. Note that we provide a zoomed in version for the categories with lower values.

Table 5.3: The comparison between the two models chosen for modeling the relationship between abrupt increases in HR and average distance to lead vehicle. The lowest AIC in GLMM with random intercept is the basis for model selection.

Model Name	AIC	BIC	Loglikelihood	Chi sq	Df	Pr
Generalized Linear Model	6256.8	6273.3	-3125.4			
GLMM with random intercept	6246.4	6268.4	-3119.2	12.397	1	0.00043

Figure 5.19 shows the number of abrupt increases in drivers' HR versus the standard deviation of distance to the lead vehicle. Visual inspection of the figure suggests that a positive relationship exists between the standard deviation in the distance to the lead vehicle and the count of abrupt increases in drivers' HR. We test this relationship by using a generalized linear mixed-effect model (GLMM) with a negative binomial process, in which the independent variable is the standard deviation of the distance to the lead vehicle. The dependent variable is the count of the number of abrupt increases in HR while considering the random effect of different participants' baselines with separate intercepts for each participant. We chose this model specifically by comparing it with a generalized model with no random effect. We chose the aforementioned model based on the AIC model comparison. As shown in Table 5.5, the GLMM model has the lowest AIC value.

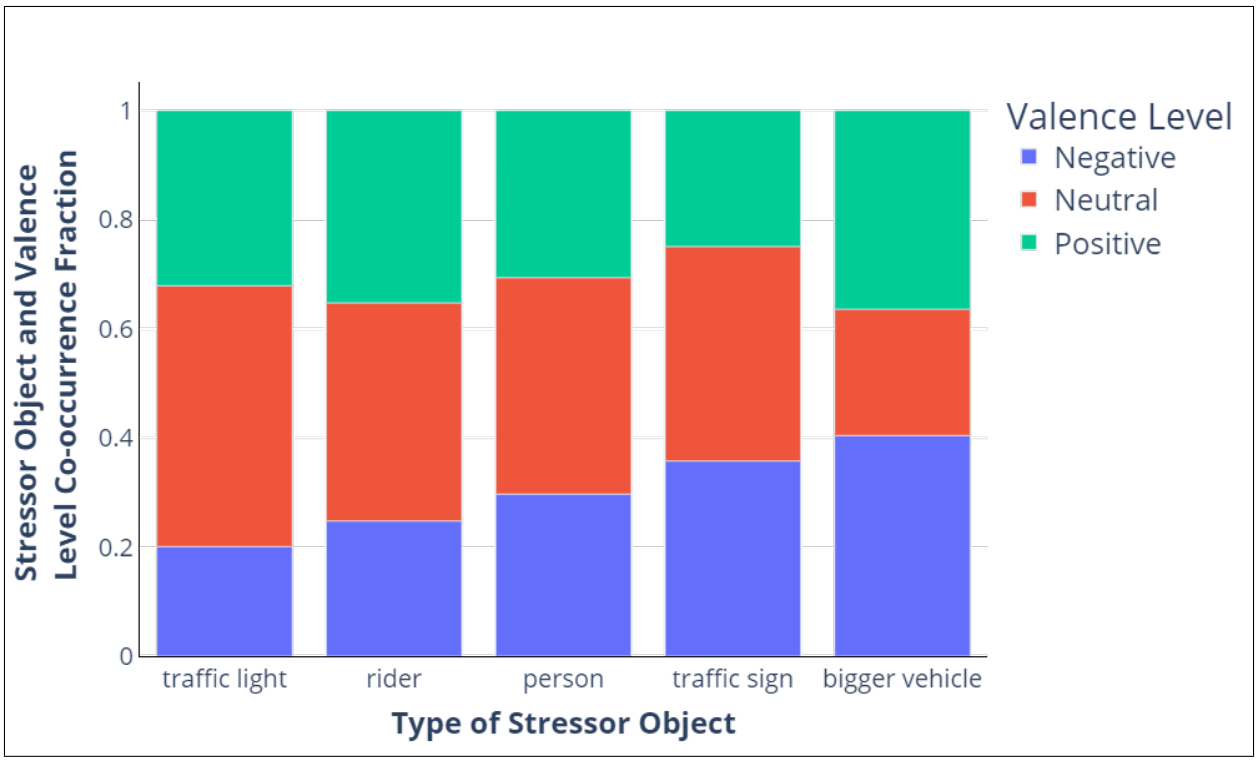


Figure 5.14: The presence of each road object in each category of valence. Note that we provide a zoomed in version for the categories with lower values.

Table 5.6 shows the result of the chosen GLMM model for the distance to lead vehicle data. The results show that participants' HR had a higher number of abrupt changes as the distance to the lead vehicle changed more sporadically (standard error = 0.01235, z-value = 11.308, p-value < 0.0001).

5.2.4.4 Relationship between Drivers' Speed and Psychophysiology

Lastly, we analyze the relationship between the drivers' speed and the level of increase in HR as well as facial expressions. Figure 5.20 shows the change in HR in the vicinity of change points in HR for two different environments of the city versus highway. While both environments show a declining trend with respect to the relationship between HR and

Table 5.4: The result of applying the GLMM model with random slope on the standard deviation of distance to lead vehicle versus abrupt increases in drivers' HR. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. The significant predictors at 0.05 level are underlined in the Pr column.

Effect	Estimate	Std. Error	z-value	Pr(>[z])	CI 2.5%	CI 97.5%
Intercept	0.38707	0.06511	5.945	2.7e-9	0.23752773	0.5160669
Distance	0.07928	0.03572	2.220	0.0264	0.01033641	0.1504898

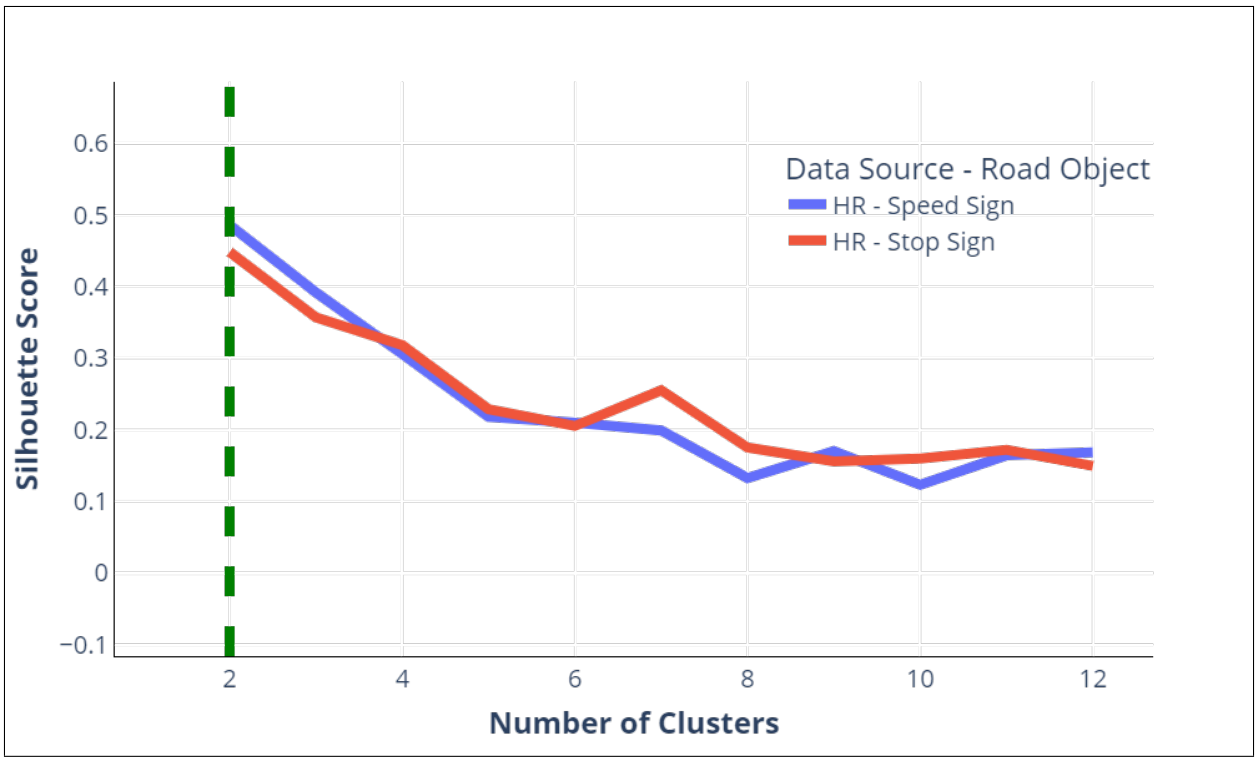


Figure 5.15: The Silhouette score for different number clusters for k-means clustering of drivers' HR in the vicinity of speed limit signs and stop signs. The optimum cluster number is 2

Table 5.5: The comparison between the two models chosen for modeling the relationship between abrupt increases in HR and standard deviation of distance to lead vehicle. The lowest AIC in GLMM with random intercept is the basis for model selection.

Model Name	AIC	BIC	Loglikelihood	Chi sq	Df	Pr
Generalized Linear Model	7335.1	7346.1	-3665.5			
GLMM with random intercept	7272.5	7289.0	-3633.2	64.590	1	9.22e-16

speed, a higher slope is observed for the highway environment. In this section, we treated the environment type as a random factor and assessed the relationship with a mixed effect model. Table 5.7 shows the result of the mixed effect model. Also, note that we chose this model as it had a lower AIC value compared to a simple linear regression (-256158 versus -256140).

Figure 5.21 shows the probability of change in HR at different speeds of the vehicle. As it is shown, higher speeds are accompanied by higher probabilities of abrupt changes in drivers' HR towards higher values. We tested this relationship using a linear regression model, and the results are shown in Table 5.8.

Figure 5.22 shows the relationship between drivers' facial expressions and speed. On

Table 5.6: The result of applying the GLMM model with random slope on the standard deviation of distance to lead vehicle versus abrupt increases in drivers' HR. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom. The significant predictors at 0.05 level are underlined in the Pr column.

Effect	Estimate	Std. Error	z-value	Pr(>[z])	CI 2.5%	CI 97.5%
Intercept	0.43270	0.01895	22.83	2e16	0.21162163	0.4832381
Distance	0.16748	0.01327	12.62	2e-16	0.08054035	0.3128414

Table 5.7: The result of the linear mixed effect model on the changes in HR at different speed in city and environment. p-values estimated via t-tests using the Satterthwaite approximations to degrees of freedom.

Effect	Estimate	Std. Error	df	t-value	Pr
Intercept	1.042	1.736e-3	1.041e+00	600.086	0.000825
Speed	-2.839e-05	4.688e-06	1.030e+04	-6.056	1.44e-09

Table 5.8: The result of the linear regression model on evaluating changes in probability of abrupt increases in HR based on drivers' speed

Effect	Estimate	Std. Error	t-value	Pr	CI 2.5%	CI 97.5%
Intercept	1.839e-01	4.963e-03	37.058	2e-16	0.1741	0.1937
Speed	2.651e-04	5.721e-05	4.633	7.91e-06	0.0001	0.0004

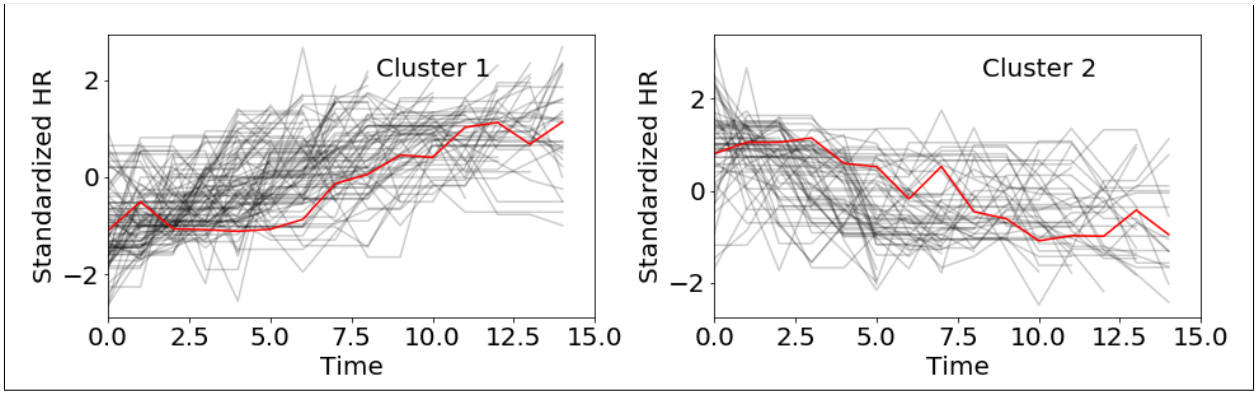


Figure 5.16: The patterns in drivers' HR in the vicinity of speed limit signs. The two clusters both have an abrupt increase in HR in the vicinity of speed limit sign

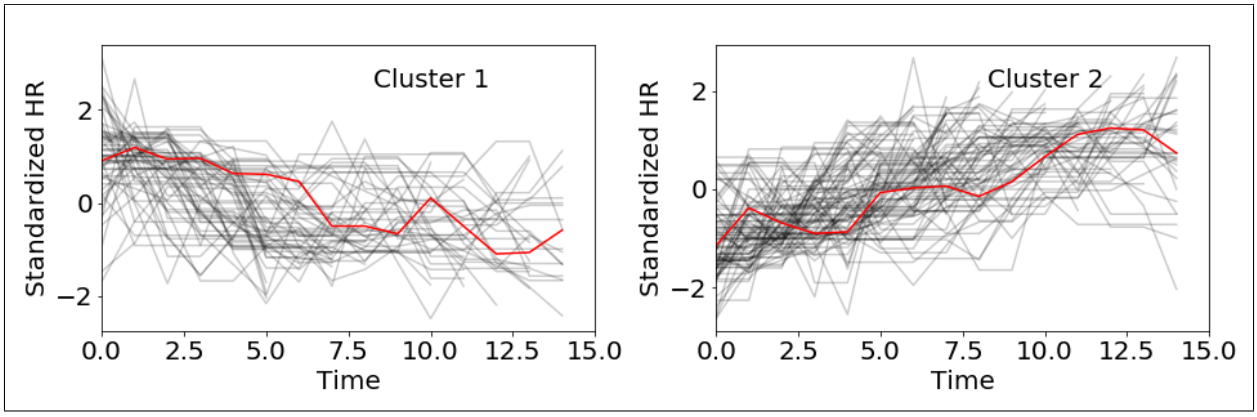


Figure 5.17: The pattern sin drivers' HR in the vicinity of speed limit signs. The two clusters both have an abrupt increase in HR in the vicinity of speed limit sign

average, drivers are less likely to show a specific movement in their facial muscles as they travel at a higher speed. The result of testing this relationship with linear regression is also shown in Table 5.9

Figure 5.23 shows the relationship between drivers' valence at varying levels of speed. As shown, drivers' valence has a more positive value at higher speeds. In other words, drivers show more positive facial expressions when they drive at higher speeds on highways. This result is also tested by using a linear regression model as shown in Table 5.10.

Table 5.9: The result of the linear regression model on evaluating changes in drivers' facial engagement based on drivers' speed

Effect	Estimate	Std. Error	t-value	Pr	CI 2.5%	CI 97.5%
Intercept	10.694699	0.328603	32.55	2e-16	10.0452	11.3441
Speed	-0.040836	0.003788	-10.78	2e-16	-0.0483	-0.0333

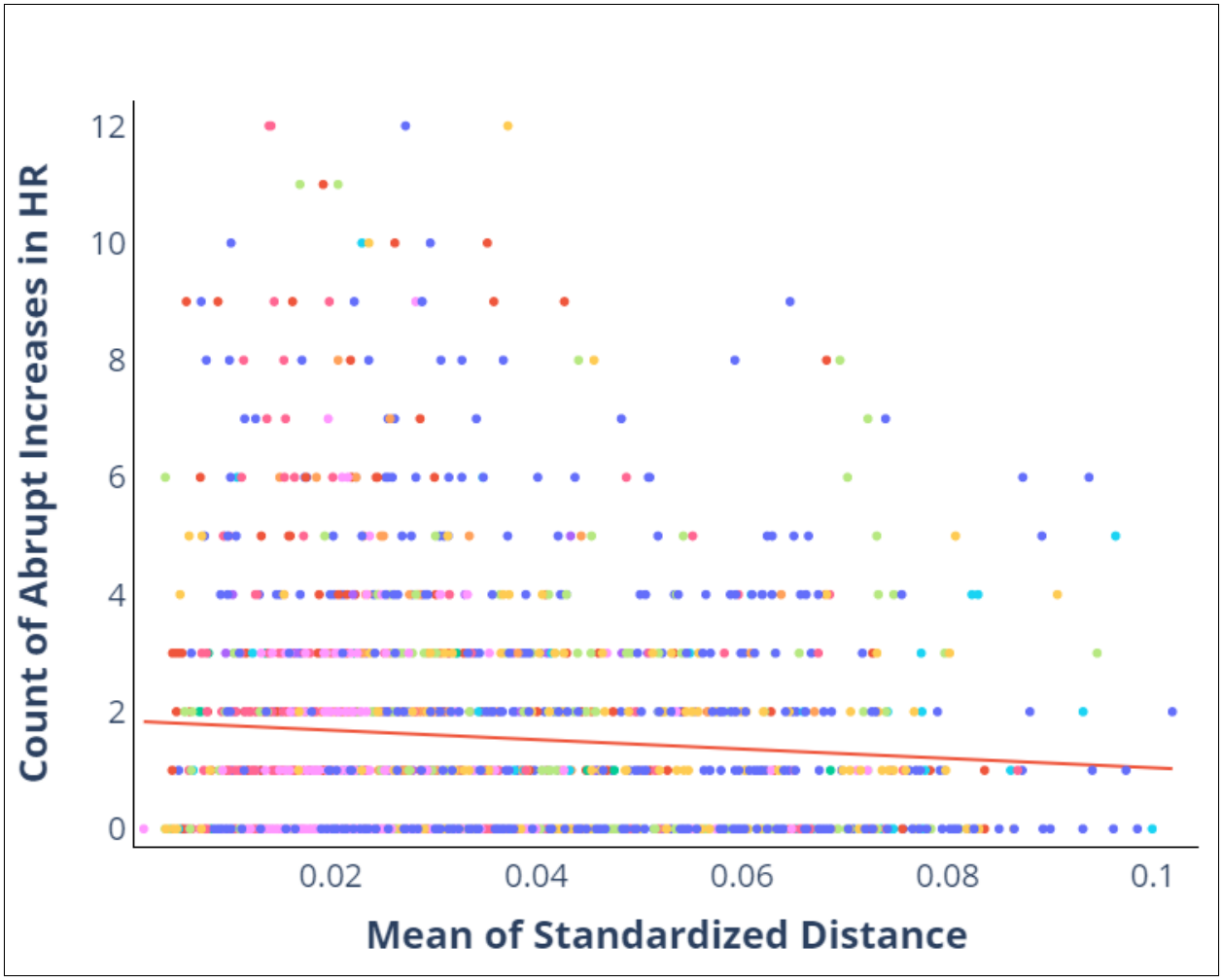


Figure 5.18: The change in the average distance to the lead vehicle versus the number of abrupt increases in drivers' heart rate. The markers show the datapoints for each participant, and the line plot is the overall trendline.

5.2.5 Discussion

In this study, we undertook an explorational approach to the relationship between drivers' HR, facial expressions, and driving context, such as the presence of certain road objects and distance to lead vehicles. As a summary, our results suggest significant relationships between changes in drivers' psychophysiological measures in the vicinity of the aforementioned road object categories. Our results, while collected materialistically and objectively, are in line with previous studies that were performed through controlled experiments and by using subjective measures such as self-reports.

Previous research showed that certain road object categories were accompanied by very high subjective stress levels. For instance, (15) showed that riders and big vehicles were accompanied by a high fraction of high stress levels. Additionally, studies such as (22) note

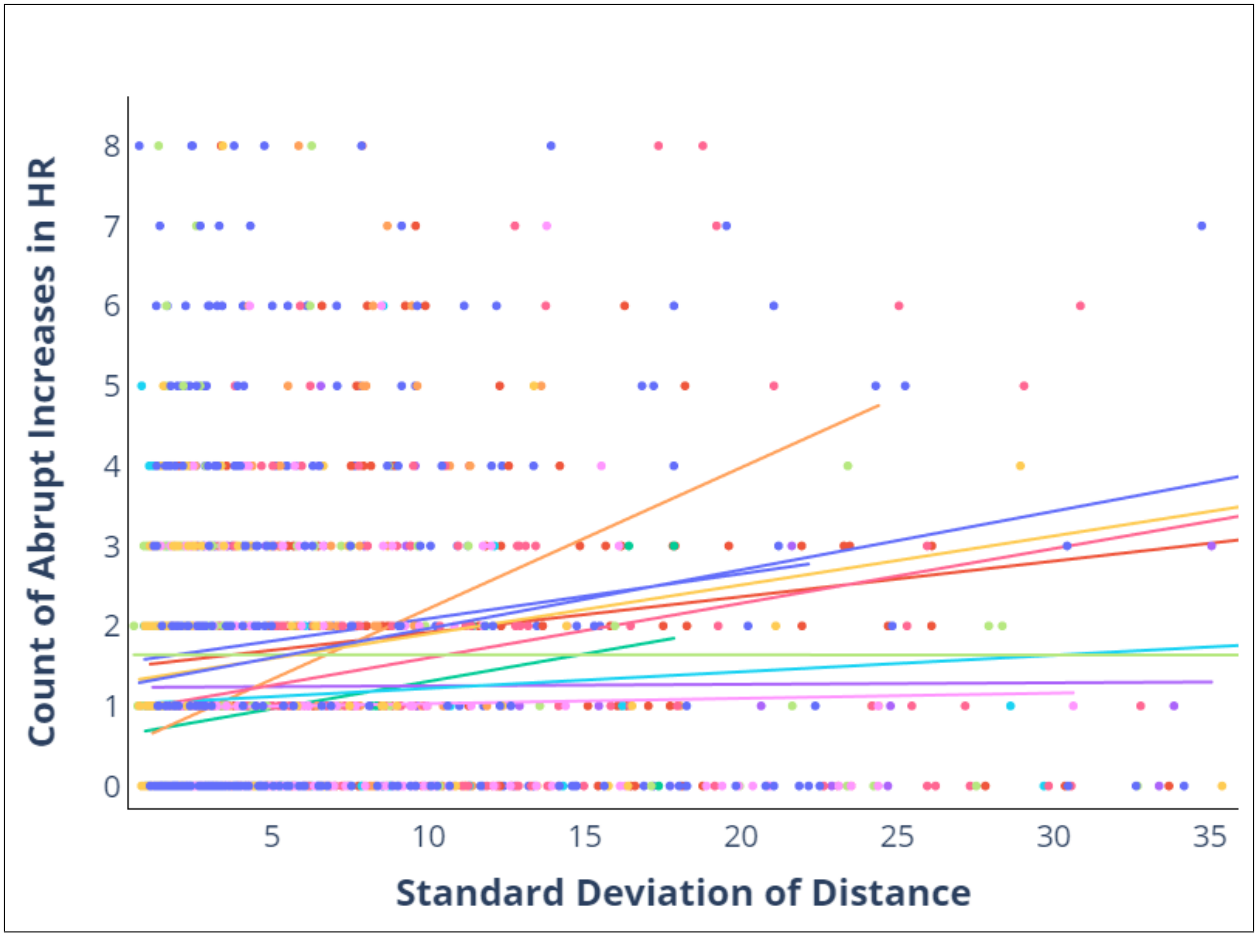


Figure 5.19: The change in the standard deviation of distance to the lead vehicle versus the number of abrupt increases in drivers' heart rate. The markers show the datapoints for each participant, and the line plots are the best fitted line for each participant's data.

that intersections are accompanied by very high subjective negative emotions. Our results show similar patterns while focusing on drivers' HR instead of subjective measures. Based on the changes in drivers' HR in the vicinity of such road objects, bigger vehicles and pedestrians are among the highest increase in HR, followed by traffic lights, traffic signs, and riders.

Moreover, changes might fall into different categories of increases in HR. For example, trucks and, in general, bigger vehicles are associated with a higher proportion of the increase in HR with more than two standard deviations away from the mean (high stress), whereas riders are often associated with a medium increase in HR (not more than two standard deviation increase in HR on figure 5.10) These results have strong implication for designing human-centered autonomous systems that may need to reason and decide for choosing routes (e.g., a route with a higher number of trucks such as highways versus a route with higher riders), as well as following or passing a road object (e.g., bicycle versus truck). Moreover, intuitively we observe that the effect of different stressors varies across participants. For

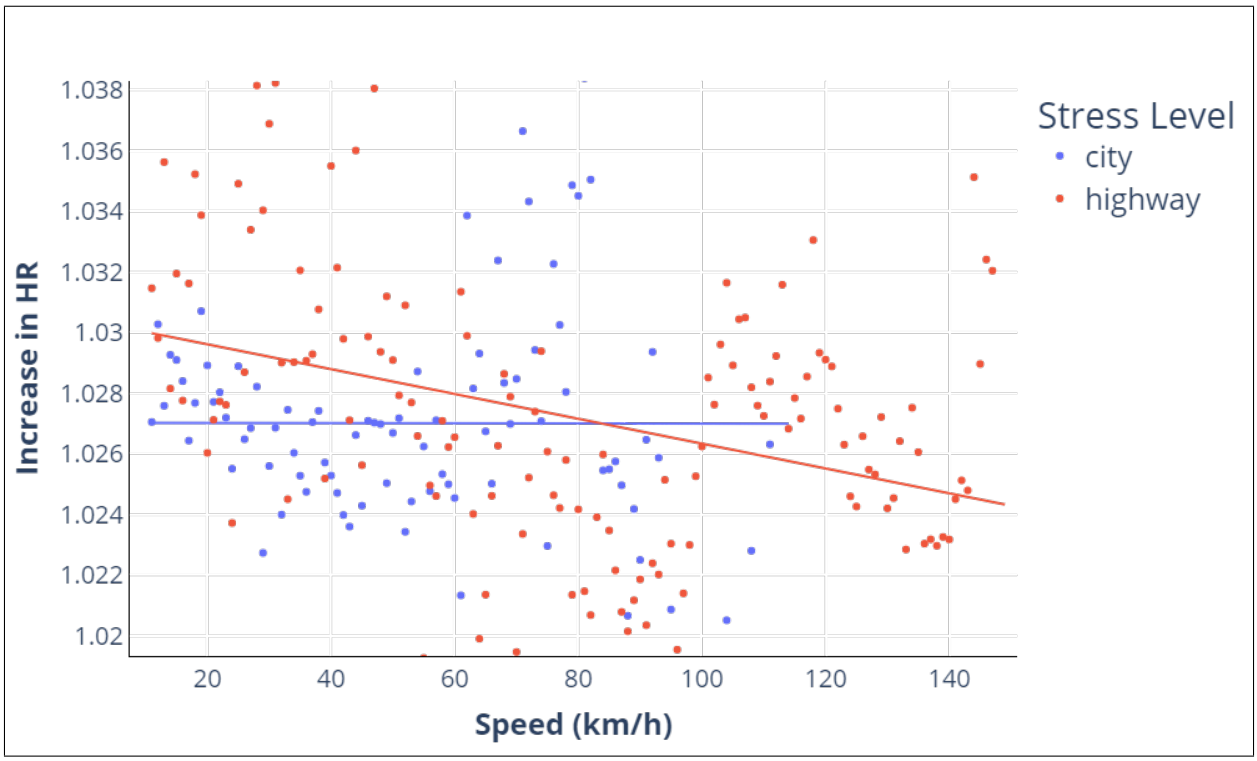


Figure 5.20: The change in heart rate at the locations of changepoints with respect to varying levels of speed for city and highway environments.

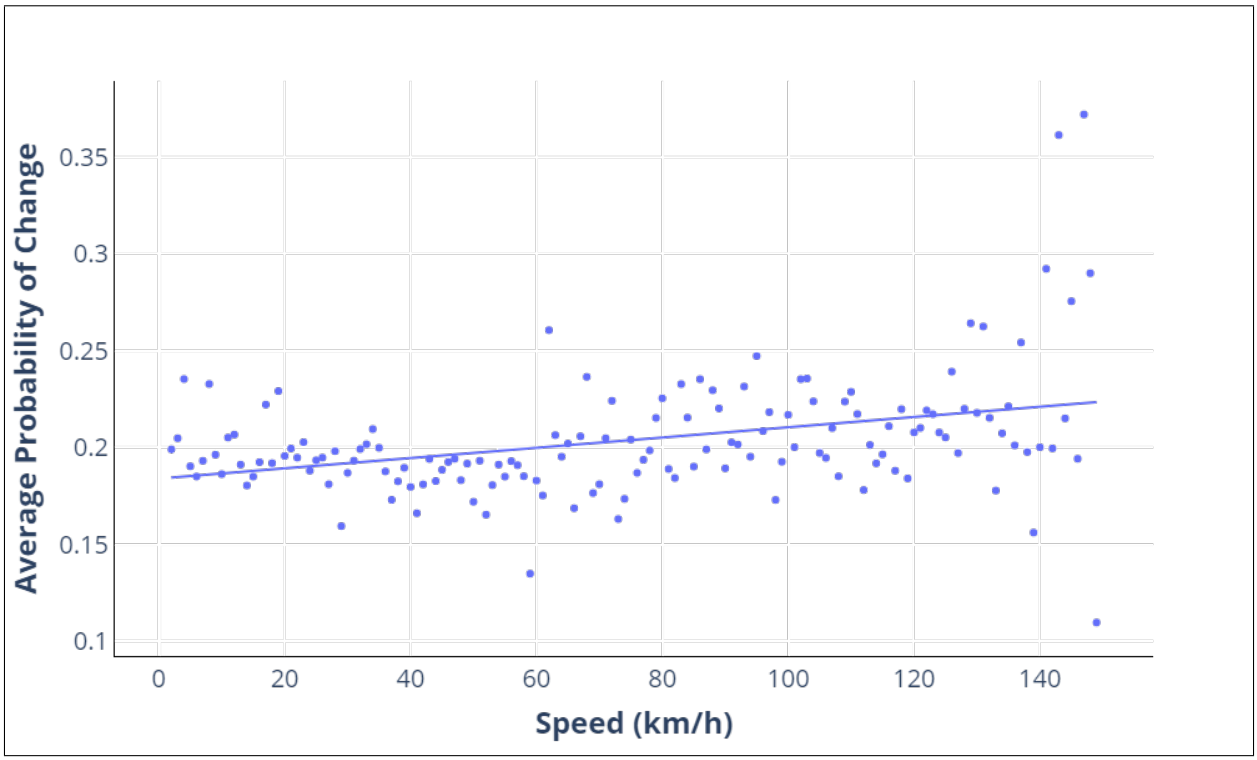


Figure 5.21: The probability of increasing change in HR at different speeds

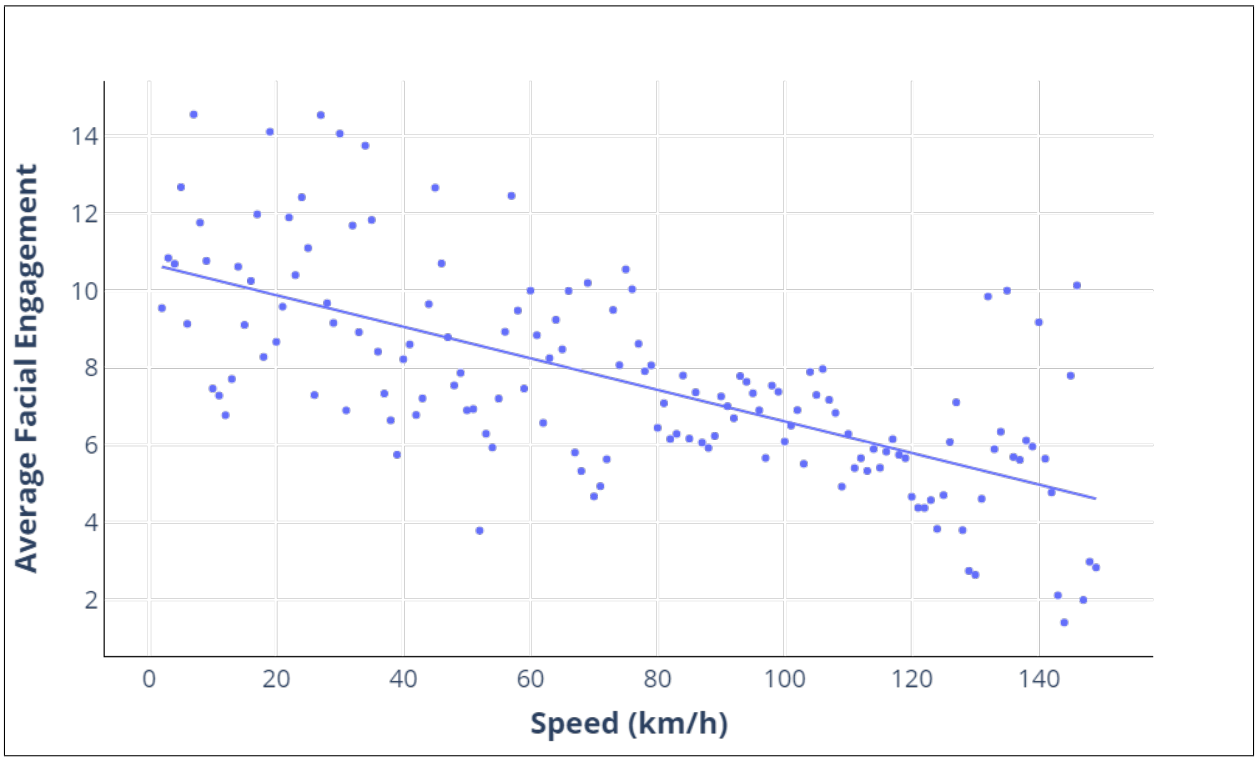


Figure 5.22: The facial engagement with respect to varying levels of speed

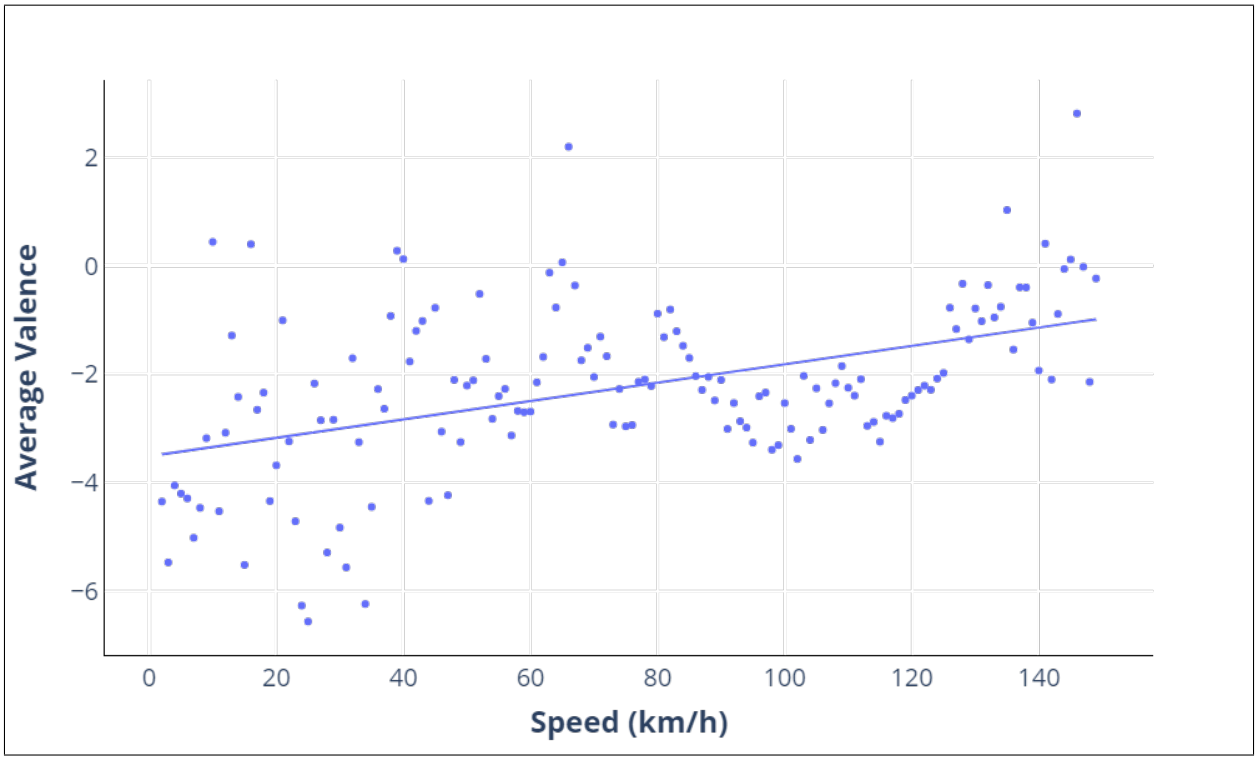


Figure 5.23: The facial valence with respect to varying levels of speed

Table 5.10: The result of the linear regression model on evaluating the relationship between in drivers' valence on drivers' speed

Effect	Estimate	Std. Error	t-value	Pr	CI 2.5%	CI 97.5%
Intercept	0.016972	0.002657	6.388	2.11e-9	-3.9709	-3.0600
Speed	-0.040836	0.003788	-10.78	2e-16	0.0117	0.0222

example, we observe the difference between participants #12 and #17 on how they respond to bigger vehicles based on their HR (figure 5.12). This has implications for designing personalized systems that can respond to each individual based on their specific profile in each specific context.

Additionally, we observe similar patterns within the engagement and valence of the drivers, where trucks and buses are among the top categories when comparing the negative emotion proportions, which is followed by traffic signs, pedestrians, riders, and traffic lights. While for some categories of road objects, the result of cardiac measures and facial expressions confirm each other (e.g., a high fraction of high-stress HR and negative valence for bigger vehicles), this is not necessarily the case for some of the other categories. For example, while traffic lights and, in general, intersections were previously shown to be associated with higher stress levels, and we observe a higher fraction of HR increases beyond two standard deviations from the mean; they exhibit the lowest fraction of high facial engagement as well as the lowest proportion of negative valence as shown on figures 5.13. Results as such might indicate that not always participants show facial expressions when they experience increases in HR. This implies the importance of multimodal sensing for human emotion and stress detection.

Analysis of patterns in drivers' HR in the vicinity of traffic signs shows that even within signs, there can be similarities and differences in how drivers perceive each sign. For instance, we observe that within each of the two traffic signs, two patterns can be detected where in both of them, drivers' HR has an abrupt increase in the vicinity of them. Additionally, we observe that even similar clusters, when comparing stop signs and speed limit signs, might have a different prolonging effect on drivers' physiology. Similar analysis should be performed for other road objects to better understand the pattern in HR in the vicinity of these objects (e.g., the difference between different types of trucks).

Our results show that a shorter distance to the lead vehicle is associated with a higher number of abrupt increases in drivers' HR. Additionally, we showed that a higher standard deviation of the distance with the following lead vehicle is associated with higher levels of abrupt increases in HR. Previous research showed a positive correlation between increases in HR and stress levels. Taken the above together, our result may indicate that being closer

to the lead vehicle as well as changing the distance to abrupt values away from the mean may be accompanied by higher stress levels and unhealthy states for the driver as well as the passenger. These results are in line with very recent research showing that within a driving simulator, shorter time headways are associated with higher workloads, which might affect drivers' safety (87). The fact that not only the average distance but also the standard deviation of distance has an impact on drivers' HR highlights the importance of connected vehicles where the distance can be calibrated and kept constant based on different user profiles.

We also observed that different participants might be affected differently by each of the road objects as well as different distances to the lead vehicle. The current scheme of designing in-cabin systems often ignores the individual differences across people. Our results warrant more human-centered considerations with an individual profile approach for designing in-cabin systems that can understand how each user might be affected by environmental attributes.

Our results show that higher speed, especially in highway environments, is correlated with lower levels of increase in drivers' HR. Previous research provided evidence on the fact that highways might be associated with lower levels of subjective stress levels (22; 107). This might indicate that reason behind perceiving highways as a less stressful environment might lie in the fact that drivers are often allowed to drive faster in the highway environment. Note that the negative correlation between speed and stress level (as measured by an increase in HR) is not as strong in the city environment. This might indicate that although in the city environment, higher speed might contribute to lowering the stress level, the presence of other stressors (e.g., traffic lights, pedestrians, riders, stop signs, etc) might compensate for the increase in speed. Additionally, similar results were found in the recent work of (74) where higher speed was more correlated with lower standard deviation in HR. Moreover, we showed that higher speeds are correlated with a higher probability of change in HR. In other words, at higher speeds, there is a higher chance that drivers' HR faces an abrupt increase. Taken the two together, we can conclude that at a higher speed, the probability of change is higher, but such a change only has a minor effect on the HR.

Our results indicate the importance of different road objects and driving context as a whole in keeping a less stressful driving experience. In the context of future vehicles, our results have implications for designing autonomous vehicles that can take actions and decisions that are a better fit for a passenger. For instance, in the case of keeping the optimal distance to the lead vehicle, drivers' HR can be used as an indicator for stress level detection at different distances. In the context of smart and connected urban environments, our results highlight the importance of the design and implementation of alternative scenarios that are

capable of replacing the current urban design scheme. For instance, different alternative designs of intersections and different alternatives to stop signs (e.g., roundabouts) can be analyzed from the perspective of how much they affect the drivers. Lastly, in the context of routing, our results have implications for designing human-centered routing systems that can provide options based on users' predicted feelings within each route with respect to different objects and road characteristics that might be present within different roads (108).

5.2.6 Limitations

This study has a number of limitations that should be addressed by the future work. First and foremost, the number of participants can be increased. Increasing the poll of participants not only helps with finding individual differences to a greater extent, but it also lays the ground to find differences across age groups, different genders, and also socioeconomic backgrounds in their reaction to each road environment attribute.

Additionally, we note that this work lies heavily on the off-the-shelf computer vision algorithms and models that were developed on certain datasets that may not necessarily represent the proper set of road stressors. For example, it can be the case that participants had different reactions to different types of trucks (e.g., trailers versus regular trucks). Part of the future work should focus on developing an object detection model that can extract such features from the massive amount of collected naturalistic videos.

While previous studies have used automatic facial expression recognition software, it is not a general consensus that such software applications can truly extract all the facial expressions. This is especially more experienced with pre-recorded in-the-wild videos where the angle of the camera, lighting, and other camera-related characteristics can change as the driving happens. Items as such can degrade the accuracy of the facial expression software and also affect the result. While we leveraged data cleaning methods to remove the unusable frames from the facial expression detection point of view, in addition to making sure all of our videos were in daylight condition, future work should investigate the effect of such matters in greater detail.

While we used computer vision to detect the distance to the lead vehicle, it is intuitive that not always the pixel wise distance represents the true distance to the vehicle. For example, within different road curves, as well as uphill and downhill driving, the distance can be very different than the visual distance in the camera. Future work should integrate other sensing modalities and a higher level of computer vision integration to detect the distance to lead vehicles with higher accuracy.

Within different modalities, there can be a lag in how each modality responds to a stimulus. For instance, it can be the case that HR increase happens faster than changes in facial expressions. In one of our previous works, we have seen possible lags between different

latent constructs (i.e., cognitive load and stress) (105). While in this paper we analyzed the different modalities, we note that a limitation is to consider the possible between modalities for a more proper comparison. Additionally, the response to different environmental stressors might also be different from a time scale point of view. It can be the case that there is a delay in response to the presence of a truck in comparison to a traffic light. Part of the future work should be focused on understanding the lag in responding to each environmental perturbation between different sensing modalities.

5.2.7 Conclusion

This research takes an exploratory approach to understand the effect of road environment objects on drivers' psychophysiological metrics. In contrast to previous studies, this research aims to understand the reason behind changes in drivers' emotions and feeling, which can later be used for interventional proposes. By analyzing naturalistic driving data from 15 participants within varying city and highway environments, we find that different road objects might be associated with different levels of increase in drivers' HR as well as different proportions of negative facial emotions. Our results indicate that bigger vehicles are associated with the highest amount of increase in drivers' HR as well as negative emotions. Additionally, we showed that shorter distances to the lead vehicle in naturalistic driving as well as the higher standard deviation in the distance might be associated with a higher number of abrupt increases in drivers' HR, showing a possible increase in stress level. Lastly, our results indicate more positive emotions, more facial engagement, and a lower number of abrupt changes in HR at a higher speed of driving.

5.3 Project 3: Driver State Modeling Under External Contextual Perturbations Using Latent Variable State Space Modeling Framework

5.3.1 Introduction

Understanding driver's state, including stress level, emotions, and cognitive load, is one of the important factors to improve driver-vehicle interactions and to enhance driver safety, comfort, and experience (15; 111; 51; 23). Recent studies are pointing towards the effect of drivers' mental state on drivers' performance both in manual and automated driving (3). For instance, drivers' stress levels and negative emotions were shown to be associated with higher accident rates (21). Similarly, cognitive load due to distraction such as secondary task engagements has an adverse effect on drivers' take-over ability in automated driving and is associated with higher accident rates (82; 21). In the definition of the National Highway Traffic Safety Administration (NHTSA) of automation (from level 1, driver assistance, to

5, fully automation), both levels 2 and 3 require driver take-over control in immediate and with notice conditions, respectively. A safe vehicle at levels 2 and 3 is reliant on drivers' attention to take over in a timely fashion, which is also reliant on drivers' state. Thus, a proper take-over requires appropriate detection, modeling, and analysis of drivers' states in different driving scenarios.

Driver's states, including emotions, cognitive load, and stress level, are psychological constructs that cannot be measured directly and are often observed through a set of measurement variables such as psychophysiological records of the driver, including driver's heart rate, gaze variability, brain signals, and skin conductance. One method to model psychological constructs such as stress level, workload, and emotions is to use latent variable modeling schemes. A latent variable framework models one or multiple unobserved constructs through a set of sensory measurements (46). Each sensor measurement might reveal a part of the real construct with a certain level of error. For example, the stress level might be modeled through a set of physiological measures such as heart rate (HR), skin temperature, and skin conductance through latent variable modeling (55). Similarly, cognitive load can be measured through driver's gaze variability, brain activity, and heart rate measurements (64; 26).

Additionally, considering that psychological constructs in driving often happen simultaneously, it is thus required to analyze the driver's state in a holistic fashion where different constructs exist simultaneously and can interact with each other. For example, real-world driving, which can be accompanied by stress-inducing driving events, often happens together with cognitive tasks such as the task of driving itself as well as the secondary tasks (64; 110). The real-world analysis of driving situations often has many, if not all of the psychological constructs interacting with each other in a dynamic fashion. In a hypothetical driving scenario, a driver might become frustrated by a lead vehicle's sudden stop; this might increase the driver's stress level due to fear of hitting the lead vehicle, while the driver might also be working with a phone or mind wandering. A review on the interaction between affect and driving behaviors also points out that it is required for driving behavior research to integrate affect, cognition, and behaviors in one framework for a more holistic understanding of the interactions among these elements (50). Latent variable modeling can thus become helpful in modeling not just the psychological constructs of interest but also their possible interaction over time.

While latent variable modeling of psychological constructs helps with realistic modeling of human's state, modeling driver's state should also address the problem of the time dependency of driver state measurements (94). Traditionally, when modeling time-dependent sensory measurements such as human HR, modeling techniques such as autoregressive mov-

ing average are used where the dependency of time points can be modeled explicitly. The problem of time dependency is also present when analyzing latent constructs where the time dependency of latent variables needs to be evaluated in the modeling scheme. For instance, when modeling stress level as a latent variable based on physiological sensors, it is important to understand the changes in stress level during the driving scenario and to what degree the stress level at each time point is related to the next. This can be achieved through state-space modeling (46). In other words, state-space modeling can be imagined as a substitute for autoregressive models in a latent variable framework.

Lastly, driving happens in a multidimensional *contextual setting*. In short *context* can be defined as any information that helps to define a situation in driving that is either internal or external to the driver. External context includes sensor measurements from the environment (e.g., weather condition, temperature, and traffic density), as well as the vehicle (e.g., current speed). The internal context consists of the measurements related to the driver (e.g., stress, valence, and arousal levels) (111). Different parts of the context can interact and affect each other. Recent studies in psychology are emphasizing the importance of context in human state analysis in which emotional episodes are tied to situation-specific needs of humans (83; 41; 8). In other words, while analyzing latent human constructs, it is important to account for the effect of external contextual inputs explicitly to achieve a more realistic model. To achieve these aforementioned modeling schemes, we propose using state-space latent variable models for driver state analysis, where the effect of external context can be measured explicitly on the latent constructs and in a temporal fashion.

The rest of the paper is as follows. We first provide detailed background literature on driver's psychophysiological states (i.e., workload and stress levels) and how they affect driver sensing data (e.g., HR), as well as their interaction with the external context (e.g., traffic density). We then propose a framework for analyzing driver's psychophysiological measures together with the external context longitudinally through taking advantage of latent variable state-space models. We showcase a sample exploratory analysis of the state-space approach on real-world driving data. We build state-space models to analyze the relationship between contextual elements (i.e., number of road users on the road and driver's task demands) and a driver's stress and workload as latent variables estimated through driver's facial expressions, eye gaze patterns, and physiological responses. Through our case study, we compare two path diagram models in a state-space modeling scheme by considering (1) one latent variable for drivers' psychophysiological state, (2) two separate latent variables for driver's psychophysiological state with an interaction with each other. We assess the log-likelihood of the aforementioned models as a measure of model fitness and discuss the comparison. We then analyze the time dependency of the aforementioned latent variables.

This research takes a holistic approach to driver state analysis where multiple internal and external factors are unified through one modeling framework temporally. Through our case study, we specifically answer the following questions:

1. How do the external context (i.e., traffic density) and task demands (based on hand-movement) affect the driver's internal contextual state (i.e., stress level and workload)?

2. How does the latent state at previous timestamps (i.e., 1 through 10 seconds) of the driver affect his/her current state (i.e., stress and workload)?

3. How does the association between latent constructs (stress level and workload) change throughout the driving scenario?

5.3.2 Background Study

The background section is divided into three subsection of context (5.3.2.1) , stress level (5.3.2.2), and workload (5.3.2.3) as follows.

5.3.2.1 Context

Depending on the field of study, the context has been defined with different elements. In computing systems, (96) defined it as surrounding objects, their locations, and their variation over time. A similar study, defined aspects of context to be *who you are*, *who you are with*, and *what resources are nearby* (95). Based on this study, context is not just the location but also noise level, lighting, social situation, etc. Later (20) defined context as any relevant information that can help characterize the situation of the interaction between an entity, an application, and the environment. Two keywords in this definition are *any relevant information* and the *interaction*. This study points out that the relevant information is not always location or objects, but it is defined based on the situation that an entity is in. More importantly, it defines certain types of context to be location, identity, activity, and time as primary. Secondary context can be derived from the primary ones. For instance, a person's identity can define their home address. Another study has defined context in three layers of adaptive layer, being responsible for retrieving the context from sensors, management layer being responsible for providing this information to other devices, and application layer being the part that uses underlying information for different applications (42). The current study considers five types of context: time, location, device, user, and network. It also points out that contextual information can be physical or logical, in which the physical part refers to raw sensors readings (e.g., body temperature), and logical context is their abstract meanings (e.g., a person having a fever).

The field of psychology provides a more human-centered definition for context. For instance, (42) points out the importance of considering emotions and preferences when defining

context. However, it is recognizable that humans' emotions and actions by themselves are meaningful when considered in the external context that they are happening (38). This points towards separating context into two categories, internal and external, where internal context is related to the human, while external context is concerned with the environment. Other studies have also provided definitions for context in driving, mostly centered around the environment, which is categorized as part of the previously mentioned definitions (external context). In other words, environment and context seem to be interchangeably used in such studies. This includes road conditions, weather conditions, and anything that happens in the environment. (97) defines context as road scene and segments. (76) defines context as a combination of environment and time. (113) defines context as the presence of turning maneuvers, a lead vehicle, and oncoming vehicles.

In this paper, we define context similar to (20; 31) as a collective summary of any relevant information about a driving situation retrieved from sensor measurements. However, by using the psychological theories definition for context, we note that context can be internal or external to the human driver. Internal contextual cues are retrieved from the human driver and can affect the driving situation (e.g., driver's heartbeat demonstrating calm or stressful state in a lane change action), whereas external contextual cues are related to the environmental elements affecting the driving situation (e.g., rainy road condition on a lane change situation).

5.3.2.2 Stress

Stress is the process in which the demand of a particular situation is perceived to be more than the available resources (33). The perceived demand can be defined based on the overall situation, including the previous experiences, internal body sensations, and the external stimuli (33). Stress can be in different time scales where short-time is referred to as acute versus long term, referred to as chronic (33).

Decreasing drivers' stress is of high importance as it can contribute to human error making and possible accidents (56). Studies in the past have pointed out the effect of stress level on drivers' performance and driving behavior (66; 35). For instance, (35) showed that perceived stress might affect behaviors such as aggressive driving and drunk driving. Multiple biomarkers have been used in literature for detecting stress (6; 17). Studies have pointed out the utility of cardiovascular measures in detecting the human state. Overall, cardiovascular activity can be measured through two technologies of Electrocardiography (ECG) and photoplethysmogram (PPG). ECG measures the electrical activity of the heart through the usage of contact electrodes, whereas PPG records the same activity through measuring blood volume in the vein using infrared technology (64; 111; 108). Devices such as conventional wearable technologies often use PPG. The electrical signal of the heart can

then be used to measure the HR (i.e., beeps per minute) and to calculate the heart rate variability (HRV) features. HRV is generally referred to as a set of features that are retrieved from the sequence of beat-to-beat intervals of an individual's HR signal. These features can be calculated in time (e.g., root mean squared of successive intervals (RMSSD)), frequency (e.g., low-frequency power (LF)), and non-linear domains (e.g., sample entropy). There is a vast literature showing that such features are, in fact, correlated with human stress levels (55). An increase in stress level is shown to be correlated with a decrease in RMSSD, an increase in LF, and an increase in HR (55). Additionally, studies pointed out the significance of individual differences in the relationship between subject stress level and HR (102).

In addition to HR, certain facial action units were used to infer stress levels. For instance, (34) reported that AU1, AU6, AU12, AU15 are the most indicative of stress levels in their study. However, when comparing dependent and independent subject methods when performing automatic facial emotion detection, the same study reported that the accuracy of stress detection using AUs dropped from 91 % to 74 % (34). Such results may show that individual differences exist in how people react in different situations. In other words, if the situation defines how a person might react, mere analysis of biological responses may not provide the true human state at each time point. These findings may suggest that driver's state analysis through psychophysiological metrics should take into account the specific situation and consider individual differences.

5.3.2.3 workload

Driver's workload is mostly defined as cognitive resources that are taken from the driver by any activity other than the driving (26), although some studies have also examined the workload from the driving itself (71). The workload includes both "mind wandering" and the load imposed by "secondary task". Cognitive workload has been shown to affect driving performance metrics. Engstrom et al. point out that cognitive workload can selectively impair driving tasks that rely on cognitive control as opposed to automated tasks (26). Studies have shown that cognitive workload might impair driver's object detection response (45), especially for the objects that are novel or difficult to detect (26). Additionally, studies show that drivers' decision-making is also negatively affected by cognitive workload (26; 19). This highlights the importance of detecting and possibly mitigating drivers' cognitive workload.

Multiple biological signals such as driver's eye metrics, cardiovascular measures, and brain signals have been used extensively for workload estimation in both in-lab and real-world situations (64; 104; 106). Driver's eye metrics in these studies included patterns of blinking rates, saccades, fixations, stationary and transition entropy (58; 29; 100; 101). Here we direct our focus on a more recent feature of the driver's gaze, which is the driver's gaze entropy

metric. There are two measures of entropy for a random variable. Information Entropy refers to the uncertainty associated with a choice (101). The entropy increases with the higher levels of uncertainty in randomness in a system. This is generally calculated through Shannon's equation (99). In gaze analysis, this entropy refers to the overall predictability of fixation locations which is a measure of gaze dispersion (99), and is called Stationary Gaze Entropy (SGE). For a set of fixation locations in a sequence of eye movements, one can assign fixation locations to spatial bins of p_i and calculate the SGE as:

$$SGE = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{5.12}$$

SGE is used extensively in the literature for human state analysis. For instance, studies have pointed out the utility of SGE for detecting task demand, complexity, experience, workload, drowsiness, and being under the influence of alcohol (101). Because inferences based on SGE can be very task-related, studies have moved more towards a new measure of entropy, namely conditional entropy. In other words, when assessing cognitive load during a task using SGE, it is important to know whether the task requires high or low SGE for optimum performance.

Conditional entropy takes into account the dependency between different fixations in a temporal fashion. This results in Gaze Transition Entropy (GTE). GTE is a measure of predictability of the next fixation location given the current location. For a sequence of transitions between different spatial bins of i and j, with a probability of p_{ij} the GTE is calculated as:

$$GTE = -\sum_{i=1}^{n} p_i \sum_{j=1}^{n} p_{ij} \log_2 p_{ij} \tag{5.13}$$

Driver's GTE is shown to be correlated with higher task demand, higher scene complexity, and higher levels of workload (101; 29). Additionally, a recent theory suggests that conceptually for each specific combination of task demand and scene complexity, there exists an optimal GTE (101). The optimal GTE is the result of an interaction between a human's internal state (e.g., memory) and the level of external information provided to the human through sensory inputs through the prediction process of the outside world. Deviating with an increased level of GTE (relative to the optimum) can be due to stress, anxiety, and emotional episodes, while a decrease in the level of GTE (relative to optimum) can be due to usage of depressants such as alcohol.

Workload detection has also been done based on facial action units. For instance, (119) reported the top four correlated action units with workload detection are AU1, AU2, AU07, AU25. However, similar to stress level detection, individual differences in workload esti-

mation were shown to be important. For instance, (119) showed that when performing subject-independent tests for workload estimation using facial AUs as compared to subject dependent, the accuracy dropped from 95 % to 68 %, an indication of inter-individual differences in facial reactions to workload.

While multiple studies have advanced our knowledge on workload or stress separately in driving, not that many studies considered analyzing both constructs simultaneously. Studies in the past have provided evidence that certain human physiological measures may be impacted differently in the case of having both stress and cognitive load present in the situation. For example, a recent review (100) reports that GTE can increase from its optimal value for a specified task if states such as anxiety and stress are present while performing a task. Thus it is important to analyze drivers' cognitive load and stress level simultaneously and in a unified framework.

5.3.3 Methodology

In this section we outline details regarding the mathematics behind state space models (5.3.3.1), data collection (5.3.3.2), feature extraction (5.3.3.3), analysis environment (5.3.3.4), and model selection (5.3.3.5).

5.3.3.1 A Latent Variable State-Space Model for Driver's State

One method to analyze driver's state is to consider the system of driver, vehicle, and the environment as a dynamical system in a state-space fashion where different perturbations from the environment change driver's state momentarily (75). A state-space model (SSM) is a mathematical representation for a dynamical system consisting of a set of inputs (referred to as perturbations) and outputs which are properties of the system that evolve over time and are measurements of certain latent variables that cannot be measured on their own (63). Using the observed variables, the latent state is estimated with a certain error, and the task of SSM is to provide the latent estimation as well as the effect of perturbations on those latent states (63). This is especially important for psychological constructs that are often measured through a set of observed variables, such as detecting stress (latent variable) through changes in physiological measures such as heart rate (observed variable).

Another important benefit of using state-space models is their natural solution to the problem of time-dependent variables. First-order State-space models (SSM) analyze the system in a recursive manner where each time point is modeled based on the previous time point in a one-step Markov process fashion (46). For the purpose of this article, we only consider first order (lag 1) state-space models. In this way, state-space models handle the time dependency of observations that is often the problem when using high-frequency sensor measurements (e.g., HR data). SSMs are suitable for analyzing driver's state since (1) driver's

state is a latent variable (e.g., stress, workload, and emotion) measured through observed variables (e.g., driver's gaze, heart rate, and skin temperature); and (2) it is time-dependent in that events in the past can affect how a driver might feel and act in the future.

We provide a summary of the procedure behind SSM. The reader is referred to (46) and (63) for greater details. Based on the notation provided in (46), the general equations for a state-space model consist of two main equations of state equation 5.14, and measurement equation 5.15, as follows:

$$x_t = Ax_{t-1} + BU_t + q_t \tag{5.14}$$

$$y_t = CX_t + DU_t + r_t \tag{5.15}$$

where x is the vector of latent variables at different time points, U is the vector of perturbations to the system (i.e., inputs), q_t is the vector of dynamic noise with covariance of Q, r_t is the vector of observation noise with zero mean and covariance of R, and y_t is the observations. Additionally, A is the matrix defining the autoregressive components across time for latent variables, B is the matrix measuring the effect of perturbation on the latent variables, C is factor loading of each latent variable based on the observed variable, and D is the matrix measuring the effect of perturbations on the observed variables. A schematic graph of this SSM representation is presented on Fig. 5.24.

As mentioned before, SSM analyzes the data in a recursive manner. SSM uses Kalman filter, which alternates prediction and correction steps as follows (46). In summary, SSM works in two steps: first, it predicts the latent variable using initial values and the state model. Second, it uses the measurement model and observation variables to update the prediction. Using a latent variable matrix at each timestep, together with its covariance matrix (P), SSM predicts the latent variables at the next timestep as:

$$x_{t|t-1} = Ax_{t-1|t-1} + Bu_t \tag{5.16}$$

$$P_{t|t-1} = AP_{t-1|t-1}A^T + Q \tag{5.17}$$

The prediction is then updated using the observed variables and the measurement model. In more detail:

$$\hat{y}_t = Cx_{t|t-1} + Du_t \tag{5.18}$$

The error is then calculated as:

$$\tilde{y} = y_t - \hat{y}_t \tag{5.19}$$

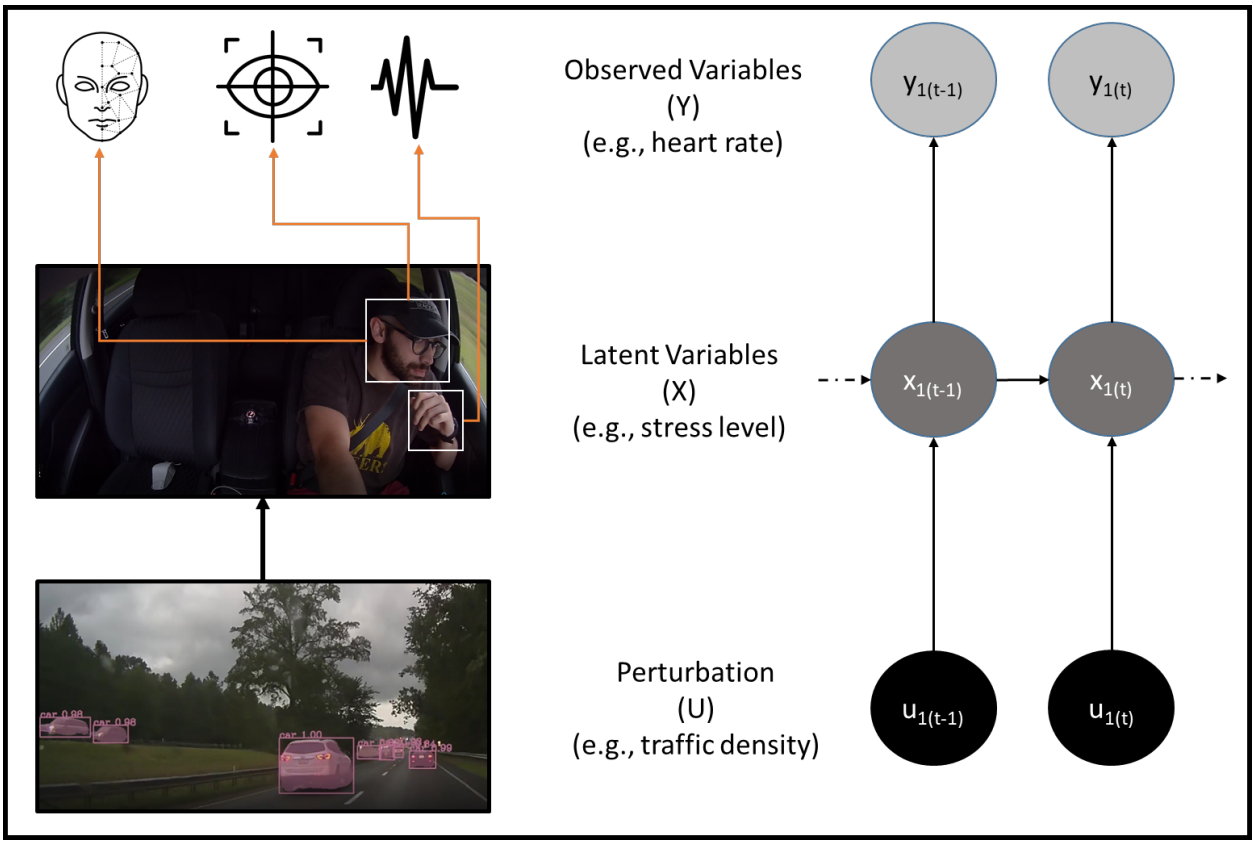


Figure 5.24: A conceptual framework of a state space model consisting of perturbations, latent and observed variables at subsequent timesteps of (t-1) and (t).

$$\hat{S}_t = CP_{t|t-1}C^T + R \tag{5.20}$$

$$K = P_{t|t-1}C^T\hat{S}_t^{-1} \tag{5.21}$$

$$x_{t|t} = x_{t|t-1} + k\tilde{y}_t \tag{5.22}$$

$$P_{t|t} = P_{t|t-1} - KCP_{t|t-1} \tag{5.23}$$

And lastly, the log likelihood (LL) of an SSM model is calculated as below (46):

$$-2LL = nlog(2\pi) + log|\hat{S}_t| + (y_t - \hat{y}_t)^T\hat{S}_t^{-1}(y_t - \hat{y}_t) \tag{5.24}$$

where n is the number of observation variables.

5.3.3.2 Data Collection

The data for this study is provided by HARMONY, a human-centered multimodal driving study in the wild (111). HARMONY is a framework that collects naturalistic longitudinal driving data through cameras, smart wearables, and multiple APIs. Through HARMONY, the driver's HR, hand movements (i.e., IMU sensors), facial expressions, gaze direction, pose direction, vehicle's speed, and location, as well as outdoor environmental videos, are collected automatically. We first focus on a long-term driving of one participant (#9) on a highway for the purpose of this paper. Also, the data from this participant is available online for further research (115). A snapshot of the driving scenario is depicted on Fig. 5.25 (109).

We then extend our findings across participants to assess the individual differences. To do so, we chose a random subset of the Harmony data from 10 other participants that drove in a long-term driving scenario (more than 1.5 hours) on a highway that was visually similar to that of the first participant. Note that because of the naturalistic nature of the data (which is not an on-road controlled study), participants drive on different roads. However, we only chose driving scenarios in highways of the state of Virginia to increase the similarity of driving scenarios across participants. More specifically, the data for this paper for each participant includes:

- **Smartwatch**: one channel of PPG signal sampled at 100 Hz, HR (HR) sampled at 1 Hz, and driver's hand acceleration and rotational velocity in three directions (i.e., X, Y, and Z) sampled at 100 Hz.
- **Camera**: in-cabin and outside videos sampled at 30 Hz, with 1080p quality.

5.3.3.3 Feature Extraction

5.3.3.3.1 Video

Using in-cabin videos, we retrieve the driver's gaze direction and facial AUs through applying OpenFace (7) on the videos. OpenFace provides gaze angles in both X (horizontal), and Y (vertical) directions. These values are measured in radian, independent of the participant's head location.

Gaze Transition Entropy: For this variable, based on method in (57), we first construct a 2D space of the range of the gaze angles in the study duration (i.e., 2-hour driving period). Then we divide the space into equally distanced areas of interest (AOI) as a 4*4 grid (100). This results in a sequence of AOIs for the driver, which is related to different areas of the frontal view (e.g., front road, center stack, left side, and etc.). We then use the method provided in (57) to calculate gaze transition entropy (GTE) for each time window in the driving scenario. In summary, for a sequence of AOIs, we first find the transition matrix

between areas of interest by assuming a first-order Markov process for the gaze sequences (57). To this end, a transition matrix is retrieved with p_{ij} being the probability of switching between AOIs i and j (in S), with a stationary probability of π_i. Then the GTE is calculated based on Shannon's entropy as:

$$\hat{H}_t = -\sum_{i \in S} \pi_i \sum_{j \in S} p_{ij} \log_2 p_{ij} \tag{5.25}$$

Facial Action Units: We limit the facial AUs in our models based on previous literature to the ones that were shown to be correlated with stress levels and workload. As mentioned in background literature, (34) reported that AU1, AU6, AU12, AU15 are the most indicative of stress levels in their study. (119) reported the top four correlated action units with workload detection to be AU1, AU2, AU07, AU25. Note that any other AU can also be used for the purpose of stress and workload estimation. We intentionally keep the number of AUs limited for a better model interpretation.

Road Object Detection: By using the outside videos and by applying Mask-RCNN (40) algorithm trained on COCO dataset (61) as implemented by (2), we retrieve the number of cars, buses, pedestrians, bikes, motorcycles, stop signs, traffic lights, and trucks in each frame of the video. Here we define the scene complexity as the total number of road users in the field of view, which is the sum of cars, buses, pedestrians, motorcycles, and trucks (Fig. 5.25). Note that due to the relatively short length of dataset per person, we have not considered the effect of momentary perturbations such as traffic lights, stop signs, etc. This will be addressed in future work with a longer dataset.

5.3.3.3.2 Smartwatch

Heart rate: Previous studies have shown the correlation between a human's heart rate and stress levels as described in section 5.3.2.2, in which higher heart rate values might be indicative of higher stress levels. Additionally, our previous studies showed that certain stressors on the road could increase a driver's HR from its baseline value, in which change points in HR time series can be used to detect such stressors (111; 108; 109). While HR values were sampled at 1 Hz, the exact frequency of sampling often changes within 0.9-1.1 Hz due to hardware issues. To address this, we resampled the HR values at 1 Hz frequency. We also apply Bayesian Change Point (BCP) detection to the driver's HR. Without going into the details of this method, BCP detects the specific moments that a change occurs in the underlying distribution of the data based on the Bayesian change point model provided in Barry and Hartigan's book (9). In summary, this model assumes that the mean of the input within different segments remains constant. Change points in driver's HR might be

Figure 5.25: A snapshot of the data used in this study including the devices (camera and smart wearable) (A), the road view (B), and the map of the case study trip (C)

correlated with stressful outside events where the posterior mean of the HR increases as a response to the stimulus(111). This is in line with previous studies that higher HR values might be associated with higher stress levels. To apply this method, we use the BCP package in R (27). The details of the BCP procedure are also provided in our former article (111).

IMU: In order to find out the moments that the driver's hand had abrupt movements, depicting activities such as working with a phone, we use the hand IMU sensor and find the magnitude of each of the gyroscope and accelerometer sensors. We then only consider the values above the average value of each sensor as timepoints when movements are abnormal.

Each row of the final data frame after feature extraction includes driver's HR and its features (i.e., probability of a change and the posterior mean detected by BCP), driver's gaze entropy, the intensity of specific facial AUs (a number between 1-5), traffic density (i.e., number of road users in the field of view) as a measure of scene complexity, and driver's hand movement as a proxy for driver's non-driving related (secondary) tasks. Driver's HR, facial AUs, and gaze features are the observation variables for the internal context latent variables (i.e., stress level and workload), while scene complexity and task demands are the

external contextual inputs to the system.

5.3.3.4 Analysis Environment

We use the MARSS package (44) developed in R programming language for estimating different SSM matrices of A, B, C, and D as described in section 5.3.3.1. We propose two different models, and we compare these models to our data. The comparison is performed using the log-likelihood of each model, where a higher value shows a better fit to the data (44). Finally, we discuss the implications of the best fit model. The analysis is performed with a 10 seconds timestep look back, meaning that each timestep is compared with 10 seconds in the past. While we could also compare every two consecutive data points, increasing the timesteps helps with a better interpretation of the model. Also, previous research in driver behavior analysis, especially in the vicinity of a crash or near-crash event, generally considers a time point of 6-20 seconds prior to the crash event for prevalence analysis of different factors (e.g., driver emotion and distractions) (21). Also, smaller numbers increase the high dependency between timepoints, which makes the model interpretation more difficult. As the timepoints were too close in time, this step lets us understand the effect of perturbations in greater detail by removing the high dependency among the data points that are close in time. Note that this step does not remove any data point and only compares each timepoint at (t) with the data point at t+10 by restructuring the data.

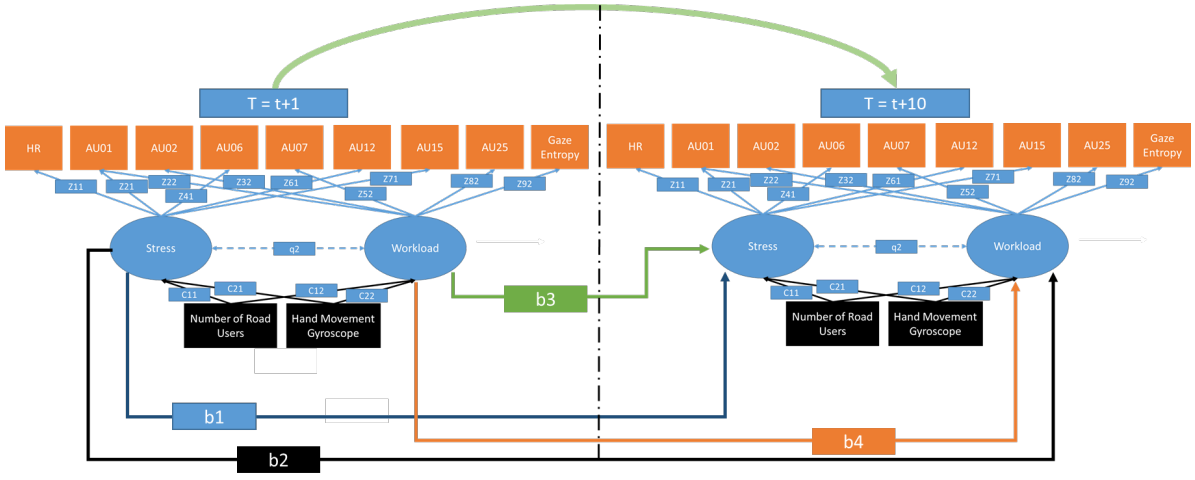


Figure 5.26: The two latent variable model. Note that in contrast with this model, in the base model, the two latent variables are identical. In this model a set of perturbations (i.e., number of road users, and hand movement shown with black) is connected to a set of latent variables (i.e., stress level and work load shown with blue) estimated through measurement variables (e.g., HR, AUs, and gaze entropy shown with orange

5.3.3.5 Model Selection

In order to answer the research questions, we define two separate models: base model and two latent variables models. In each model, a set of perturbations (i.e., number of road users, and hand movement shown with black on Fig. 5.26) is connected to a set of latent variables (i.e., stress level and work load shown with blue on Fig. 5.26) estimated through measurement variables (e.g., HR, AUs, and gaze entropy shown with orange on Fig. 5.26).

Base Model This model considers one latent variable referred to as the driver's psychophysiological state. In other words, this model assumes that one latent variable is enough to describe the multimodal data regarding a driver's internal state under the effect of the external context. In order to be able to compare this model with a model with two latent variables, we need to account for the differences in the degrees of freedom. When building this model, we assume two identical latent variables. This helps with accounting for the changes in the degrees of the freedom of the model as compared to the two separate latent variables model, which is described below.

Two Latent Variables Model: The second model includes two separate latent variables for stress and workload of the driver, namely internal context. These latent variables are measured through driver's HR, facial action units, and eye gaze measurement features. This model assumes a covariance between the two latent variables. Thus a covariance matrix between the latent variables is also estimated in the model. More specifically, as shown in Fig. 5.26, the q2 variable between stress and workload is estimated through the state space approach. The model is depicted in Fig. 5.26 for different timesteps. Note that the state-space model takes a recursive attempt at estimating the different matrices as described in section 5.3.3.1.

5.3.4 Case Study and Results

In order to choose between the two models, we compare using their -2 $*$ Log Likelihood (-2LL). Table 5.11 shows the -2LL of each model fitted for 11 different participants. As shown, the -2LL of the model with the two latent variable models is considerably lower as compared to the base models. This suggests that the model with interacting two latent variables can better describe the variability in the data for each individual participant. This comparison led us to choose the interacting latent variable model for the rest of our analysis.

Let us focus on participant #9 for a better description of state-space results (Table 5.12)). We then extend our findings to other participants. The latent stress is captured through the higher HR, AU1, AU6, AU15, and lower probability in AU12. The workload is captured through the higher intensity of the AU1, AU2, and lower intensity in AU7, as well as higher gaze entropy. The association between the two latent variables show that higher workload is

Table 5.11: A comparison between the two models based on their -2LL across different participants. The model with two latent variables has a lower -2LL

Participant ID	Interacting Latent Variable Model	Base Model	ΔLL
2	170459.8	172879	2419.2
3	299674	305882.4	6208.4
9	272412	276434.4	4022.4
10	101985	102780.7	795.7
12	187456.8	189293.8	1837
14	120738.6	122577.1	1838.5
16	93854.2	94914.72	1060.52
17	153230.4	155310.3	2079.9
18	167789.8	170418.8	2629
19	83357.6	84523.7	1166.1
20	178416.4	179560.6	1144.2
22	208460	210691.6	2231.6

associated with lower stress level in this participant, which the coefficient is ($Q.q1 = -0.068$ on Table 5.12).

The association between the two latent variables varies across participants (Fig. 5.27). Note that these numbers are the result of dividing the association coefficient by the product of the standard deviation of each latent variable. This is required for the correct comparison of the association coefficient across participants. The association coefficient varies between both positive and negative numbers across participants. This can imply that there are two groups of participants. In the first group, an increase in participants' stress level is accompanied by a decrease in workload, while in the second group, the two latent variables change in the same direction. Such high variability across participants shows the importance of individual profiles when considering matters such as distraction, stress levels, and driver state monitoring within different contexts.

Research Question 1: How do the external context (i.e., traffic density) and task demands (based on hand-movement) affect the driver's internal contextual state (i.e., stress level and workload)?

Participant #9: Let us first start with one participant to illustrate the results. The state-space model for participant 9 shows that the number of road users can be associated with an increase in stress levels and a decrease in workload (compare C.C11 with C.C12 on Table 5.12). Due to being stuck in higher traffic density with the increase in the number of road users, the mean value of the driver's heart rate increases, which in turn increases

Table 5.12: Estimates of the interacting latent variable model with their confidence intervals for participant #9. We specifically focus on one participant to illustrate the results of state space. The parameters shown here are based on the Fig. 5.26

Parameter	ML.Est	Std Error	low CI	up CI
Z11	1.76e+00	0.57898	0.62989	2.899433
Z21	2.26e-01	0.08057	0.06782	0.383666
Z41	-2.50e-02	0.03421	-0.09203	0.042064
Z61	8.47e-02	0.04596	-0.00538	0.174772
Z71	2.00e-01	0.07329	0.05608	0.343378
Z22	-1.87e-05	0.01258	-0.02467	0.024630
Z32	2.62e-02	0.01502	-0.00324	0.055640
Z52	-1.37e-01	0.04654	-0.22820	-0.045754
Z82	-1.21e-01	0.04164	-0.20307	-0.039844
Z92	1.03e+00	0.33772	0.36807	1.691904
B.b1	8.81e-01	0.00716	0.86746	0.895518
B.b2	8.91e-03	0.00905	-0.00884	0.026650
B.b3	-1.58e-03	0.00271	-0.00689	0.003724
B.b4	9.76e-01	0.00244	0.97149	0.981065
Q.q2	-6.08e-03	0.00422	-0.01435	0.002193
C.C11	9.94e-03	0.00398	0.00215	0.017732
C.C21	-6.90e-03	0.00326	-0.01328	-0.000517
C.C12	-7.34e-03	0.00330	-0.01381	-0.000870
C.C22	-7.53e-03	0.00339	-0.01418	-0.000877

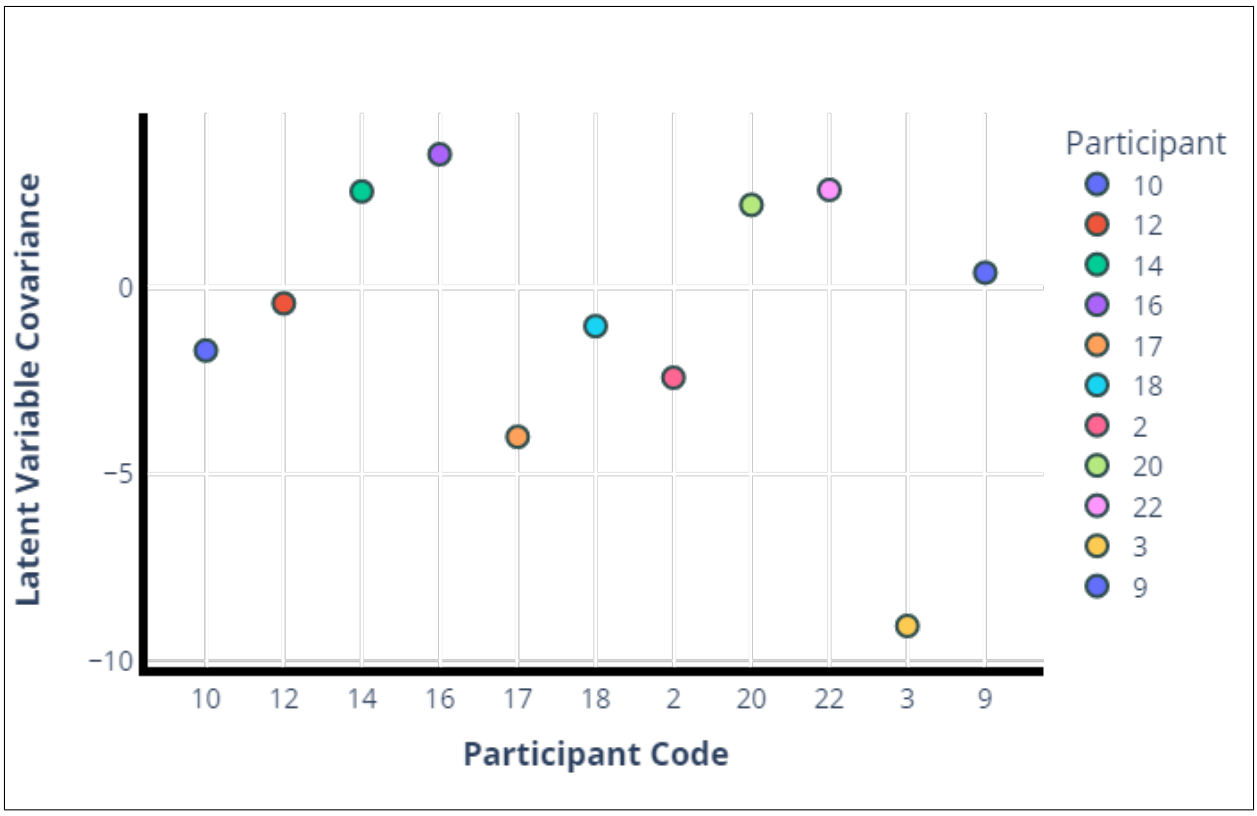


Figure 5.27: The association between the two latent variables of stress and workload across different participants. Note the high individual variability between participants. For some participants the association is positive leading to a synchronized increase in stress level and workload, while for others the two constructs change in different directions.

the stress level. Previous research has shown that traffic density negatively affects a driver's well-being (77).

Additionally, our model shows that an increase in the number of road users is accompanied by a lower workload (see Table 5.12). When considering the workload, we have measured the gaze entropy through different regions in the field of view. This method of analysis considers the whole secondary task area (i.e., using the vehicle radio, phone, etc.) as one region. Thus it cannot analyze a person's finer gaze patterns when working with a phone. This is important to consider when analyzing the workload under different environmental perturbations. An explanation for this can be due to the fact that higher traffic density often is accompanied by a stationary state for the vehicle, which motivates the participant to perform secondary tasks more often (118). This decreases the gaze dispersion with respect to the driving scene and diverges it to the secondary task area (e.g., phone and center stack), which decreases the workload as measured with respect to driving and increases the workload spent on the secondary task. The videos accompanying our analysis were reviewed to confirm

this finding qualitatively. Hand movement is associated with a decrease in stress level as well as workload. Note that higher hand movements result in performing secondary tasks in which take the attention from normal driving and bring it into one specific region (i.e., phone and center stack location in the vehicle) (compare C.C21 with C.C22 on Table 5.12).

Comparison Across Participants: We then consider comparing different participants with respect to the effect of the external context on their stress level and workload. Fig. 5.28 - A and B depict the effect of number of road users and task demands (i.e., C values on Fig. 5.26) on driver's stress and workload respectively. As shown, participants' stress level and work load are associated with different impacts by the number of road users and task demands. While this difference can certainly be due to the different contextual elements that were present in each participant's driving scenario (e.g., weather condition and distance to other vehicles), which we did not account for, it can also be due to individual differences in how they react the number of road users and task demands. In some of the participants, an increase in the number of road users is associated with an increase in their stress level (e.g., participant number 9, 10). For some of the other participants, the increase in the number of road users is associated with a decrease in their stress level (e.g., participants number 16 and 18). There are also some participants that their stress level are not affected by the increase in number of road users (e.g., participant number 3). Simultaneously, the number of road users is associated with an increase in the workload for some participants (e.g., participant numbers 17, and 18) while not for the other participants. Similar results can also be observed in the effect of task demands when comparing different participants (Fig. 5.28 - B).

While confirming the reason behind the possible individual differences is not possible in a naturalistic study without controlled experiments, one possible explanation can be that an increase in the number of road users might make it harder for some participants to drive, which increases his/her stress level. However, the other group of participants might be using different feedback loops to drain the pressure from the increase in the number of road users, such as using their phones, listening to music, or talking to a passenger. For instance, considering Fig. 5.28 - A, for participant 3, the increase in the number of road users increases their stress while decreasing the workload of driving as measured through gaze entropy. This participant might be using their phone more often in these situations, which is the reason behind the decrease in the level of workload imposed by driving and diverges their attention to the secondary tasks (e.g., phone). The future direction of our research will analyze different feedback loops in participants through differential equation modeling.

Research Question 2: How does the latent state at previous timestamps (i.e., 1 through 10 seconds) of the driver affect his/her current state (i.e., stress and

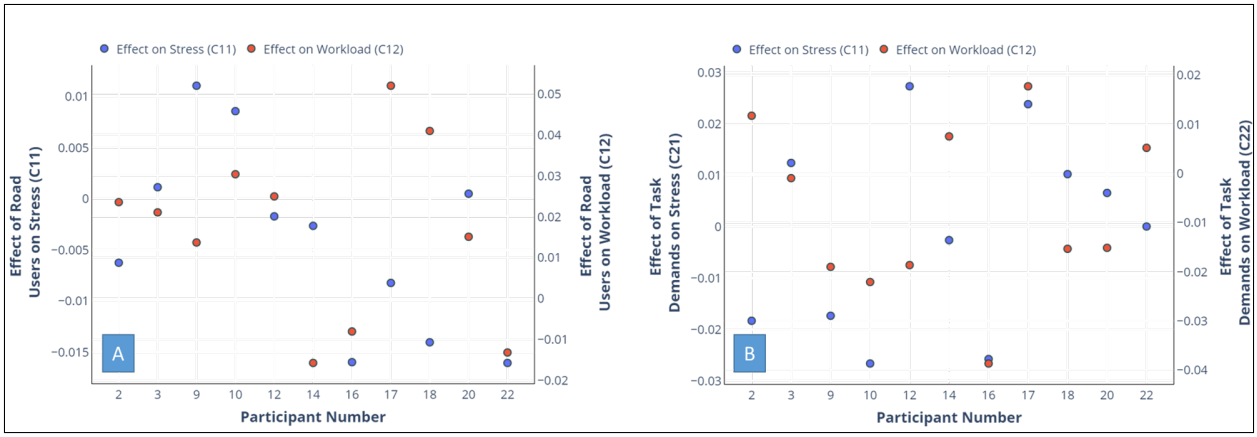


Figure 5.28: The effect of number of road users (A) and task demands (B) on stress and workload across different participants. The number on each marker shows the participant number.

workload)?

Participant #9: In order to answer this question we focus on the variables connecting the different time steps which are B.b1 and B.b4 on Table 5.12. It is interesting to observe very high temporal dependency values for drivers' stress and workload. Note that we have already applied a 10-second look ahead for the state-space model. Even with a 10-second look ahead, the model provides evidence that a driver's states (both stress level and workload) are highly dependent on its historical values, meaning that momentary changes can well affect the future state of the driver.

Comparison Across Participants: The high dependency is also observed through the data collected from other participants. Fig. 5.29 - A and B depict the stress and workload transition coefficients across participants. The high values of coefficients imply that drivers' previous states are well predictive of their future state with a very high dependency when modeled through a state-space framework.

One alternative explanation to the high temporal dependency is that the presence of rare events in driving can affect the analysis and may result in inaccurate high dependency values between the driver's states. To detect rare events, we use the results of the change point detection on drivers' HR and find the locations that HR changes abruptly. This analysis is performed based on previous research showing the effect of external context on drivers' HR using changepoint detection methods (for example, see (111; 39; 53)). To showcase the rare event analysis, let us focus on one participant, #9. Using the change points in driver's HR, we have segmented the driving scenario for participant 9 to avoid the rare events statistics. Here we define a rare event in driving as an event that is associated with an increase in

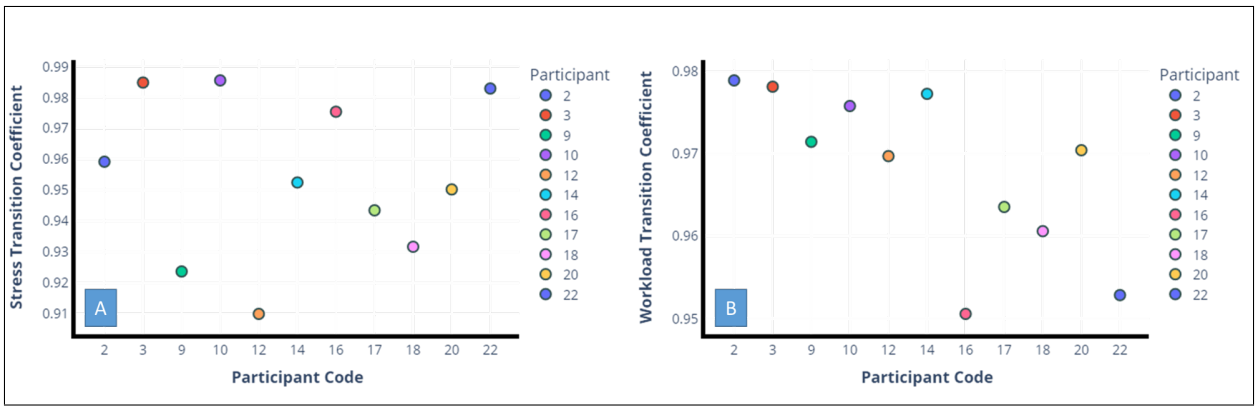


Figure 5.29: The transition coefficient of stress latent variable across participants (A), and the transition coefficient of workload latent variable across participants (B). Note the high dependency of stress latent variable on the previous values across participants

driver's HR beyond twice the standard deviation of HR in the whole driving scenario. The locations of change points associated with such rare events are shown in Fig. 5.30 with vertical red dashed lines. For each segment, we then ran the state-space model. The results of the stress and workload transition coefficients are shown in Fig. 5.30. There are two main observations drawn from Fig. 5.30. Firstly, within different segments as defined by change points, the transition coefficients are still very high (above 0.7), depicting a high dependency on the previous states. Second, the two coefficients do not move in a synchronized fashion. For instance, moving from segment 2 to 3, the workload transition stays constant while stress transition decreases. For this segment, this implies that although the dependency of workload on its historical values does not change, the stress level is becoming more and more unpredictable based upon its historical values. This confirms our finding in the model selection section that this model, in fact, captures two separate constructs through human sensing data, which do not always act in a synchronized fashion. Additionally, this finding lays the ground for possible information transmission between the two constructs as in some cases, they perform synchronized and in other cases, they do not. This brings us to our third research question.

Research Question 3: How does the association between latent constructs change throughout the driving scenario?

To further analyze the association between the two transition coefficients, we have applied a rolling window on the data collected from participant #9 and assessed a state-space model for each rolling window. More specifically, a sliding window with a length of half an hour of data was considered with an overlap of 1 minute. At each time step, half an hour of data is considered, a state-space model is assessed for the data, and then the window moves for

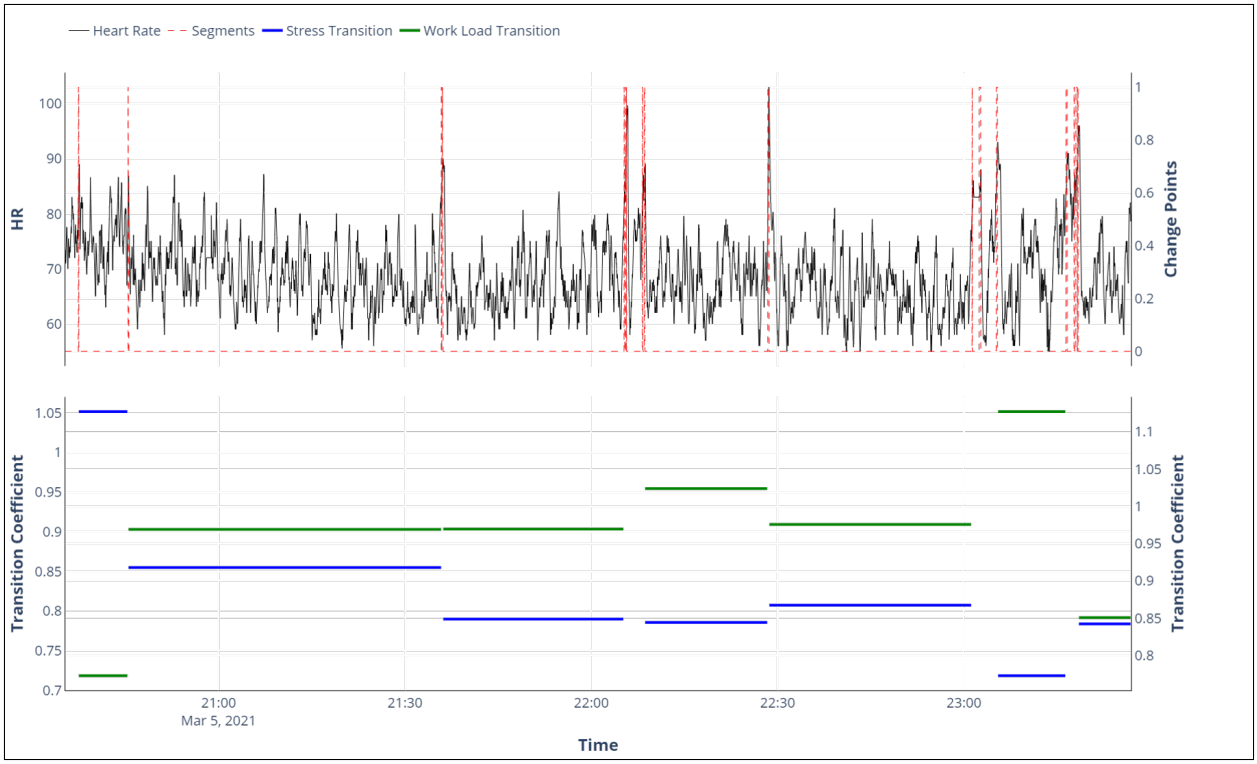


Figure 5.30: Driver's HR together with the locations of segmentation, shown with red lines. These locations are chosen based on the occurrence of a change point with an abrupt increase in HR as much as 1.5 times the standard deviation of HR in the trip.

one minute in the forward direction in time, and the analysis is repeated. We have then calculated the cross-correlation between the two signals at each rolling window. This is depicted on Fig. 5.31. The association between the two signals varies across the driving scenario duration, in which it moves between positive to negative values. This might imply that sometimes changes in the two constructs are synchronized, and sometimes they are in opposite direction.

However, the cross-correlation by itself does not show the exact direction of correlation. In other words, on Fig. 5.31 it is not possible to understand whether stress level is leading or following the workload. To further find the answer to the sequence of stress level and workload events, we calculated the windowed cross-correlation and peak picking analysis as provided by (12). In summary, in order to find out the variation in the association between two behavioral signals, one can perform a windowed cross-correlation (WCC) and peak picking analysis on the data. This analysis technique attempts to find the lag that locally maximizes the correlation between the two signals. This is performed by moving a certain window of each signal in different directions for different lags and assessing the

correlation between the two signals. Note that this is different than merely finding the lag that maximizes the correlation, as it performs locally. If the two signals are synchronized, the lag at which the correlation maximizes should always be at zero.

On the other hand, if the dynamics of maximum lag varies during the time, this can imply that the sequence of the two signals can change throughout time. In other words, in our case, it can imply that sometimes events in stress level proceed the events in workload and sometimes follows. Although this does not necessarily suggest evidence on causality, as both latent variable events can be caused by a third variable, one possible explanation in such situations can be that stress causes workload at some time points, and in the other ones, workload causes stress. Another possible explanation is that the response in stress level and workload to different inputs happen at different time scales in which sometimes the stress level responds faster and in other times the workload leads the response.

After performing the WCC analysis on the workload and stress transition coefficients, the results are plotted on Fig. 5.31 as purple dots showing the lag at which the correlation at each window maximizes. There are two main observations in this plot that helps with understanding the relationship between stress level and workload transition coefficients. Firstly, the lag between the two signals moves from positive to negative values showing changes in the sequence of the two signals of stress and workload transition. This might imply that there is a flow of information between the two constructs that move in different directions at different time points. Second, it is also visible that as the direction starts to change, the correlation between the two signals calculated through cross-correlation starts to increase.

5.3.5 Discussion

In this study, we have analyzed the interaction between the driver's state (i.e., internal context) and the environmental attributes (i.e., external context) through modeling the system of the driver and the environment using a state-space fashion. Using a state-space modeling approach, we demonstrated that a two-latent variable model could better describe the multimodal driver psychophysiological sensing data, pointing out the possibility of having multiple psychological constructs that interact with one another. Additionally, our model takes a holistic approach in analyzing the relationship between the external and internal context by analyzing the state of the driver in the environment through state-space latent variable modeling.

In our study, we observed a strong dependency of the latent variables on their values from 10 seconds prior. This has strong implications for designing autonomous vehicles where drivers' state is sought to be estimated and predicted for a safer shared autonomy. Current guidelines of National Highway Traffic Safety Administration (4), define different levels of automation from 0 to 5, where in level 3 of automation, the driver is not required to monitor

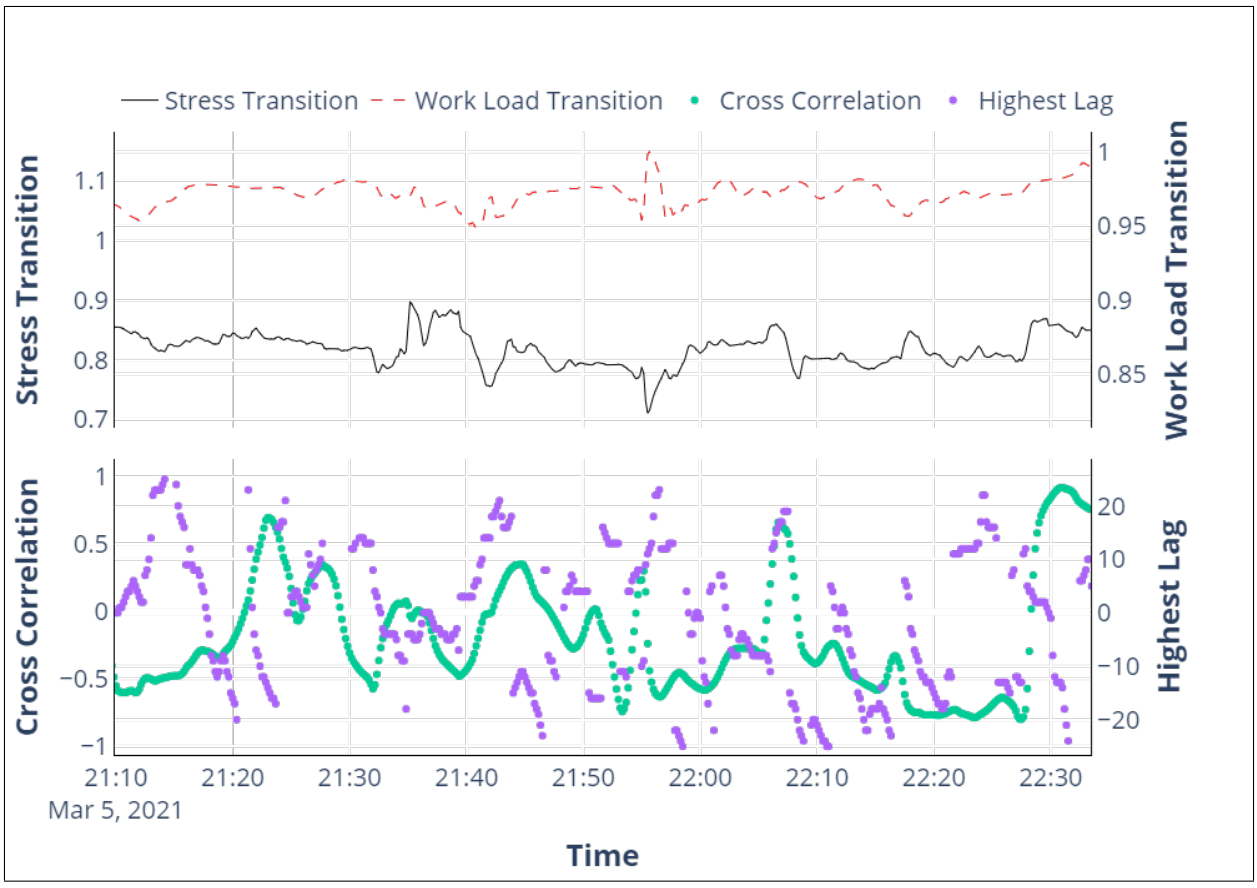


Figure 5.31: The time series of changes in driver's stress and workload transition coefficients. Note that the cross-correlation between the two signals does not always follow the same trend implying that two latent variables are, in fact, measuring two possibly different constructs. Additionally, note that the cross-correlation between the two signals of stress and workload transition coefficient varies during the trip, in which different directions of the flow of information between the two constructs can be seen through the lags across the two signals that maximizes the synchronous behavior.

the road but has to take-over the control with prior notice. Our study has implications for defining a proper notice period for take-over control while considering drivers' state. While we have focused on a 10 second look-ahead time window, more studies are required to find the dependency between drivers' state, and its historical values for a bigger population and within different contexts (e.g., city versus highway). When considering the effect of emotions on a driver's take-over control, it is important to consider the autoregressive nature of different driver state constructs the fact that prior states such as stress level are associated with the current state. Additionally, when analyzing prior naturalistic driving data for gaining insights regarding the effect of environmental factors on safety and crash-related matters, the analysis should move beyond a couple of seconds prior to events. For example, when

considering the effect of driver's emotions or secondary task engagements on the prevalence of accidents (e.g., for a sample prevalence analysis, see (21)), it is important to move beyond a couple of seconds prior to the accident, as drivers' state is highly associated to its previous values. Thus a stressful event can well propagate throughout the future timesteps.

Driving is a cognitive task that can often be accompanied by acute stress. Certain on-road (e.g., getting closer to a lead vehicle), and internal characteristics of the driver (e.g., remembering a stressful event) can all affect the driver's state simultaneously. In our study, we observe a better model fit for two latent variables as compared to one when considering driver's state measures such as HR, facial, and gaze. A two latent variable model provides a better description for the underlying variability in the driver's state measures. Considering previous studies pointing to a need for modeling schemes that integrate workload and affective states together (50), this might provide evidence for the existence of different constructs when analyzing driver state data. This implies that driver state modeling should take a holistic approach and consider multiple constructs rather than isolating each construct. Thus a naturalistic study focused on emotions in driving may need to also consider the workload and vice versa.

The association between stress level and workload varies throughout the study. While we did not analyze the reason behind variation, there seems to be a flow of information between different constructs that, in some cases, one leads the other one. Although this might not provide full evidence for a causal relationship between stress level and workload, it might imply that the response to the environment happens with different lags within the two constructs. The value and also the direction of the lag are not constant and vary in different driving situations. This implies the importance of considering context when analyzing psychological constructs in a holistic approach.

The state-space approach estimates the latent variables through different observatory variables; thus, any new measurement can also be added to the modeling scheme as they become available. This emphasizes the modular nature of the state-space approach in estimating the latent variable and the effect of the perturbations upon them. For instance, different sensors are being added to the wearable devices continuously, such as skin temperature, skin conductance, etc. Every new sensor can be added as a measurement variable. Similarly, environmental measurements such as distance to other vehicles, road complexity measures, etc., can also be added as perturbations to the system in the current modeling scheme.

Lastly, recent psychological theories and experimental findings are providing more and more evidence on the importance of individualizing profiles for driver state analysis. In other words, without considering individual differences, building generalizable frameworks

for driver state detection might easily fail when applying them in subject-independent situations. In our study, we observe that even within similar environments, the association between participants' psychological constructs can be very different across our participants. Additionally, we observe that external contextual elements such as traffic density can have different effects on different participants. In some participants, it can increase the stress level, while in other participants, it might decrease the stress level. It is important to note that higher traffic density, while being stressful for many individuals, can actually decrease the speed of travel, which might be pleasant to some other participants. This might explain the differences across participants in the coefficients retrieved through state-space models. More studies are required to analyze the relationship between the driver and the environment within varying contexts.

5.3.6 Limitations

There are currently a set of limitations that will be addressed in our future work. We will expand the number of participants to understand the individual differences across different scenarios. For instance, for our study, we have focused on one highway driving scenario for each participant, whereas other scenarios might provide additional information on the individual differences in the interaction between the two constructs. While in this study, we did not find meaningful differences across age and gender, future work will also consider the variation in the impact of the environment with respect to different ages and gender.

Additionally, we will increase the number of features to include distance to the other vehicles, lane position, and speed that are detected from the external environment. Adding more modalities might provide information on the differences across participants on the impact of environment on their stress and workload. For example, it could be the case that two drivers are facing a similar number of road users but with different distances to the lead vehicles. Additionally, this will help us better understand the interaction between the contextual elements. Using other behavioral metrics such as driver's speed, we can better analyze the driver's feedback to the elements of external context. The feedback loop will be analyzed in greater detail using second-order differential equation models where we account explicitly for a participant's resilience in different driving events.

5.3.7 Conclusion

In this research, we propose using the latent variable state-space modeling approach for analyzing the impact of the external context on drivers' states. Through applying this modeling scheme on naturalistic driving data retrieved through HARMONY, (1) we estimate driver's stress level and workload by using the data from driver's cardiovascular measures as well as gaze variability and facial expression data; (2) we estimate the effect of the number

of cars and drivers task demands as perturbations to this dynamical system on driver's stress level and workload; and (3) we analyze the temporal dependency of driver's state during a driving scenario. Our work paves the way for designing human-centered driver-vehicle interaction systems that can understand and respond to changes in the driver's state resulting from the driving environment and provide a safer driving experience.

Bibliography

[1] Abdic, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B. and Schuller, B. (2016). Driver frustration detection from audio and video in the wild, *Springer*, Vol. 9904, Springer, p. 237.

[2] Abdulla, W. (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow, `https://github.com/matterport/Mask_RCNN`.

[3] Agrawal, S. and Peeta, S. (2021). Evaluating the impacts of situational awareness and mental stress on takeover performance under conditional automation, *Transportation research part F: traffic psychology and behaviour* **83**: 210–225.

[4] *Automated vehicles for safety* (n.d.).
URL: *https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety: :text=A%20vehicle%20that%20is%20fully,controls%20for%20an%20actual%20driver.*

[5] Balali, V. and Golparvar-Fard, M. (2016). Evaluation of multiclass traffic sign detection and classification methods for us roadway asset inventory management, *Journal of Computing in Civil Engineering* **30**(2): 04015022.

[6] Balters, S., Gowda, N., Ordonez, F. and Paredes, P. E. (2021). Individualized stress detection using an unmodified car steering wheel, *Scientific reports* **11**(1): 1–11.

[7] Baltrusaitis, T., Zadeh, A., Lim, Y. C. and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, pp. 59–66.

[8] Barrett, L. F., Mesquita, B. and Gendron, M. (2011). Context in emotion perception, *Current Directions in Psychological Science* **20**(5): 286–290.

[9] Barry, D. and Hartigan, J. A. (1993). A bayesian analysis for change point problems, *Journal of the American Statistical Association* **88**(421): 309–319.

[10] Bates, D., Sarkar, D., Bates, M. D. and Matrix, L. (2007). The lme4 package, *R package version* **2**(1): 74.

[11] Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series., *KDD workshop*, Vol. 10, Seattle, WA, USA:, pp. 359–370.

[12] Boker, S. M., Rotondo, J. L., Xu, M. and King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series., *Psychological methods* **7**(3): 338.

[13] Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution, *Trends in ecology & evolution* **24**(3): 127–135.

[14] Brown, V. A. (2021). An introduction to linear mixed-effects modeling in r, *Advances in Methods and Practices in Psychological Science* **4**(1): 2515245920960351.

[15] Bustos, C., Elhaouij, N., Solé-Ribalta, A., Borge-Holthoefer, J., Lapedriza, A. and Picard, R. (2021). Predicting driver self-reported stress by analyzing the road scene, *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, pp. 1–8.

[16] Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique, *Journal of artificial intelligence research* **16**: 321–357.

[17] Chesnut, M., Harati, S., Paredes, P., Khan, Y., Foudeh, A., Kim, J., Bao, Z. and Williams, L. M. (2021). Stress markers for mental states and biotypes of depression and anxiety: A scoping review and preliminary illustrative analysis, *Chronic Stress* **5**: 24705470211000338.

[18] Chung, W.-Y., Chong, T.-W. and Lee, B.-G. (2019). Methods to detect and reduce driver stress: a review, *International journal of automotive technology* **20**(5): 1051–1063.

[19] Cooper, P. J., Zheng, Y., Richard, C., Vavrik, J., Heinrichs, B. and Siegmund, G. (2003). The impact of hands-free message reception/response on driving task performance, *Accident Analysis & Prevention* **35**(1): 23–35.

[20] Dey, A. K., Abowd, G. D. and Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications, *Human–Computer Interaction* **16**(2-4): 97–166.

[21] Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M. and Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data, *Proceedings of the National Academy of Sciences* **113**(10): 2636–2641.

[22] Dittrich, M. (2021). Why drivers feel the way they do: An on-the-road study using self-reports and geo-tagging, *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 116–125.

[23] Du, N., Yang, X. J. and Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving, *arXiv preprint arXiv:2010.03047* .

[24] *Easy OCR* (2021).
URL: *https://github.com/JaidedAI/EasyOCR*

[25] Ekman, P. and Cordaro, D. (2011). What is meant by calling emotions basic, *Emotion review* **3**(4): 364–370.

[26] Engström, J., Markkula, G., Victor, T. and Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis, *Human factors* **59**(5): 734–764.

[27] Erdman, C. and Emerson, J. W. (2007). bcp: an r package for performing a bayesian analysis of change point problems, *Journal of Statistical Software* **23**(1): 1–13.

[28] et. al., G. J. (2021). ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support.
URL: *https://doi.org/10.5281/zenodo.5563715*

[29] Fabio, R. A., Incorpora, C., Errante, A., Mohammadhasni, N., Caprì, T., Carrozza, C., De Santis, S. and Falzone, A. (2015). The influence of cognitive load and amount of stimuli on entropy through eye tracking measures., *EAPCogSci*.

[30] FakhrHosseini, S. M. and Jeon, M. (2019). How do angry drivers respond to emotional music? a comprehensive perspective on assessing emotion, *Journal on multimodal user interfaces* **13**(2): 137–150.

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

[32] Fox, J. (2002). Linear mixed models, *Appendix to an R and S-plus Companion to Applied Regression* **16**: 2349–2380.

[33] Francis, A. L. (2018). The embodied theory of stress: A constructionist perspective on the experience of stress, *Review of General Psychology* **22**(4): 398–405.

[34] Gavrilescu, M. and Vizireanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system, *Sensors* **19**(17): 3693.

[35] Ge, Y., Qu, W., Jiang, C., Du, F., Sun, X. and Zhang, K. (2014). The effect of stress and personality on dangerous driving behavior among chinese drivers, *Accident Analysis & Prevention* **73**: 34–40.

[36] Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., Marias, K. and Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos, *Biomedical Signal Processing and Control* **31**: 89–101.

[37] Gilks, W. R. (2005). Markov chain monte carlo, *Encyclopedia of biostatistics* **4**.

[38] Greenaway, K. H., Kalokerinos, E. K. and Williams, L. A. (2018). Context is everything (in emotion research), *Social and Personality Psychology Compass* **12**(6): e12393.

[39] Guo, X., Angulo, A., Robartes, E., Chen, T. D. and Heydarian, A. (2021). Orclsim: A system architecture for studying bicyclist and pedestrian physiological behavior through immersive virtual environments, *arXiv preprint arXiv:2112.03420* .

[40] He, K., Gkioxari, G., Dollár, P. and Girshick, R. B. (2017). Mask r-cnn. corr abs/1703.06870 (2017), *arXiv preprint arXiv:1703.06870* .

[41] Hoemann, K., Khan, Z., Feldman, M., Nielson, C., Devlin, M., Dy, J., Barrett, L. F., Wormwood, J. B. and Quigley, K. (2020). Context-aware experience sampling reveals the scale of variation in affective experience.

[42] Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G., Altmann, J. and Retschitzegger, W. (2003). Context-awareness on mobile devices-the hydrogen approach, *36th annual Hawaii international conference on system sciences, 2003. Proceedings of the*, IEEE, pp. 10–pp.

[43] Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scandinavian journal of statistics* pp. 65–70.

[44] Holmes, E. E., Ward, E. J. and Wills, K. (2012). Marss: Multivariate autoregressive state-space models for analyzing time-series data., *R journal* **4**(1).

[45] Horrey, W. J. and Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques, *Human factors* **48**(1): 196–205.

[46] Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment, *Structural Equation Modeling: A Multidisciplinary Journal* **25**(2): 307–324.

[47] Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues, *Annual review of psychology* **60**: 1–25.

[48] Jain, A. K., Murty, M. N. and Flynn, P. J. (1999). Data clustering: a review, *ACM computing surveys (CSUR)* **31**(3): 264–323.

[49] Jensen, M. B., Philipsen, M. P., Møgelmose, A., Moeslund, T. B. and Trivedi, M. M. (2016). Vision for looking at traffic lights: Issues, survey, and perspectives, *IEEE Transactions on Intelligent Transportation Systems* **17**(7): 1800–1815.

[50] Jeon, M. (2015). Towards affect-integrated driving behaviour research, *Theoretical Issues in Ergonomics Science* **16**(6): 553–585.

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

[52] Jocher, G., Nishimura, K., Mineeva, T. and Vilariño, R. (2020). Yolov5, *Code repository https://github. com/ultralytics/yolov5* .

[53] Kerautret, L., Dabic, S. and Navarro, J. (2022). Detecting driver stress and hazard anticipation using real-time cardiac measurement: A simulator study, *Brain and behavior* p. e2424.

[54] Kerautret, L. and Navarro, J. (2022). Detecting driver stress and hazard anticipation using real-time cardiac measurement: A simulator study, *Brain and Behavior* .

[55] Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. and Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature, *Psychiatry investigation* **15**(3): 235.

[56] Krahnen, L. S., Bauernfeind, G., Leiber, P. and Jipp, M. (2022). Evaluation of two short-term stress interventions in the context of mobility, *Transportation research part F: traffic psychology and behaviour* **84**: 155–164.

[57] Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., Vilaro, A. and Villalobos, N. (2015). Gaze transition entropy, *ACM Transactions on Applied Perception (TAP)* **13**(1): 1–20.

[58] Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C. and Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze, *PloS one* **13**(9): e0203629.

[59] Kulke, L., Feyerabend, D. and Schacht, A. (2020). A comparison of the affectiva imotions facial expression analysis software with emg for identifying facial expressions of emotion, *Frontiers in psychology* **11**: 329.

[60] Kumar, S., Datta, D., Dong, G., Cai, L., Barnes, L. and Boukhechba, M. (2021). Leveraging mobile sensing and bayesian change point analysis to monitor community-scale behavioral interventions: a case study on covid-19.

[61] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L. (2014). Microsoft coco: Common objects in context, *European conference on computer vision*, Springer, pp. 740–755.

[62] Liu, S., Koch, K., Zhou, Z., Föll, S., He, X., Menke, T., Fleisch, E. and Wortmann, F. (2021). The empathetic car: Exploring emotion inference via driver behaviour and traffic context, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **5**(3): 1–34.

[63] Lodewyckx, T., Tuerlinckx, F., Kuppens, P., Allen, N. B. and Sheeber, L. (2011). A hierarchical state space approach to affective dynamics, *Journal of mathematical psychology* **55**(1): 68–83.

[64] Lohani, M., Payne, B. R. and Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving, *Frontiers in human neuroscience* **13**: 57.

[65] Malladi, R., Kalamangalam, G. P. and Aazhang, B. (2013). Online bayesian change point detection algorithms for segmentation of epileptic activity, *2013 Asilomar Conference on Signals, Systems and Computers*, IEEE, pp. 1833–1837.

[66] Matthews, G., Dorn, L., Hoyes, T. W., Davies, D. R., Glendon, A. I. and Taylor, R. G. (1998). Driver stress and performance on a driving simulator, *Human Factors* **40**(1): 136–149.

[67] Mauss, I. B. and Robinson, M. D. (2009). Measures of emotion: A review, *Cognition and emotion* **23**(2): 209–237.

[68] McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J. and Picard, R. (2013). Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 881–888.

[69] McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J. and Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit, *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pp. 3723–3726.

[70] McKight, P. E. and Najab, J. (2010). Kruskal-wallis test, *The corsini encyclopedia of psychology* pp. 1–1.

[71] Mehler, B. and Reimer, B. (2019). How demanding is" just driving?" a cognitive workload-psychophysiological reference evaluation.

[72] Mehta, A., Sharma, C., Kanala, M., Thakur, M., Harrison, R. and Torrico, D. D. (2021). Self-reported emotions and facial expressions on consumer acceptability: A study using energy drinks, *Foods* **10**(2): 330.

[73] Mesken, J., Hagenzieker, M. P., Rothengatter, T. and de Waard, D. (2007). Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports,(observed) behaviour, and physiology, *Transportation research part F: traffic psychology and behaviour* **10**(6): 458–475.

[74] Milardo, S., Rathore, P., Amorim, M., Fugiglando, U., Santi, P. and Ratti, C. (2021). Understanding drivers' stress and interactions with vehicle systems through naturalistic data analysis, *IEEE Transactions on Intelligent Transportation Systems* .

[75] Mirman, J. H. (2019). A dynamical systems perspective on driver behavior, *Transportation research part F: traffic psychology and behaviour* **63**: 193–203.

[76] Moosavi, S., Omidvar-Tehrani, B., Craig, R. B., Nandi, A. and Ramnath, R. (2017). Characterizing driving context from driver behavior, *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1–4.

[77] Morris, E. A. and Hirsch, J. A. (2016). Does rush hour see a rush of emotions? driver mood in conditions likely to exhibit congestion, *Travel behaviour and society* **5**: 5–13.

[78] Müller, M. (2007). Dynamic time warping, *Information retrieval for music and motion* pp. 69–84.

[79] Nacpil, E. J. C., Wang, Z. and Nakano, K. (2021). Application of physiological sensors for personalization in semi-autonomous driving: A review, *IEEE Sensors Journal* .

[80] Napoletano, P. and Rossi, S. (2018). Combining heart and breathing rate for car driver stress recognition, *2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, IEEE, pp. 1–5.

[81] Nguyen, T. T. H., Jatowt, A., Coustaty, M. and Doucet, A. (2021). Survey of post-ocr processing approaches, *ACM Computing Surveys (CSUR)* **54**(6): 1–37.

[82] Odachowska, E., Ucińska, M., Kruszewski, M. and Gąsiorek, K. (2021). Psychological factors of the transfer of control in an automated vehicle, *Open Engineering* **11**(1): 419–424.

[83] Ortony, A., Clore, G. L. and Collins, A. (1990). *The cognitive structure of emotions*, Cambridge university press.

[84] Park, S. Y., Moore, D. J. and Sirkin, D. (2020). What a driver wants: User preferences in semi-autonomous vehicle decision-making, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13.

[85] *Python Tesseract* (2021).
URL: *https://pypi.org/project/pytesseract/*

[86] R Core Team (2017). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

[87] Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., Guillen, P. P. and Lenné, M. G. (2022). Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving, *Transportation Research Part F: Traffic Psychology and Behaviour* **87**: 149–163.

[88] Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R., Torrao, G., Lv, W., Puente Guillen, P. and Lenné, M. (2022). Physiological indicators of driver workload

during car-following scenarios and takeovers in highly automated driving, *Transportation Research Part F Traffic Psychology and Behaviour* **87**: 149–163.

[89] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2015). You only look once: Unified, real-time object detection. arxiv 2015, *arXiv preprint arXiv:1506.02640* .

[90] Reinares-Lara, P., Rodríguez-Fuertes, A. and Garcia-Henche, B. (2019). The cognitive dimension and the affective dimension in the patient's experience, *Frontiers in psychology* **10**: 2177.

[91] Roidl, E., Frehse, B., Oehl, M. and Höger, R. (2013). The emotional spectrum in traffic situations: Results of two online-studies, *Transportation research part F: traffic psychology and behaviour* **18**: 168–188.

[92] Russell, J. A. (1980). A circumplex model of affect., *Journal of personality and social psychology* **39**(6): 1161.

[93] Russell, J. A. (2003). Core affect and the psychological construction of emotion., *Psychological review* **110**(1): 145.

[94] Sadeghi, M., McDonald, A. D. and Sasangohar, F. (2021). Posttraumatic stress disorder hyperarousal event detection using smartwatch physiological and activity data, *arXiv preprint arXiv:2109.14743* .

[95] Schilit, B., Adams, N. and Want, R. (1994). Context-aware computing applications, *1994 First Workshop on Mobile Computing Systems and Applications*, IEEE, pp. 85–90.

[96] Schilit, B. N. and Theimer, M. M. (1994). Disseminating active map information to mobile hosts, *IEEE network* **8**(5): 22–32.

[97] Schneemann, F. and Heinemann, P. (2016). Context-based detection of pedestrian crossing intention for autonomous driving in urban environments, *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, pp. 2243–2248.

[98] Shahapure, K. R. and Nicholas, C. (2020). Cluster quality analysis using silhouette score, *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, pp. 747–748.

[99] Shannon, C. E. (1948). A mathematical theory of communication, *The Bell system technical journal* **27**(3): 379–423.

[100] Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M., Berlowitz, D. J., Swann, P. and Howard, M. E. (2018). Stationary gaze entropy predicts lane departure events in sleep-deprived drivers, *Scientific reports* **8**(1): 1–10.

[101] Shiferaw, B., Downey, L. and Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency, *Neuroscience & Biobehavioral Reviews* **96**: 353–366.

[102] Sommerfeldt, S. L., Schaefer, S. M., Brauer, M., Ryff, C. D. and Davidson, R. J. (2019). Individual differences in the association between subjective stress and heart rate are related to psychological and physical well-being, *Psychological science* **30**(7): 1016–1029.

[103] Tavakoli, A. (2022). Code for stop sign detection.
URL: *https://github.com/arashtavakoli/*

[104] Tavakoli, A., Balali, V. and Heydarian, A. (2019). A multimodal approach for monitoring driving behavior and emotions, *Technical report*, Transportation Research Board, Washington DC, United States.

[105] Tavakoli, A., Boker, S. and Heydarian, A. (2021). Driver state modeling through latent variable state space framework in the wild, *UNDER REVIEW - Transportation Research Part F* .

[106] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2020a). Personalized driver state profiles: A naturalistic data-driven study,, *11th International Conference on Applied Human Factors and Ergonomics*.

[107] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2020b). Personalized driver state profiles: A naturalistic data-driven study, *International Conference on Applied Human Factors and Ergonomics*, Springer, pp. 32–39.

[108] Tavakoli, A., Boukhechba, M. and Heydarian, A. (2021). Leveraging ubiquitous computing for empathetic routing: A naturalistic data-driven approach, *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[109] Tavakoli, A. and Heydarian, A. (2021). Multimodal driver state modeling through unsupervised learning, *arXiv preprint arXiv:2110.01727, UNDER REVIEW - Transportation Research Part F* .

[110] Tavakoli, A., Kumar, S., Boukhechba, M. and Heydarian, A. (2021). Driver state and behavior detection through smart wearables, *2021 IEEE Intelligent Vehicles Symposium (IV)*, IEEE.
URL: *https://doi.org/10.1109/iv48863.2021.9575431*

[111] Tavakoli, A., Kumar, S., Guo, X., Balali, V., Boukhechba, M. and Heydarian, A. (2021). Harmony: A human-centered multimodal driving study in the wild, *IEEE Access* **9**: 23956–23978.

[112] Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K. and Woods, E. (2020). Tslearn, a machine learning toolkit for time series data, *Journal of Machine Learning Research* **21**(118): 1–6.
URL: *http://jmlr.org/papers/v21/20-091.html*

[113] Tivesten, E. and Dozza, M. (2014). Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving, *Transportation research part F: traffic psychology and behaviour* **26**: 258–272.

[114] Tracy, J. L. and Randles, D. (2011). Four models of basic emotions: a review of ekman and cordaro, izard, levenson, and panksepp and watt, *Emotion review* **3**(4): 397–405.

[115] UVABRAINLAB (2020). Harmony case study.
URL: *https://osf.io/zextd/*

[116] Wang, Z., Xiong, H., Zhang, J., Yang, S., Boukhechba, M., Barnes, L. E., Zhang, D. and Dou, D. (2021). From personalized medicine to population health: A survey of mhealth sensing techniques, *arXiv preprint arXiv:2107.00948* .

[117] Wu, D., Liao, M., Zhang, W. and Wang, X. (2021). Yolop: You only look once for panoptic driving perception, *arXiv preprint arXiv:2108.11250* .

[118] Young, K. L., Osborne, R., Koppel, S., Charlton, J. L., Grzebieta, R., Williamson, A., Haworth, N., Woolley, J. and Senserrick, T. (2019). What contextual and demographic factors predict drivers' decision to engage in secondary tasks?, *IET Intelligent Transport Systems* **13**(8): 1218–1223.

[119] Yüce, A., Gao, H., Cuendet, G. L. and Thiran, J.-P. (2016). Action units and their cross-correlations for prediction of cognitive load during driving, *IEEE Transactions on Affective Computing* **8**(2): 161–175.

[120] Zepf, S., Dittrich, M., Hernandez, J. and Schmitt, A. (2019). Towards empathetic car interfaces: Emotional triggers while driving, *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.

[121] Zheng, R., Yamabe, S., Nakano, K. and Suda, Y. (2015). Biosignal analysis to assess mental stress in automatic driving of trucks: Palmar perspiration and masseter electromyography, *Sensors* **15**(3): 5136–5150.

List of Publications

[1] Arash Tavakoli and Arsalan Heydarian. Multimodal driver state modeling through unsupervised learning. *Accident Analysis & Prevention*, 170:106640, 2022.

[2] Arash Tavakoli, Shashwat Kumar, Xiang Guo, Vahid Balali, Mehdi Boukhechba, and Arsalan Heydarian. Harmony: A human-centered multimodal driving study in the wild. *IEEE Access*, 9:23956–23978, 2021.

[3] Arash Tavakoli, Vahid Balali, and Arsalan Heydarian. How do environmental factors affect drivers' gaze and head movements? Technical report, Mineta Transportation Institute, 2021.

[4] Arash Tavakoli, Shashwat Kumar, Mehdi Boukhechba, and Arsalan Heydarian. Driver state and behavior detection through smart wearables. In *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, July 2021.

[5] Arash Tavakoli, Mehdi Boukhechba, and Arsalan Heydarian. Leveraging ubiquitous computing for empathetic routing: A naturalistic data-driven approach. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2021.

[6] Arash Tavakoli, Mehdi Boukhechba, and Arsalan Heydarian. Personalized driver state profiles: A naturalistic data-driven study. In *International Conference on Applied Human Factors and Ergonomics*, pages 32–39. Springer, 2020.

[7] Arash Tavakoli, Vahid Balali, and Arsalan Heydarian. A multimodal approach for monitoring driving behavior and emotions. Technical report, Transportation Research Board, Washington DC, United States, 2019.

[8] Benjamin D Bowes, Arash Tavakoli, Cheng Wang, Arsalan Heydarian, Madhur Behl, Peter A Beling, and Jonathan L Goodall. Flood mitigation in coastal urban catchments using real-time stormwater infrastructure control and reinforcement learning. *Journal of Hydroinformatics*, 23(3):529–547, 2021.

[9] Cheng Wang, Benjamin Bowes, Arash Tavakoli, Stephen Adams, Jonathan Goodall, and Peter Beling. Smart stormwater control systems: A reinforcement learning approach. In *Proceedings of the ISCRAM Conference Proceedings—17th International Conference on Information Systems for Crisis Response and Management, Blacksburg, VA, USA*, pages 24–27, 2020.

[10] Arash Tavakoli, Amir Ashrafi, Arsalan Heydarian, and Madhur Behl. The internet of wasted things (iowt). In *Proceedings of the 8th International Conference on the Internet of Things*, pages 1–3, 2018.