

Emotional Impacts on Driver Behavior: An Emo-Psychophysical Car-Following Model

Bryan James Higgs

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Civil Engineering

Montasir M. Abbas, Chair
Antonio A. Trani
Bruce H. Friedman
Feng Guo

July 10th, 2014
Blacksburg, VA

Keywords: naturalistic data, psychophysiology, data clustering, driving simulator, emotion

Emotional Impacts on Driver Behavior: An Emo-Psychophysical Car-Following Model

Bryan Higgs

Abstract

This research effort aims to create a new car-following model that accounts for the effects of emotion on driver behavior. This research effort is divided into eight research milestones: (1) the development of a segmentation and clustering algorithm to perform new investigations into driver behavior; (2) the finding that driver behavior is different between drivers, between car-following periods, and within a car-following period; (3) the finding that there are patterns in the distribution of driving behaviors; (4) the finding that driving states can result in different driving actions and that the same driving action can be the result of multiple driving states; (5) the finding that the performance of car-following models can be improved by calibration to state-action clusters; (6) the development of a psychophysiological driving simulator study; (7) the finding that the distribution of driving behavior is affected by emotional states; and (8) the development of a car-following model that incorporates the influence of emotions.

Dedication

I would like to dedicate this work to my beloved wife, Arielle. Her constant support and love has greatly strengthened me in the pursuit of my goals. I would also like to thank all those that have taken their time to teach and encourage me as I have matured into the person I am today. I would like to especially thank my advisor, Montasir Abbas, for his wise direction and mentoring that has greatly aided me in the fulfillment of this work and has greatly prepared me for the future I have before me.

Table of Contents

Abstract	ii
Dedication	iii
Table of Contents	iv
List of Figures	viii
List of Tables	ix
1. Introduction.....	1
1.1 Research Objectives	1
1.2 Dissertation Contribution	1
1.3 Dissertation Organization.....	1
2. A Two-Step Segmentation Algorithm for Behavioral Clustering of Naturalistic Driving Styles 3	
Abstract	3
2.1 Introduction	4
2.1.1 Naturalistic Driving Data	4
2.1.2 Synthesis of Past Efforts	5
2.2 Methodology	6
2.2.1 Car-Following Period Extraction	7
2.2.2 Segmentation Algorithm.....	7
2.2.3 Clustering Algorithm	8
2.3 Results	10
2.4 Conclusions	12
Acknowledgment	12
References	13
3. Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns 14	
Abstract	14
3.1 Introduction	15
3.2 Synthesis of Past Efforts	16
3.2.1 Car-Following Data Collection Methods.....	16
3.2.2 Car-Following Models	18
3.2.3 Data Clustering	18
3.3 Methodology	19
3.3.1 Naturalistic Driving Data	20
3.3.2 Extraction of Car-Following Periods	22

3.3.3	Segmentation Optimization Equation	22
3.3.4	Clustering	23
3.3.5	Optimization of Car-Following model.....	24
3.4	Results	25
3.4.1	Segment Duration Analysis	25
3.4.2	Multi-resolution Comparison of Error	25
3.4.3	Car Drivers.....	26
3.4.4	Truck Drivers	34
3.5	Discussion and Implications.....	39
3.6	Conclusions	40
	References	41
4.	Identification and Classification of State-Action Clusters of Car-Following Behavior	43
	Abstract	43
4.1	Introduction	44
4.1.1	Synthesis of Past Efforts	44
4.2	Methodology	44
4.2.1	Naturalistic Driving Data	45
4.2.2	Extraction of Car-Following Periods	45
4.2.3	Segmentation Optimization Equation	45
4.2.4	Clustering.....	46
4.2.5	Discriminant Analysis.....	46
4.3	Results	46
4.4	Discussion and Implications.....	53
4.5	Conclusions	55
	References.....	55
5.	Multi-Resolution Comparison of Car-Following Models using Naturalistic Data.....	57
	Abstract	57
5.1	Introduction	58
5.1.1	Naturalistic Driving Data	58
5.1.2	Synthesis of Past Efforts	59
5.2	Methodology	61
5.3	Results	65
5.3.1	Driver Level of Analysis.....	65
5.3.2	Car-Following Period Level of Analysis	66

5.3.3	Cluster Level of Analysis.....	68
5.4	Implications.....	68
5.5	Conclusions.....	69
	References.....	70
6.	Experimental Design for a Psychophysiological Driving Simulator Study.....	72
	Abstract.....	72
6.1	Introduction.....	73
6.1.1	Synthesis of Past Efforts.....	73
6.2	Methodology.....	76
6.2.1	Personality Survey.....	76
6.2.2	Emotion Surveys.....	77
6.2.3	Physiological Data Collection.....	77
6.2.4	Driving Simulator Scenarios.....	79
6.2.5	Classification of Psychological Data.....	81
6.2.6	Study Procedure.....	81
6.3	Results.....	82
6.4	Discussion and Implications.....	84
6.5	Conclusions.....	84
	References.....	85
7.	Development of an Emotional Car-Following Model.....	86
	Abstract.....	86
7.1	Introduction.....	87
7.1.1	Synthesis of Past Efforts.....	87
7.2	Methodology.....	88
7.2.1	Driving Simulator Experimental Design.....	88
7.2.2	Car-Following Data Segmentation and Clustering.....	89
7.2.3	Psychological Data Classification.....	89
7.3	Results.....	90
7.4	Discussion and Implications.....	93
7.5	Conclusions.....	94
	References.....	94
8.	Investigation of the Effects of Physiological States on Driver Behavior.....	96
	Abstract.....	96
8.1	Introduction.....	97

8.1.1	Synthesis of Past Efforts	97
8.2	Methodology	97
8.3	Results	98
8.4	Conclusions	101
	References	102
9.	Development of an Emo-Psychophysical Car-Following Model	104
	Abstract	104
9.1	Introduction	105
9.1.1	Synthesis of Past Efforts	105
9.1.2	Effects of Emotion on Car-Following Behavior	108
9.2	Methodology	108
9.2.1	Psychophysiological Driving Simulator Study	108
9.2.2	Wiedemann Car-Following Model	108
9.2.3	Wiedemann-GHR Model	109
9.3	Results	110
9.4	Conclusions	112
	References	112
10.	Summary of Findings	115
10.1	Findings	115
10.2	Limitations of the Study	116
10.3	Recommendations for Future Research	116

List of Figures

Figure 2-1: Conceptual Representation of the Algorithm Design	9
Figure 2-2: Cluster Distribution of High-Risk Driver	11
Figure 2-3: Cluster Distribution of Medium-Risk Driver.....	11
Figure 2-4: Cluster Distribution of Low-Risk Driver	11
Figure 2-5: Cluster Distributions of Car-Following Period 1 for the High-Risk Driver	12
Figure 2-6: Cluster Distributions of Car-Following Period 2 for the High-Risk Driver	12
Figure 3-1: Relationship between Driving States and Driving Actions	15
Figure 3-2: Conceptual Representation of the Algorithm Design	20
Figure 3-3: Sum of Squared Error for Different Numbers of Clusters	24
Figure 3-4: Comparison of Error for Different Levels of Calibrated Parameters of the GHR Model	26
Figure 3-5: Circular Plot of Cluster Frequencies for Car Drivers (A-J).....	27
Figure 3-6: Example Segmented and Clustered Car-following Period	31
Figure 3-7: Time-Space Comparison of Clusters	34
Figure 3-8: Circular Plot of Cluster Frequencies for Truck Drivers (A-J)	35
Figure 4-1: Example Segmented and Clustered Car-following Period	46
Figure 4-2h: Discriminant Analysis Classification Matrix for Yaw Angle.....	51
Figure 5-1: Comparison of GHR Model Calibrations to Different Levels of Analysis.....	65
Figure 5-2: Histogram of RMSE of Car-Following Periods using Parameters Calibrated to Driver F	66
Figure 5-3: Comparison of Models for Example Car-Following Period.....	67
Figure 5-4: Comparison of Top 3 Models for Example Car-Following Period	67
Figure 6-1: Psychlab Physiological Data Collection System	78
Figure 6-2: PsychLab Acquire Software.....	79
Figure 6-3: Picture of Drive Safety Driving Simulator	80
Figure 6-4: HyperDrive Screenshot.....	80
Figure 6-5: Comparison of Example Car-following Periods for Driver A.....	83
Figure 6-6: Comparison of Example Car-following Periods for Driver B	84
Figure 7-1: Locations of Various Emotional States in Valence (Pleasure) and Arousal Space [23]	90
Figure 7-2: Cluster Distributions by Emotion	90
Figure 7-3: Range vs. Range Rate for Three Largest Clusters	91
Figure 7-4: Transition Probability Trees for Multiple Emotions.....	93
Figure 8-1: Distribution of Physiological Clusters	98
Figure 8-2: R-R Interval over Time for Anger Scenario	99
Figure 9-1: Wiedemann 74 Car Following Logic [38]	109
Figure 9-2: Wiedemann Thresholds of the W-GHR Model for Different Emotions.....	112

List of Tables

Table 2-1: Number of Segments for Each Car-Following Period by Driver	10
Table 3-1: Types of Vehicle Trajectory Data Collection.....	18
Table 3-2: Synthesis of Car-Following Models.....	18
Table 3-3: Time-Series Clustering Methods.....	19
Table 3-4: Naturalistic Data Car-following Variables.....	22
Table 3-5: Cluster Occurrences, Average Lengths, and Standard Deviations for Car Drivers (A-J)	30
Table 3-6: Calibrated GHR Parameters for Each Cluster for Car Drivers	33
Table 3-7: Cluster Occurrences, Average Lengths, and Standard Deviations for Truck Drivers (A-J).....	37
Table 3-8: Calibrated GHR Parameters for Each Cluster for Truck Drivers.....	39
Table 4-1: Classification of Clusters.....	53
Table 5-1: Multiple Resolutions of Analysis.....	62
Table 5-2: Table of Car Following models and parameters	64
Table 5-3: RMSE by Model for All Car-Following Periods of Each Driver.....	66
Table 5-4: RMSE of Segmented Models by Cluster for Example Car-Following Period.....	68
Table 6-1: Types of Vehicle Trajectory Data Collection.....	75
Table 6-2: Psychological State and Trait Inventories	75
Table 6-3: Methods for Physiological Data Collection [14].....	76
Table 6-4: Simulator Data Variables	82
Table 7-1: Psychological State and Trait Inventories	88
Table 8-1: Distribution of State-Action Clusters for Each Physiological Cluster	101
Table 9-1: Root Mean Squared Error of Wiedemann and Wiedemann-GHR Models for Different Emotions	110

1. Introduction

The accuracy of current car-following models is compromised by one key element that is not fully considered, the human element. There is wide variability within human drivers that goes unexplained by current car-following modeling techniques resulting in high error values and thus inaccurate traffic simulations. The variability of driver behavior is one of the most difficult aspects to consider in modeling. The real challenge to conquering this difficulty is that the variability can stem from numerous sources including: emotion, hunger, sensitivity to light conditions, personality, thirst, etc. Emotion is potentially the most influential source of variability as evidenced by the existence of road rage and driving anger. Road rage is very intriguing as it is a combination of personality traits and emotional state. Road rage occurs more frequently in some people versus others which indicates that some personalities are more resistant to road rage than others. Also, different personalities have different ways of venting their anger. For example, some people express their anger verbally (yelling and cursing) while others express it physically (erratic accelerations), and in some cases people focus their anger in a vengeful manner (attacking a certain vehicle). This effort proposes a new car-following model that captures the effects of emotion on driver behavior.

1.1 Research Objectives

The objectives of this research effort are as follows:

- Investigate the variability that exists in driver behavior by examining the difference in behavior that exist: (1) between drivers, (2) between car-following periods, and (3) within a car-following period.
- Investigate the strengths and weaknesses of existing car-following models in terms of driver variability.
- Develop a driving simulator study that also collects psychophysiological data
- Investigate the effects of emotion on driver behavior and represent those effects in a new car-following model.

1.2 Dissertation Contribution

This dissertation uses both Naturalistic and driving simulator data to contribute new findings in the variability of driver behavior and the role of emotions. There are two main contributions of this dissertation: (1) a segmentation and clustering method for the analysis of car-following behavior and (2) a new emo-psychophysical car-following model. The segmentation and clustering method adds the ability to examine car-following behavior in more depth by adding a new resolution (within a car-following period) to the data. The new emo-psychophysical car-following model gives an existing model the ability to model the differences in driver behavior due to different emotional states.

1.3 Dissertation Organization

This dissertation is composed of ten chapters. Chapter 1 presents an introduction, the research objectives, the contribution, and the organization of the Dissertation. Chapter 2 presents a multi-level analysis of the heterogeneity of the car-following behavior of drivers. Chapter 3 presents an investigation of driving patterns that different drivers exhibit. Chapter 4 presents an analysis of the defining characteristics of the state-action clusters found in previous research. Chapter 5 presents a multi-resolution comparison of car-following models. Chapter 6 presents the

development of a psychophysiological driving simulator study. Chapter 7 presents the development of a car-following model that accounts for differences in behavior due to the emotional state. Chapter 8 presents an investigation of the influence of physiological changes on driver behavior. Chapter 9 presents an emo-psychophysical car-following model. Lastly, Chapter 10 presents the dissertation conclusion, the limitations of the study, and future recommendations.

2. A Two-Step Segmentation Algorithm for Behavioral Clustering of Naturalistic Driving Styles¹

Abstract

This research effort aimed to investigate the hypothesis that drivers apply different driving styles in their daily driving tasks. A two-step algorithm was used for segmentation and clustering. First, a car-following period was broken into different duration segments that account for their temporal distribution. Second, the segments produced by the previous step were clustered based on similarity. Variations of k-means clustering and optimization techniques were used in this process. The segments centroids, used for clustering, were 8-dimensional and were produced by taking the average of the data points in each segment based on longitudinal acceleration, lateral acceleration, gyro (yaw rate), vehicle speed, lane offset, gamma (yaw angle), range, and range rate. The results of this methodology were continuous segments of car-following behavior as well as clusters of segments that show similar data and thus similar behaviors. The sample used in this paper included three different truck drivers that are representative of a high-risk driver, a medium-risk driver, and a low-risk driver. In summary, the results revealed behavior that changed within a car-following period, between car-following periods, and between drivers. Each driver showed a unique distribution of behavior, but some of the behaviors existed in more than one driver but at different frequencies.

¹ Paper has been published in the proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC 2013)

2.1 Introduction

Several assumptions are made in the modeling of the car-following behaviors of drivers. These assumptions make a significant impact upon agent-based simulations of transportation networks. The biggest of these assumptions is homogeneity among drivers. This means that all drivers are assumed to have the same behavior (homogeneity) in all car-following periods. Typically, heterogeneity means that different drivers have different behaviors, but this still assumes homogeneity within the driver over all car-following periods. Human behavior is attributed as erratic which would contradict an assumption of homogeneity. The fact is that any time data are grouped; there is an assumption of homogeneity. This research effort aims to analyze different levels of the assumption of homogeneity as follows:

1. Homogeneity amongst all drivers (i.e. all drivers have the same behavior)
2. Heterogeneity amongst drivers, but with homogeneity within the driver (i.e. a driver has the same behavior in all car-following periods)
3. The behavior of drivers remains constant in a car-following period (i.e. a driver has the same behavior throughout a car-following period)

The analysis in this paper addresses each of these assumptions through a segmentation and clustering algorithm. The algorithm first divides each car-following period into segments then the algorithm clusters the segments together based upon similarities in the data or behavior. For analysis of the first assumption, homogeneity among drivers, the segments and clusters are grouped by driver to highlight the similarities and differences between the behaviors of drivers. For analysis of the second assumption, heterogeneity among drivers with homogeneity within a driver, the segments and clusters are grouped by car-following period to highlight the differences between car-following periods. The analysis of the third assumption, driver behavior is constant during a car-following period, is the basis for segmentation and clustering and is thus covered through analysis of the other two assumptions.

This effort proposes a methodology for segmenting car-following periods into different regimes. However, the work does not stop at this point because the regimes may repeat and thus clustering of the segments is needed in order to properly define the structure of the regimes. In the following sections, we first present a synthesis of past effort, then we describe the naturalistic driving data that were previously collected by Virginia Tech Transportation Institute (VTTI), followed by a description of our methodology, and finally we present our findings and conclusion.

2.1.1 Naturalistic Driving Data

Naturalistic driving data refers to the process of collecting data related to drivers' actions as they operate vehicles that have been equipped with specialized sensors along with processing and recording equipment. In effect, the vehicle becomes a moving data collection device. The drivers operate and interact with these vehicles during their normal driving routines while the data collection equipment is continuously recording numerous items of interest during the entire driving. Naturalistic data collection methods require a sophisticated network of sensor, processing, and recording systems. This system provides a diverse collection of both on-road driving and driver (participant, non-driving) data, including measures such as driver input and performance (e.g., lane position, headway, etc.), four camera video views, and driver activity data. This information may be supplemented by subjective data, such as questionnaire data.

As part of a naturalistic truck driving study (NTDS) conducted by VTTI [1], one hundred drivers were recruited from four different trucking fleets across seven terminals and one to three trucks at each trucking fleet were instrumented (nine trucks total). After a participant finished 4 consecutive weeks of data collection, another participant started driving the instrumented truck. Three forms of data were collected by the NTDS DAS: video, dynamic performance, and audio. Approximately 14,500 driving-data hours covering 735,000 miles traveled were collected. Nine trucks were instrumented with the DAS.

2.1.2 Synthesis of Past Efforts

Ossen and Hoogendorn [2] studied the car-following behavior of individual drivers using vehicle trajectory data that were extracted from high-resolution digital images collected at a high frequency from a helicopter. The analysis was performed by estimating the parameters of different specifications of the GHR car-following rule for individual drivers. In 80 % of the cases, a statistical relation between stimuli and response could be established. The Gipps (a safe distance model) and Tampere (stimulus-response model) models and a synthetic data based approach were used for assessing the impact of measurement errors on calibration results. According to the authors, the main contribution of their study was that considerable differences between the car-following behaviors of individual drivers were identified that can be expressed in terms of different optimal parameters and also as different car-following models that appear to be optimal based on the individual driver data. This is an important result taking into account that in most models a single car-following rule is used. The authors also proposed for future research to apply more advanced statistical methods and to use larger databases. Brackstone et. al. [3] used data collected with an instrumented vehicle that was assembled at TRG Southampton to parameterize the Wiedemann's threshold for a typical following spiral. As a result they represent the action points as a function of a probability distribution based on ground speed.

Micro-simulation software packages use a variety of car-following models including Gipps' (AIMSUN, SISTM, and DRACULA), Wiedemann's (VISSIM), Pipe's (CORSIM), and Fritzsche's (PARAMICS). And different automated calibration parameters such as genetic algorithms have been used to calibrate the distribution of car-following sensitivity parameters [4]. Panwai and Dia [5] compared the car-following models between different simulation software, including AIMSUN, PARAMICS and VISSIM using an instrumented vehicle to record differences in speed and headway (Leading speed, relative distance, relative speed, follower acceleration were recorded). The EM shows similar values for psychophysical models in VISSIM and PARAMICS and lower values in AIMSUN. The RMS error and qualitative drift and goal-seeking analyses also showed a substantially different car-following behavior for PARAMICS. Siuhi and Kaseko [6] demonstrated the need for separate models for acceleration and deceleration responses by developing a family of car-following models and addressing the shortcomings of the GM model. Previous work from Osaki [7] and Subranmanian [8] modified the GM model separating the acceleration and deceleration responses. Ahmed [9], following some work from Subranmanian assumed non linearity in the stimulus term and introduced traffic density. Results from Ahmed [9] and Toledo [10] showed, against popular belief, that acceleration increases with speed but decreases with vehicle separation. Due to statistical insignificance, Ahmed and Toledo also removed speed from their deceleration models. Siuhi and Kasvo [6] addressed some of these shortcomings by developing separate models, not only for acceleration and deceleration, but also for steady-state responses. Nonlinear regression with robust standard errors was used to estimate the model parameters and obtain the distributions across drivers. The stimulus response thresholds

that delimit the acceleration and deceleration responses were determined based on signal detection theory.

Using two models of similar complexity (number of parameters): the “Intelligent Driver Model” (IDM) and the “Velocity Difference Model” (VDIFF), Kesting and Treiber [11] researched car-following behaviors on individual drivers using publicly available trajectory data for a straight one-lane road in Stuttgart, Germany. They used a nonlinear optimization procedure based on a genetic algorithm to minimize the deviations between the observed driving dynamics and the simulated trajectory. One of the major findings of the study was that a significant part of the deviations between measured and simulated trajectories can be attributed to the inter-driver variability and the intra-driver variability (stipulating that human drivers do not drive constantly over time, and their behavioral driving parameters change)—the latter accounts for a large part of the deviations between simulations and empirical observations.

Menneni et al [12] presented a calibration methodology of the VISSIM Wiedemann car-following model based on integrated use of microscopic and macroscopic data using NGSIM Relative distance vs. relative speed graphs were used for the microscopic calibration, specifically to determine the action points (it is important to note that action points were not identical to perception threshold). Scatter and distribution of action points on relative distance versus relative velocity graphs also showed similarity in driver behavior between the two freeways.

Hoogendoorn and Hoogendoorn [13] proposed a generic calibration framework for joint estimation of car-following models. The method employed relies on the generic form of most models and weights each model based on its complexity. This new approach can cross-compare models of varying complexity and even use multiple trajectories when individual trajectory data is scarce. Prior information can also be used to realistically estimate parameter values.

The research into clustering data is divided into two main groups: univariate clustering and multivariate clustering. Within each group there are many facets, but for the application to modeling car-following behavior, the facet of time-series clustering is the most important. Most of the research into clustering has occurred outside of the field of transportation engineering. Cui et al. presented a methodology of clustering drivers based upon questionnaire results [14]. This clustering is simple and not as advanced as the method being researched in other fields.

Univariate clustering is easier than multivariate clustering thus there is more extensive research into using univariate clustering methods. Fan et al. presented a methodology for clustering time-series data. This methodology has two main parts: the segmentation of the data and then the clustering of the segments. The paper suggests various algorithms for both segmentation and clustering [15]. Kremer et al. also worked with time-series clustering, but the focus of the effort was placed upon the changes or evolution of different clusters over time [16].

The research into multivariate clustering is limited, but the impacts of the researched methods have a much broader range of applications than univariate clustering methods. Plant et al. worked with the clustering of multivariate time-series. The researchers emphasized clustering based upon the interactions of the different variables [17]. Wang et al. conducted research on the clustering of multivariate time-series data. Their research was based upon the reduction of various segments of the data to statistical measures which could then be clustered by normal k-means methods [18].

2.2 Methodology

The methodology used in this study first extracted car-following periods from the exorbitant amount of data contained within the Naturalistic Database. The methodology then goes into detail

about the algorithm used for the creation of segments or regimes in each car-following period. Lastly, the clustering algorithm is discussed.

2.2.1 Car-Following Period Extraction

Car-following situations were automatically extracted from the enormous volume of driving data in the database in order to analyze the car-following driver behavior. The filtering process was an iterative process where initial values and conditions were used and after the events were flagged they are reviewed in the video data to adjust the values accordingly in order to obtain minimum noise. Visual inspection of the first subsets created revealed some non car-following events so additional filtering was thus performed to remove these events from the database.

Specifically, car-following periods were extracted automatically according to these conditions:

- Radar Target ID>0

This eliminates the points in time without a radar target detected

- Radar Range<=120 meters

This represents four seconds of headway at 70 mph

- $-1.9 \text{ meters} < \text{Range} * \sin(\text{Azimuth}) < 1.9 \text{ meters}$

This restricts the data to only one lane in front of the lead vehicle

- Speed>=20km/h

This speed was used in order to minimize the effect of traffic jams, but still leave the influence of congestion in the data

- Rho-inverse <=1/610 meters⁻¹

This limits the curvature of the roadway such that vehicles are not misidentified as being in the same lane as the subject vehicle when roadway curvature is present.

- Length of car-following period while the Range is less than 61 meters >= 30 seconds

This criterion was found by trial and error using video analysis to verify positive or negative results.

The automatic extraction process was verified from a sample of events through video analysis. For the random sample of 400 periods, 392 were valid car-following periods.

2.2.2 Segmentation Algorithm

The segmentation part of the algorithm was based upon k-means clustering, but was adapted for time-dependent clustering. Normal k-means clustering assigns data points to the nearest cluster centroid, whereas this adaptation of k-means found segment lengths with each segment having its own centroid. Equation 1 and Equation 2 show the mathematics used in the segmentation algorithm.

Equation 1: Multivariate Time-Series Centroid Clustering

$$\min_s \sum_{i=1}^k \sum_{a=1}^m \sum_{j=1}^n \frac{\|x_{aj} - \mu_{i_a}\|}{\max_j x_{aj}}$$

Where:

x_{aj} is the j th observation of variable x_a

μ_{i_a} is the centroid of cluster i for variable a

n is the number of observations

k is the number of clusters

a is the number of variables

Equation 2: Multivariate Time-Series Centroid

$$\mu_{i_a} = \frac{\sum_{j=1}^n x_{aj}}{n}$$

Where:

x_{aj} is the j th observation of variable x_a

μ_{i_a} is the centroid of cluster i for variable a

n is the number of observations

k is the number of clusters

a is the number of variables

The centroids had eight dimensions that included the following: Longitudinal Acceleration, Lateral Acceleration, Gyro (Yaw Rate), Vehicle Speed, Lane Offset, Gamma (Yaw Angle), Range, and Range Rate. The distance or error calculation divided the difference between the data points and the centroid value by the maximum value for the corresponding variable, thus scaling all the variable values. This part of the algorithm was coded into the optimization software, AIMMS, with the variables being the segment durations and all other variable being dependent upon the segment durations. The model in AIMMS changes the segment durations, calculates the segment centroids, and calculates the total error simultaneously.

2.2.3 Clustering Algorithm

The second part of this methodology clustered the segments, found in the first part of the algorithm, in order to find similar segments or behaviors in the car-following data. This part of the methodology used the statistical software, JMP, as it includes an established k-means clustering method.

The clustering was based upon the segment centroids, found in the previous step. This process is illustrated in Figure 2-1 below.



Segmentation



Clustering

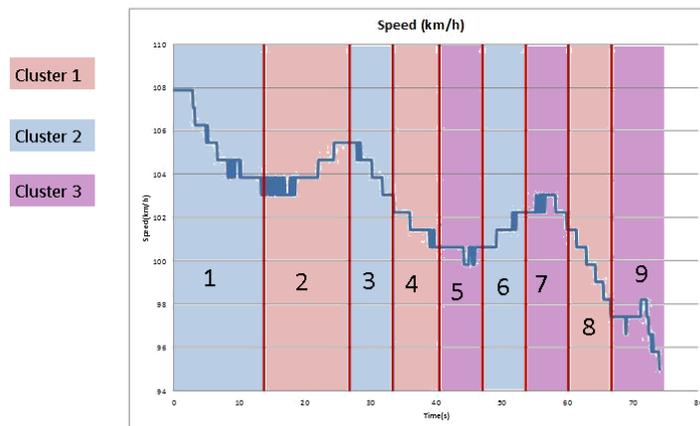


Figure 2-1: Conceptual Representation of the Algorithm Design

In this paper, the car-following behaviors of three different drivers were compared. These drivers were chosen because they represent high-risk, medium-risk, and low-risk drivers. The risk was assessed through the number of conflicts each driver experienced during the collection of the

naturalistic data. Ten car-following periods were used for each driver and the length of these periods were a minimum of 30 seconds. The aforementioned algorithm was used to segment and cluster the behavior of the three drivers.

2.3 Results

Table 2-1 shows that the number of segments created by the segmentation part of the algorithm varies by driver and by car-following period. For example, car-following period 1 for the high-risk driver has 9 segments while car-following period 3 for the high-risk driver only has 4 segments. The results in this table contradict the three main assumptions that were listed in the introduction. The first assumption, homogeneity among drivers, was violated by the fact that the average number of segments was not the same for all drivers. The second assumption, heterogeneity among drivers with homogeneity within a driver, was violated by the fact that the number of segments was not constant for all the car-following periods of each driver. The third assumption, constant behavior in a car-following period, was violated by the fact that segments exist in the data.

Table 2-1: Number of Segments for Each Car-Following Period by Driver

Car-Following Period	High-Risk	Medium-Risk	Low-Risk
1	9	10	10
2	8	9	9
3	4	7	9
4	4	7	9
5	4	6	10
6	10	6	10
7	9	5	10
8	8	4	9
9	7	3	8
10	7	3	6
Average	7	6	9
Standard Deviation	2.26	2.36	1.25

Further analysis of the formed clusters provided more insight into evaluation of the main assumptions. To continue the analysis, the clusters had to be grouped by driver (assumption 1) and by car-following period (assumption 2).

Figure 2-2 shows that the cluster distribution of the High-Risk Driver had a strong dependence upon clusters 8 and 24 and there was a total of 10 clusters. Figure 2-3 shows the cluster distribution of the medium-risk driver. It is interesting to note that though the two drivers both exhibited 10 total clusters, the cluster distributions were different. The same clusters that dominated the behavior of the high-risk driver, clusters 8 and 24, appeared in the behavior of the medium-risk driver, but in a relatively low amount. Figure 2-4 shows the cluster distribution of the low-risk driver which exhibited 21 total clusters, more than twice as many as the other drivers indicating very complex behaviors. Comparison of the distributions of all three drivers revealed that there were some clusters that were unique to each driver, but some clusters were shared by

more than one driver. In consideration of the assumption of homogeneity among drivers, the three figures showed heterogeneous or driver dependent distributions.

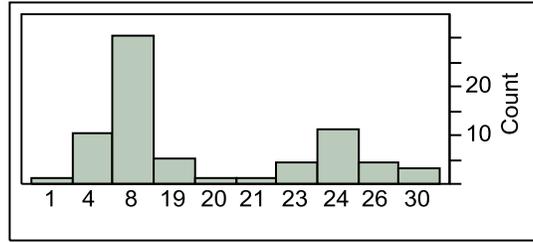


Figure 2-2: Cluster Distribution of High-Risk Driver

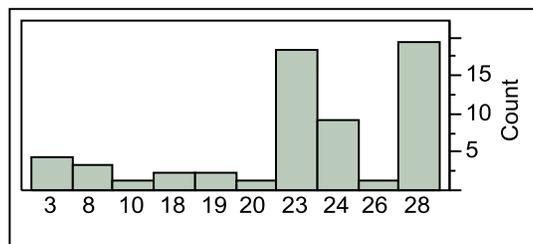


Figure 2-3: Cluster Distribution of Medium-Risk Driver

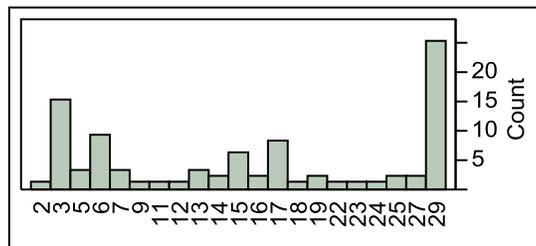


Figure 2-4: Cluster Distribution of Low-Risk Driver

Figure 2-5 shows that three separate clusters composed the behavior of the High-Risk Driver in the first car-following period which indicated a very simple behavior. Figure 2-6 shows the cluster distributions for the second car-following period of the High-Risk Driver. Comparison of these two figures revealed that the second car-following period was a more complex variation of the first period by simply adding two additional clusters. These two clusters could appear for a number of reasons including: driver adaptation to a new situation, driver fatigue, driver distraction, etc. In regard to the assumption of heterogeneity among drivers with homogeneity within a driver, the two figures show a difference which could be attributed to any number of variables including: emotions of the driver, different environmental conditions, following a larger vehicle, reaction to the behaviors of other drivers, etc. Each of the figures show clusters of different behavior within a car-following period which violated the assumption of constant behavior within a car-following period.

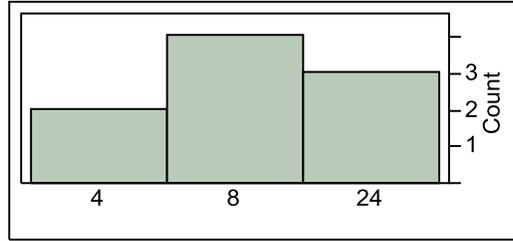


Figure 2-5: Cluster Distributions of Car-Following Period 1 for the High-Risk Driver

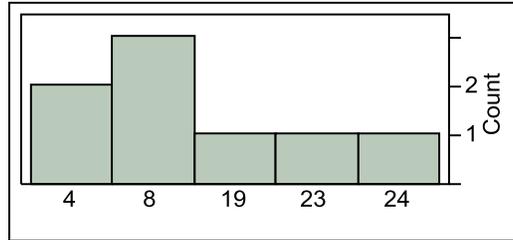


Figure 2-6: Cluster Distributions of Car-Following Period 2 for the High-Risk Driver

2.4 Conclusions

This research effort aimed to investigate three main assumptions of the car-following behavior of drivers: homogeneity among drivers, heterogeneity among drivers with homogeneity within a driver, and constant behavior during a car-following period. The methodology used in this effort was a 2-step algorithm that would first segment car-following periods and then cluster these individual segments based upon longitudinal acceleration, lateral acceleration, gyro (yaw rate), speed, lane offset, gamma (yaw angle), range, and range rate. The sample for this study was three drivers with ten car-following periods each. These drivers were chosen because they represented a low-risk driver, a medium-risk driver, and a high-risk driver. The research findings show that all three assumptions were violated by the data. In summary, the results revealed behavior that changed within a car-following period, between car-following periods, and between drivers. Each driver showed a unique distribution of behavior, but some of the behaviors existed in more than one driver but at different frequencies. This suggests that a measure of the safety of each driver can be derived by analyzing and attributing certain behaviors to certain clusters. For example, cluster 8 could be the behavior of tailgating which was shown in high frequency in the high-risk driver, low frequency in the medium-risk driver, and not at all in the low-risk driver. The width and composition of the distributions change between car-following periods suggesting adaptability in the behavior of the drivers and a very strong dependence upon additional factors that need to be identified. Further research should include more car-following periods and analyze the difference between the car-following periods in depth in order to identify key variables that affect the behavior distributions of drivers. These results show that this methodology was very effective in exploring and discovering correlation and patterns in car-following behaviors that are otherwise concealed within the data.

Acknowledgment

This material is based upon work supported by the Federal Highway Administration under Agreement No. DTFH61-09-H-00007. Any opinions, findings, and conclusions or

recommendations expressed in this publication are those of the Author(s) and do not necessarily reflect the view of the Federal Highway Administration.

The authors would like to express thanks to individuals at Virginia Tech and the Virginia Tech Transportation Institute who contributed to the study in various ways: Greg Fitch, Shane McLaughlin, Zain Adam, Brian Daily, and Rebecca Olson.

References

- [1] R. Olson, R. J. Bocanegra, R. Hanowski, and J. Hickman, "Driver Distraction in Commercial Vehicle Operations," Center for Truck and Bus Safety, Virginia Tech Transportation Institute, Blacksburg, VA FMCSA-RRR-09-042, 2009.
- [2] S. Ossen and S. P. Hoogendoorn, "Validity of trajectory-based calibration approach of car-following models in presence of measurement errors," *Transportation Research Record*, pp. 117-125, 2008.
- [3] M. Brackstone, "Driver Psychological Types and Car Following: Is There a Correlation? Results of a Pilot Study," in *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2003, p. 6.
- [4] G. G. Schultz and L. R. Rilett, "Analysis of Distribution and Calibration of Car-Following Sensitivity Parameters in Microscopic Traffic Simulation Models," p. 11, 2004.
- [5] S. Panwai and H. Dia, "Comparative evaluation of microscopic car-following behavior," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 6, pp. 314-325, 2005.
- [6] S. Siuhi and M. Kaseko, "Parametric Study of Stimulus-Response Behavior for Car-Following Models," in *89th Transportation Research Board Annual Meeting*, Washington, D.C., 2010.
- [7] H. Osaki, "Reaction and anticipation in the Car-Following Behavior," in *12th International Symposium on the Theory of Traffic Flow and Transportation*, 1993.
- [8] H. Subranmanian, "Estimation of Car-Following Models," Master of Science in Transportation Master Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1996.
- [9] K. I. Ahmed, "Modeling Drivers' Acceleration and Lane Changing Behavior," Doctor of Science in Transportation Systems and Decision Sciences PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1999.
- [10] T. Toledo, "Integrating Driving Behavior," Doctor of Philosophy PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2003.
- [11] A. Kesting and M. Treiber, "Calibrating Car-Following Models using Trajectory Data: Methodological Study," *Transportation Research Record*, vol. 2088, pp. 148-156, 2008.
- [12] S. Menni, C. P. D. Sun, P.E., and P. Vortisch, "An Integrated Microscopic and Macroscopic Calibration for Psycho-Physical Car Following Models " *TRB 2009 Annual Meeting CD-ROM* p. 17, 2008.
- [13] S. Hoogendoorn and R. Hoogendoorn, "A Generic Calibration Framework for Joint Estimation of Car Following Models using Microscopic Data," *TRB 2010 Annual Meeting CD-ROM*, p. 17, 2010.
- [14] Y. Cui, Y. Zhang, y. Ma, and Y. Feng, "Cluster analysis on evaluation indicators of driver characteristics," in *2007 IEEE Intelligent Transportation Systems Conference, 30 Sept.-3 Oct. 2007*, Piscataway, NJ, USA, 2007, pp. 70-5.
- [15] Z.-H. Fan, D.-K. Huang, J.-Z. Li, and K.-H. Wang, "Morphing Cluster Dynamics: Time series clustering with multipartite graph," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics, August 26, 2004 - August 29, 2004*, Shanghai, China, 2004, pp. 1749-1754.
- [16] H. Kremer, S. Gunnemann, and T. Seidl, "Detecting Climate Change in Multivariate Time Series Data by Novel Clustering and Cluster Tracing Techniques," in *10th IEEE International Conference on Data Mining Workshops*, Sydney, NSW, Australia, 2010, pp. 96-97.
- [17] C. Plant, A. M. Wohlschlagel, and A. Zherdin, "Interaction-Based Clustering of Multivariate Time Series," in *Data Mining, 2009. ICDM '09. Ninth IEEE International Conference on*, 2009, pp. 914-919.
- [18] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in *7th IEEE International Conference on Data Mining*, Omaha, NE, United states, 2007, pp. 351-360.

3. Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns²

Abstract

Driving behavior can be influenced by many factors that are not feasible to collect in driving behavior studies. The research presented in this paper investigated the characteristics of a wide range of driving behaviors linking driving states to the drivers' actions. The proposed methodology was structured such that a known state can be linked to multiple actions, thus accounting for the effects of the unknown driving state factors. A two-step algorithm was developed and used for segmentation and clustering of driving behaviors. The algorithm segmented and clustered car-following behavior based on eight state-action variables: longitudinal acceleration, lateral acceleration, yaw rate, vehicle speed, lane offset, yaw angle, range, and range rate. The results of this methodology were state-action clusters that define the driving pattern of drivers. The sample used in this paper included twenty different drivers—ten car and ten truck drivers. The results revealed that behavior patterns were different between car drivers, but not truck drivers. The results also showed that car drivers exhibit behaviors that are unique to each driver while truck drivers show a common driving pattern. Characteristics and frequency of recognized driving patterns were provided in the paper, along with the corresponding modeling parameters of each pattern.

² Paper has been accepted for journal publication in IEEE Intelligent Transportation Systems Transactions: Special Issue on Uncertainty in Computational Traffic Models

3.1 Introduction

A driver's behavior, as defined in this paper, is the way a driver responds to their existing driving state (e.g., vehicle's speed, distance to the vehicle in front, etc.) by implementing a certain action (e.g., accelerate, steer, etc.). A driver behavior can therefore be formally defined as the function that maps traffic states to driver's actions. Typically, driver behavior studies collect data that encompass both states and actions, and attempt to develop a more accurate mapping between them. These efforts usually do not include many influencing factors of human behavior, such as emotion, personality, hunger, thirst, etc. [1].

This concept is illustrated in Figure 3-1, where the driver actions are shown to depend on the driving states through a mapping function. There are two major points that need to be pointed out in the figure: (1) data collection methods usually capture only a subset of the variables that fully define the driving states and (2) despite this fact, current modeling techniques assume a single mapping function for each driver. As a result, and regardless of how complex the mapping function is, existing methods will still produce large errors due to the attempt of pooling what are essentially different states (differing primarily in unmeasured variables, such as driver's emotion, weather, location, etc.) into the same states, and producing functions that map the average of these states to the observed actions, with reduced accuracy.

Our proposed concept, as shown in the same figure (shown as the divided pie chart), allows the decomposition of the single mapping function assigned to the driver, into several functions. Each of these functions would cover a different part of the state space to minimize the overall modeling error. The number of mapping functions and the corresponding identification and size of covered state space is what we define in this paper as a *driving pattern*.

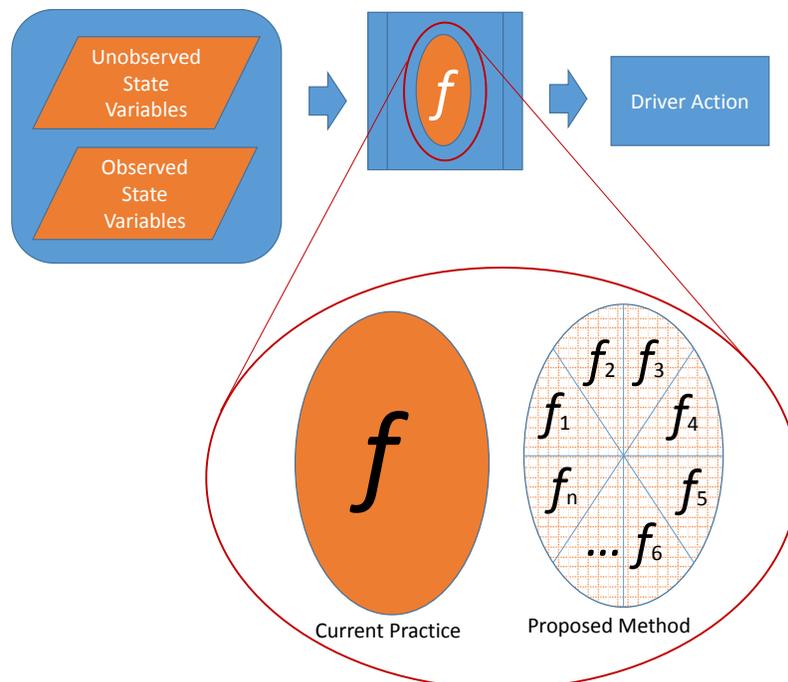


Figure 3-1: Relationship between Driving States and Driving Actions

The proposed methodology created this new structure through two main steps: segmentation and clustering of car-following data. Segmentation divided each car-following period into segments of similar driving states and actions. The same states and actions can repeat

over time, thus clustering found and clustered the repeated segments. The results were clusters that contain similar sets of driving states and corresponding actions. This will potentially reduce the modeling error due to the unknown state parameters, albeit without identifying what these unknown parameters were.

There were three objectives of this research effort: (1) to find a prescriptive set of state-action clusters that can be used to characterize possible driving patterns of car and truck drivers, (2) to compare the different patterns associated with multiple drivers, and examine the similarities and differences between different drivers, and (3) to show that the performance of car-following models can be improved by considering the proposed segmentation and clustering technique. The key to the success of this effort was that the dataset includes a wide variety of driving conditions such that most of the potential driving states, known and unknown, are captured in the data.

In the following sections, we first present a synthesis of past research that summarized previous work that influenced and guided this effort. Next, we describe in detail our methodology, which included a synopsis of the Naturalistic Driving Data that were previously collected by Virginia Tech Transportation Institute (VTTI), a description of the method used to extract car-following periods from the dataset, the segmentation equation, and the clustering method. Lastly, we present the implications of our findings and our conclusions.

3.2 Synthesis of Past Efforts

The past efforts that influenced this work are divided into three categories: (1) car-following data collection methods, (2) modeling car-following behaviors, and (3) data clustering. The data collection methods focus on the strengths and weaknesses of various data collection methods. The modeling of car-following behaviors focuses on the numerous car-following models that have been developed and improved over the years. Finally, data clustering focuses on the techniques employed in clustering data that have been applied in multiple fields.

3.2.1 Car-Following Data Collection Methods

Car-following studies typically collect vehicle trajectory data through various means, including Naturalistic, simulator, and video data collection methods.

Table 3-1 highlights the strengths and weaknesses of each data collection method, which ultimately affect the accuracy of the modeling results obtained through calibration of car-following models. For example, naturalistic and simulator data are useful for observing the range of behaviors of one driver, but fixed camera data would be useful for observing multiple drivers in the same traffic stream.

Table 3-1: Types of Vehicle Trajectory Data Collection

Data Type	Description	Strengths	Weaknesses
Naturalistic [2]	-An instrumented vehicle is driven in normal driving routines	-Driver is in natural environment -Multiple trajectories are observed for each driver	-Drivers know they are being observed
Simulator [3, 4]	-Drivers drive in a simulated environment	-Driving environment in controlled -Multiple trajectories are observed for each driver	-Drivers know they are being observed -Drivers might perform differently in a simulated environment
Video [5, 6]	-A video camera is used to collect data within a certain area	-Drivers do not know they are being observed	-Observation period and area are fixed -Only one trajectory is observed per vehicle

3.2.2 Car-Following Models

Car-following models are designed to process various stimuli, such as the headway distance between vehicles, and produce a response or action, such as the driver decelerating to maintain a certain following distance. The main categories of car following models are: action-point or psychophysical models, linear models, non-linear models, and combination models. Action-point or psychophysical models divide car-following periods into different regimes that represent the drivers intended action. For example, a driver intending to follow a lead vehicle will accelerate and decelerate in an attempt to maintain a desired following distance. Linear and non-linear car-following models are very similar in that they mainly process the headway and difference in speed between the lead and following vehicles with some calibration parameters to create the action of the subject vehicle usually in terms of acceleration. Combination models are the car-following models that combine characteristics of both psychophysical and non-linear or linear models. TABLE 3-2 highlights the strengths and weaknesses of each data collection method

Table 3-2: Synthesis of Car-Following Models

Model Type	Example Model	Relevant Efforts
Action-Point or Psychophysical Models	Wiedemann [7] (used in VISSIM [8]), and Fritzsche [9] (used in PARAMICS [10])	-Evaluating the thresholds through calibration or comparison to the action points [11, 12].
Linear and Non-Linear Models	- Gazis-Herman-Rothery (GHR) model [13], Intelligent Driver Model (IDM) [14], and the Velocity Difference model (VDIFF) [15].	-Inter- and intra-driver variability[16] -Parameter calibration and parameter correlation [17] [18]
Combination Models	hybrid car-following models [19]	-Using specific calibrated GHR parameters to define acceleration behavior within psychophysical models

3.2.3 Data Clustering

The research into clustering data is divided into two main groups: univariate clustering and multivariate clustering. There are many facets within each group, but for the application to modeling car-following behavior, the facet of time-series clustering is the most important. Most

of the research into clustering has occurred outside of the field of transportation engineering. Table 3-3 presents a list of the various time-series clustering techniques covered in the book by Aggarwal and Reddy [20].

Table 3-3: Time-Series Clustering Methods

Method	Metrics Used	Relevant Efforts
Univariate Clustering	-Lp Distance -Dynamic Time-Warping Distance -EDIT Distance -Longest Common Subsequence	-methodology for clustering time-series data [21]
Multivariate Clustering	-Multidimensional Lp Distance -Multidimensional Dynamic Time-Warping Distance -Multidimensional EDIT Distance -Multidimensional Longest Common Subsequence -Multidimensional Subsequence Matching	-Reduction of various segments of the data to statistical measures which could then be clustered by normal k-means methods [22]. -Exploration of behavior heterogeneity of three drivers [23]

3.3 Methodology

The methodology used in this study consisted of a segmentation and clustering technique that was used to capture the full range of state-action clusters for ten car and truck drivers. First, we present a brief overview of the Naturalistic Database used in this methodology, followed by the method used to extract car-following periods from the exorbitant amount of data. Next, we present the segmentation method followed by the clustering method. Lastly, we present the calibration of a car-following model to the data.

A car-following period is defined as a period during which a subject vehicle is reacting to a lead vehicle in the same direction of travel. Each car-following period consists of multiple regimes which, in reference to psychophysical car-following models, are separated by the drivers intended action. For example, the Wiedemann model includes an approaching regime where the subject driver slows down to fall in line with the lead vehicle and a following regime where the subject driver accelerates or decelerates in order to maintain a safe following distance. Each regime consists of two main parts: state and action. The behavior is the link between the state and action that the driver expresses (accelerate, decelerate, and turning). The segmentation process included in this paper divided each car-following period into multiple segments and each segment was defined as a regime. The clustering process grouped similar segments into a single cluster accounting for the fact that regimes could repeat. Thus, a cluster was defined as the grouping of a certain behavior. The segmentation and clustering process is illustrated in Figure 3-2 for a sample car-following period.

In the figure, three of the eight variables are shown to keep the illustration simple. The middle graphs show the results of the segmentation step where segments of similar data have been formed. These segments are then processed and clustered in the clustering step to create the bottom graphs. The clustering step is more constricting than the segmentation step, thus some adjacent segments are placed in the same cluster. The algorithms for segmentation and clustering are data-driven with the maximum number of segments and clusters. This segmentation and clustering

technique was designed to not be unique to the dataset of this study explained in the next subsection.

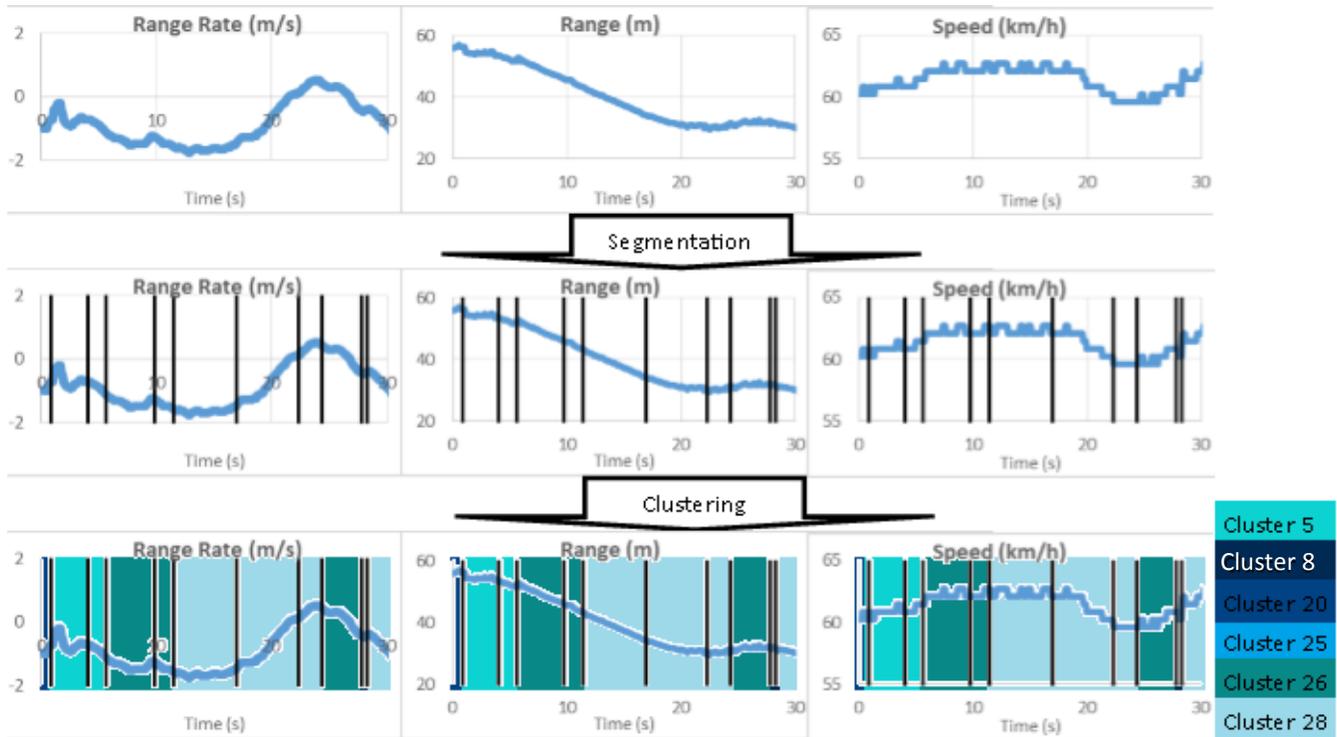


Figure 3-2: Conceptual Representation of the Algorithm Design

3.3.1 Naturalistic Driving Data

Naturalistic driving data refers to data collected from drivers in their natural environment. This was accomplished by equipping vehicles with specialized sensors and “vehicle network” recording equipment and then allowing the participants to drive the vehicles as they see fit. The equipment records a large number of variables (e.g., speed, acceleration, steering wheel positions), and, of specific interest to car-following, a radar positioned at the front of the vehicle records the differences in position and speed between the subject vehicle and the lead vehicle. The equipment also included cameras that recorded: what the driver saw from the front, what the driver saw from the two side mirrors, the drivers’ face, and what the driver was doing inside the vehicle. The naturalistic data used in this research effort were collected by VTTI [2].

TABLE 3-4 shows the naturalistic driving data variables and their definitions. This data were collected from nine trucks and 100 cars that were used by multiple drivers and was recorded at a rate of 10-hertz or 10 samples per second.

Table 3-4: Naturalistic Data Car-following Variables

Variable	Definition
Radar Target ID	Unique identification number assigned to each lead vehicle
Longitudinal Acceleration (g)	Acceleration of the subject vehicle along the roadway
Lateral Acceleration (g)	Acceleration of the subject vehicle across the roadway
Yaw Angle (rad)	Angle between the vehicle heading and the center of the roadway
Yaw Rate (rad/s)	Rate of the change in the yaw angle
Vehicle Speed (km/h)	The speed of the subject vehicle
Lane Offset (in)	The distance between the center of the vehicle and the center of the lane of travel
Range (ft)	The distance from the front of the subject vehicle to the back of the lead vehicle
Range Rate (ft/s)	The rate of change in the range or the difference in speed between the subject vehicle and the lead vehicle

3.3.2 Extraction of Car-Following Periods

Car-following periods were extracted automatically according to specific conditions that fall into two categories. The first set of conditions served as filters that extracted data relevant to car-following periods. The second set of conditions served two purposes: to extract car-following periods that are long enough for analysis and to filter out false-positives. The first set of conditions were the following: there was a lead vehicle, the lead vehicle was within 120 meters, the lead vehicle was in the same lane as the subject vehicle, the roadway curvature was small, and the vehicle speed was greater than 20 kilometers per hour. The limitation on the speed was in place to separate car-following data from traffic jam data because traffic jams present a large number of variables that were not collected in the naturalistic data. The limitation on the roadway curvature was to limit the cases where the roadway in front of the subject vehicle was outside of the cone from which the radar collects data. The second set of conditions were the following: while the range was less than 61 meters, the length of the car-following period was greater than 30 seconds. Multiple lengths of time were tested to define these conditions, but 30 seconds showed the highest positive rate (98%) as compared to lower lengths of time like 25 seconds (80%).

The automatic extraction process was verified from a sample of events through video analysis. For the random sample of 400 periods, 392 were valid car-following periods. The eight periods deemed to not be classified as car-following periods still represented situations where the driver would react to stimuli from another vehicle. For example, a vehicle would merge into the roadway ahead of the subject vehicle and while the lead vehicle was trying to accelerate up to speed, the subject vehicle's driver showed no reaction to the lead vehicle even though the distance to the lead vehicle was within normal car-following ranges. The automatic extraction process was then used to extract car-following periods from 10 car drivers and 10 truck drivers resulting in over 3000 total car-following periods used in the analysis of this paper.

3.3.3 Segmentation Optimization Equation

Time-series segmentation was accomplished by the use of segment lengths and segment centroids. Segment lengths were variables that define the length in time of each segment. The segment lengths were optimized such that the variance within each segment was minimized. The segment centroid was defined as the average of the data points contained within the segment. In this case the centroid had eight dimensions including: average Longitudinal Acceleration, average Lateral Acceleration, average Yaw Rate, average Vehicle Speed, average Lane Offset, average Yaw Angle, average Range, and average Range Rate. These eight dimensions were chosen because

they captured the driver's state and the corresponding actions. The Vehicle Speed, Lane Offset, Yaw Angle, Range, and Range Rate represented the state variables. The lateral acceleration and yaw rate represented the driver action of steering. The longitudinal acceleration represented the driver actions of accelerating or braking.

Each centroid consisted of multiple variables and thus the distance of each variable of each data point to the segment centroid was considered. This was accomplished by converting each of the distances to a normal Z-scale. The equation converting each variable to a Z-scale is shown in (1), accounting for the standard deviation as shown in (2). The Z-scale converts all the variables to the same scale and puts each data point on terms of deviation from the mean which allowed this equation to be used to minimize each data point's deviation from the mean on multiple variables.

$$\text{Min } Z = \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n \frac{|x_{ijk} - \bar{x}_{ij}|}{s_{ij}} \quad (1)$$

Subject to:

$$\begin{aligned} t_i &= [0, T] \forall i \\ \sum t_i &= T \end{aligned}$$

Where:

x_{ijk} is the k th observation of variable j in segment i

\bar{x}_{ij} is the centroid of segment i for variable j

l is the number of segments in a car-following period

m is the number of variables

n is the number of observations in segment i

t_i is the length, in time, of segment i

T is the total length, in time, of a car-following period

$$s_{ij} = \sqrt{\frac{\sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2}{n-1}} \quad (2)$$

This optimization equation was coded into MATLAB [24] and the segment lengths were optimized using a Metropolis [25] algorithm with an initial condition of three second segments.

3.3.4 Clustering

The second part of this methodology clustered the segments, found by the segmentation algorithm, in order to find similar segments or behaviors in the car-following data. Clustering the segments shows how certain behaviors repeat throughout the data within and between drivers. K-means clustering is used in this step with the segment centroids as summary statistics that represented each segment. The statistical software, JMP [26], was used for this step, due to the fact that JMP has a set of tools and techniques for analysis that go beyond simple clustering.

The sensitivity of the number of clusters was analyzed by clustering the segments into different numbers of clusters and obtaining the sum of squared distance from the cluster means to the cluster data points. Figure 3-3 shows that using more than 30 clusters had little reward in terms of the sum of squared distance, thus, 30 clusters were used to represent the behaviors of drivers.

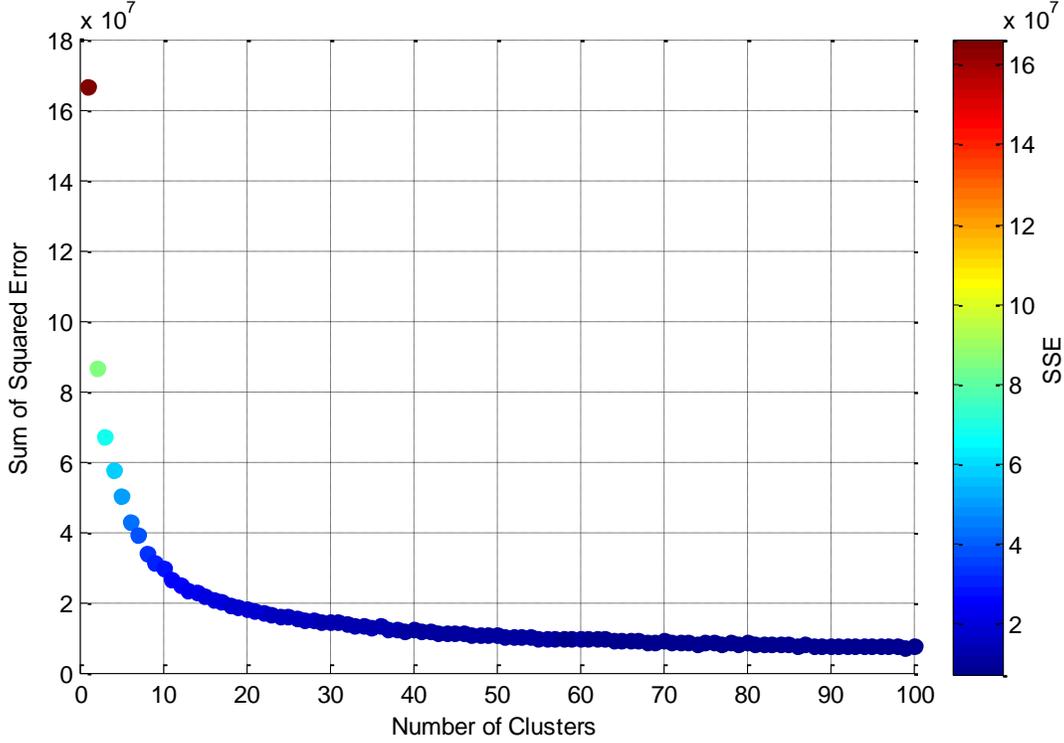


Figure 3-3: Sum of Squared Error for Different Numbers of Clusters

3.3.5 Optimization of Car-Following model

In addition to the clustering effort in this paper, we attempted to calibrate an example car-following model for each cluster. This later task was intended to facilitate the implementation of the research results without impacting the validity of cluster determination or accuracy. The car-following model chosen for use in this paper was the GHR model, shown in (3). The GHR model relates the acceleration to the current speed, relative speed, and space headway. With numerous clusters, the resulting number of calibration parameters was still 120 parameters (4 model parameters times 30 clusters) for car drivers and 120 parameters for truck drivers. The use of more complicated models would potentially decrease the error at the sacrifice of additional complexity.

$$\mathbf{a}_n(t) = c\mathbf{v}_n^m(t) \frac{\Delta v(t-\tau)}{\Delta x^l(t-\tau)} \quad (3)$$

$a_n(t)$ is the acceleration of the subject vehicle at time t

$v_n(t)$ is the speed of the subject vehicle at time t

τ is the perception reaction time of the driver

$\Delta v(t - \tau)$ is the relative speed at time t minus τ

$\Delta x(t - \tau)$ is the space headway at time t minus τ

n is the vehicle index of the follower vehicle

c, l, m are model parameters

The GHR model was used in this paper in a manner that was similar to the hybrid Wiedemann-GHR car-following model [19]. In the original hybrid model, the thresholds were defined by the Wiedemann model, but in this paper the thresholds were found using the segmentation and clustering technique discussed earlier. The GHR model was used in this paper

for two purposes. The first was to express the transfer function between states and actions, and the second was to show that the GHR model could improve in performance through the application of a segmentation and clustering methodology.

A genetic algorithm was used to minimize the Root Mean Squared Error (RMSE) between the GHR output and the observed data. The optimal parameters are provided in this paper for reference purposes. The data from ten car drivers and ten truck drivers were used to highlight similarities and differences between the drivers and between car and truck drivers. Car and truck drivers were analyzed separately due to the fact that cars and trucks have extremely different vehicle characteristics and dynamics.

3.4 Results

3.4.1 Segment Duration Analysis

An analysis of the duration of the clusters showed that a large majority of the clusters last more than a minute. The rest of the clusters were at least five seconds long which was equivalent to 50 data points. This meant that all of the clusters contained adequate information for a behavioral analysis of each cluster.

3.4.2 Multi-resolution Comparison of Error

The validation of the methodology consisted of a cross-comparison of the root mean squared error of the GHR car-following model at different levels of inquiry. The levels of inquiry were as follows: (1) ten car drivers, (2) one car driver, (3) one car-following period, and (4) one segmented and clustered car-following period. The 2nd, 3rd, and 4th levels were optimized to multiple drivers and multiple car-following periods in order to find the average error. For each level, the GHR model was optimized to find the minimum root mean squared error for the difference in speed between the model and the data. For the last level, each cluster was optimized in order to find parameters that apply to each cluster individually.

Figure 3-4 shows that using the GHR car-following model with a segmented car-following period drastically reduced the error as compared to the other three levels of inquiry. It is interesting to note that the error bar for the optimizations to one car-following period is the widest while the error bar for the optimizations to a segmented period is the narrowest indicating more consistency in the optimizations to segmented car-following periods.

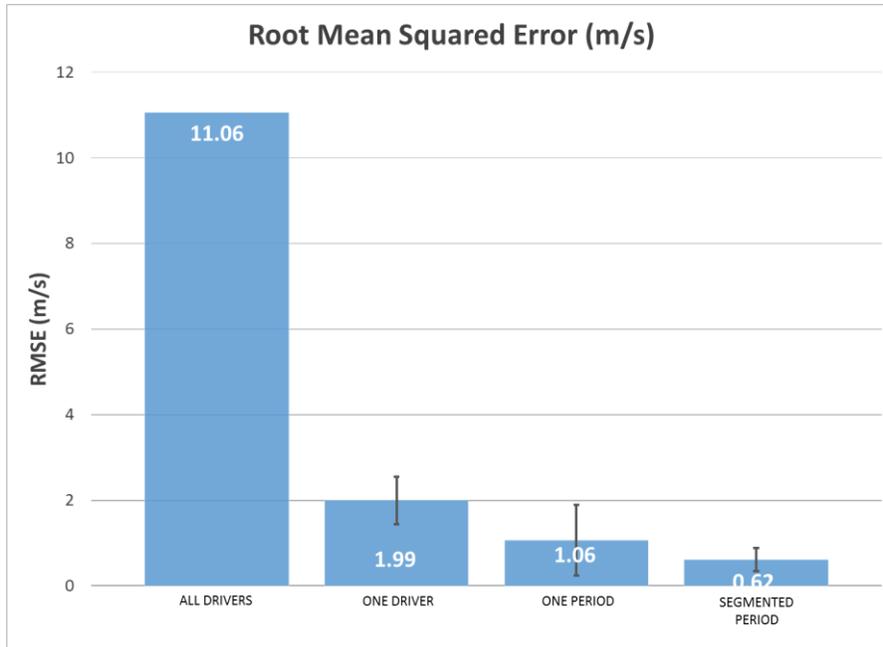


Figure 3-4: Comparison of Error for Different Levels of Calibrated Parameters of the GHR Model

3.4.3 Car Drivers

The results of segmentation and clustering for car drivers revealed some very interesting inclinations. Figure 3-5 represents the cluster distributions or “driving patterns” of each driver. Each division on the circle represents a specific cluster (1-30). The colors in the figure represent the number of occurrence of each cluster by driver. The number of occurrences of a cluster is divided by the maximum occurrences per driver in order to standardize the scale into a 0 to 1 scale as shown. For example, Cluster 1 for Driver J has 6 occurrences, but Cluster 10 has the maximum number of occurrences of 1222, thus giving Cluster 1 a 0.005 (6/1222) and Cluster 10 a 1 (1222/1222). This calculation protects against bias caused by long datasets or drivers with longer car-following periods. The different rings in Figure 3-5 represent the different drivers as shown on the ring labels (A-J) with the first driver (A) being the smallest ring and the last driver (J) being the largest ring.

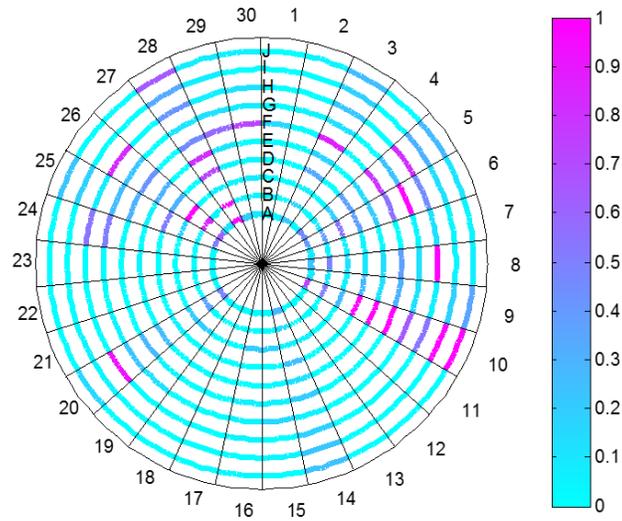


Figure 3-5: Circular Plot of Cluster Frequencies for Car Drivers (A-J)

Figure 3-5 shows that the behaviors of car drivers were scattered with some areas of commonality. Cluster 10 was the most prominent area of commonality, but most of the drivers show behaviors that are manifested in multiple clusters. For example, Driver F showed high frequencies in four different clusters (3, 5, 10, and 30). This also highlights that each driver had a specific combination of high frequency clusters that did not repeat between car drivers.

Table 3-5 shows that each cluster has a very high standard deviation in reference to segment lengths. This indicated that some unknown factors were causing drivers to switch between the clusters at various time steps and these factors were not correlated to temporal durations.

Table 3-5 also shows a number of clusters that did not occur very often like clusters 11, 22, and 18, which were investigated. The findings were that the clustering method produced clusters of data errors. Both clusters 11 and 22 had data errors where one sensor produced data that conflicted with another sensor. For example, in cluster 11, the data showed high lateral acceleration indicating that the vehicle was making a very sharp turn, but a very low yaw rate indicated that the vehicle was continuing to move forward. Cluster 18 showed a very hard turn and brake suggesting conflict avoidance behavior. The clustering method appeared to be successful in not only grouping car-following behaviors, but also in extracting and grouping data errors that occur over all ten car drivers. The clusters containing data errors were removed from further analysis. Figure 3-6 shows an example segmented and clustered car-following period in the same format as the Wiedemann model. The right side of the graph represents differences in speed, between the lead and following vehicles, which result in the following vehicle getting closer to the lead vehicle while the left side of the graph represents cases where the headway is increasing.

Table 3-5: Cluster Occurrences, Average Lengths, and Standard Deviations for Car Drivers (A-J)

Cluster	Count	Average Cluster Length (s)	Standard Deviation of Cluster Length (s)
1	24	16.59	11.67
2	38	85.13	256.95
3	787	26.63	22.67
4	10	12.10	10.30
5	483	86.64	98.67
6	1186	70.59	80.22
7	179	26.45	37.40
8	644	67.06	85.65
9	490	42.83	49.86
10	1644	68.63	77.09
11	1	15.02	0.00
12	14	19.61	19.85
13	4	338.47	369.58
14	734	23.68	19.51
15	111	65.46	123.72
16	159	35.93	38.57
17	5	11.84	5.03
18	2	33.26	3.78
19	205	42.68	58.89
20	371	107.35	146.66
21	15	39.49	49.53
22	1	5.65	0.00
23	6	271.63	386.02
24	4	3194.95	2016.44
25	636	49.56	56.71
26	941	105.87	145.97
27	14	13.52	10.44
28	718	99.73	165.00
29	327	39.48	40.73
30	178	36.21	95.20

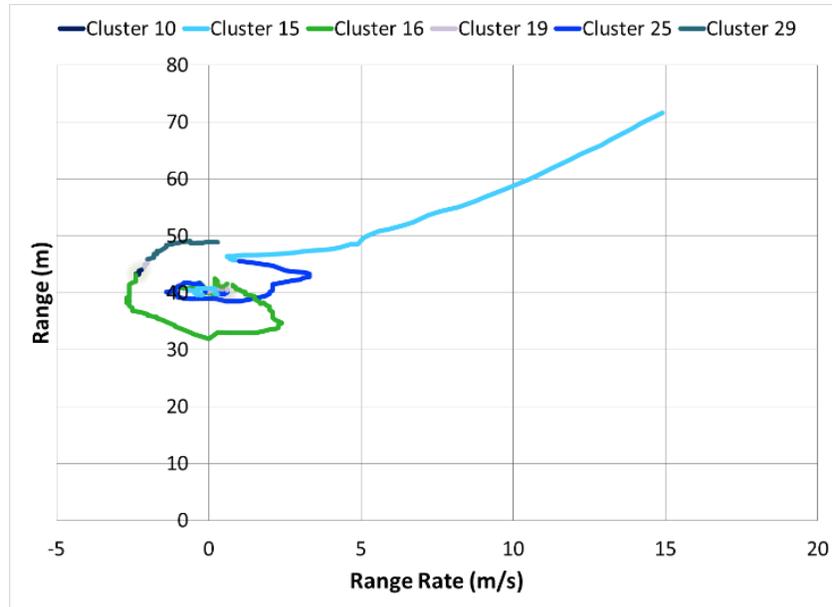


Figure 3-6: Example Segmented and Clustered Car-following Period

The first cluster, 15 (light blue), shows that the following vehicle was catching up to the lead vehicle and decelerated to match the speed of the lead vehicle. The behavior shifted to cluster 25 (red) as the lead vehicle decelerated and then the following vehicle decelerated after a small delay. The car-following period then went into a few loops around the range of 40 meters where both the lead and following vehicles were accelerating. The behavior shifts out of the loop into cluster 16 (pink) as the lead vehicle stopped accelerating, but the following vehicle continued accelerating. The car-following period then shifted between clusters 10, 19, and 29 as the following vehicle was attempting to pass the lead vehicle.

The changes of behavior in this example were rapid and mostly stem from different actions that the driver took like decelerating and accelerating. This revealed that each cluster contains a specific action of the driver. The example also showed that some of the clusters shifts were due to a changing environment like the lead vehicle decelerating. Gathering all of this together elucidated the view of the clusters as actions in certain environments. The driver behavior was the accumulation of all of the driver's actions.

Table 3-6 presents the optimized GHR parameters for each cluster that was shown in Figure 3-5. The parameters were varied which supports that each cluster represented different behaviors. Some of the perception-reaction times were very small which could have been the result of the driver correctly anticipating the actions of the lead vehicle. Also, note that these parameters were optimized to a specific regime of behavior and each set of parameters did not represent the behaviors of an entire car-following period. All of the clusters showed RMSE values that were far below those shown in Figure 3-4. This indicates a large reduction in the modeling error that could be attributed to the segmentation and clustering technique.

Table 3-6: Calibrated GHR Parameters for Each Cluster for Car Drivers

Cluster	c	m	l	τ (s)	RMSE (m/s)
1	60.07	6.40	7.95	2.95	0.06
2	-42.41	4.00	5.73	3.00	0.69
3	-70.01	-2.39	0.43	3.00	0.78
4	-41.66	-5.25	-2.15	0.20	0.16
5	-4.35	2.34	3.11	3.00	1.13
6	-3.54	0.00	1.00	0.11	1.38
7	-58.81	-6.08	-1.76	2.87	0.74
8	-8.26	-0.24	1.10	1.68	0.61
9	-10.97	0.00	1.00	0.01	1.21
10	-90.72	-3.05	1.00	2.94	1.70
12	-27.02	10.00	9.41	2.91	0.20
13	-0.07	0.52	-0.14	0.00	0.71
14	-22.28	-0.67	0.83	0.06	0.94
15	-12.96	-1.15	0.34	0.62	0.74
16	-27.86	-0.65	1.08	0.01	0.79
17	-18.54	9.18	-3.97	1.93	0.04
19	-31.99	-0.43	1.24	0.53	1.07
20	-72.83	2.86	4.16	3.00	0.63
21	35.27	8.03	10.00	3.00	0.26
23	-21.70	-4.34	-1.85	2.75	0.13
24	38.90	-2.04	1.53	1.31	0.43
25	-1.77	0.51	1.10	1.40	0.80
26	-49.54	-1.70	1.00	2.17	0.95
27	23.76	2.04	7.26	2.58	0.01
28	-94.66	-0.55	1.37	0.08	0.86
29	-72.12	-1.25	0.94	1.28	0.73
30	-16.03	0.06	2.03	1.64	0.79

Figure 3-7 shows a time-space plot of hypothetical car-following trajectories using some of the resulting clusters.

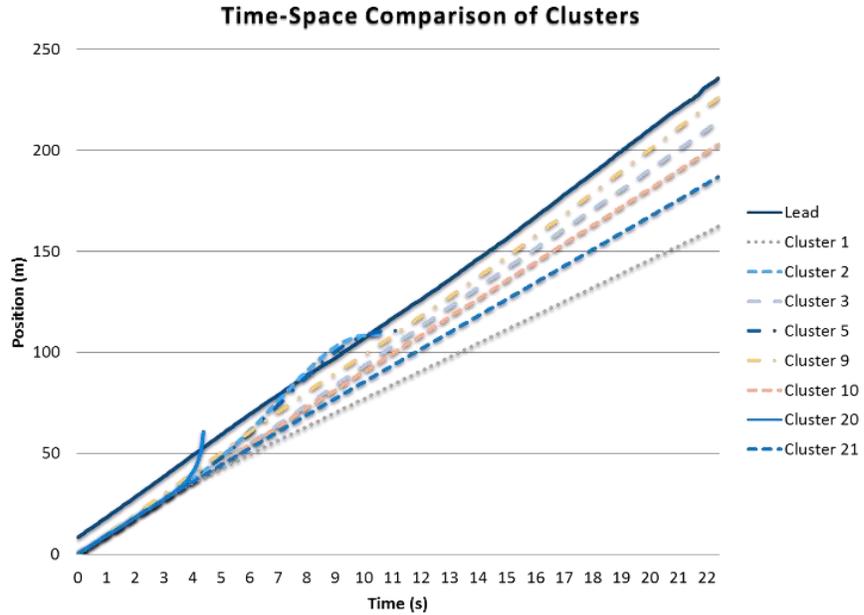


Figure 3-7: Time-Space Comparison of Clusters

The data consisted of a full car-following period with only 22 seconds shown in the graph to help highlight the differences between the clusters. The difference between the clusters was very small at first, less than 0.3 meters, but after a few seconds the differences became visually noticeable. Some clusters ended early due to passing the lead vehicle as is the case for clusters 20 and 11. Some clusters maintained a constant following distance while others show increasing following distance.

3.4.4 Truck Drivers

Figure 3-8 shows that unlike the car drivers, most of the truck drivers showed a specific combination of high frequency clusters, 14 and 19, which repeated between most of the drivers. These two clusters also appeared to occur in equal frequency which indicated a dual nature in the behavior of truck drivers.

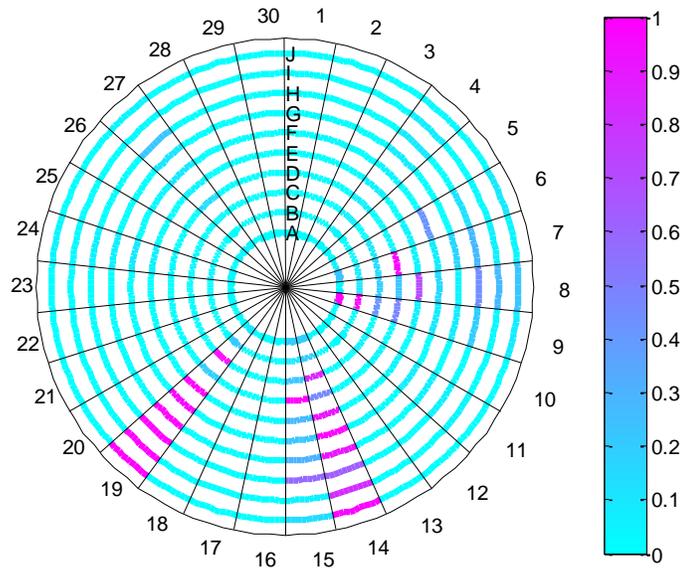


Figure 3-8: Circular Plot of Cluster Frequencies for Truck Drivers (A-J)

Table 3-7 shows that the clusters for truck drivers also had large standard deviations in reference to segment lengths. Also, there were a high number of low occurrence clusters which indicated that truck driver behavior is not as diverse as car driver behavior and thus using thirty clusters for truck drivers revealed more artefacts.

Table 3-7: Cluster Occurrences, Average Lengths, and Standard Deviations for Truck Drivers (A-J)

Cluster	Count	Average Cluster Length (s)	Standard Deviation of Cluster Length (s)
1	7	48.56	21.13
2	53	30.74	36.53
3	33	30.39	26.58
4	447	14.67	15.21
5	4	8.00	2.94
6	635	63.44	86.40
7	325	137.52	190.28
8	978	64.09	69.32
9	1260	88.58	128.13
10	7	16.26	14.32
11	326	28.70	17.64
12	2	7.50	2.12
13	97	42.12	46.19
14	1730	113.61	119.28
15	613	126.38	129.56
16	1	2.00	0.00
17	3	10.67	4.51
18	2	1.00	0.00
19	2170	105.21	117.50
20	3	7.00	4.36
21	1	6.00	0.00
22	5	12.40	13.58
23	22	31.47	23.06
24	391	30.34	20.74
25	3	13.00	3.46
26	56	42.61	41.36
27	320	52.60	45.21
28	2	5.50	2.12
29	5	9.47	7.25
30	1	3.00	0.00

Table 3-8 presents the calibrated GHR parameters for the truck drivers. This table also shows varied parameters which support that each cluster is representative of different behaviors.

Table 3-8: Calibrated GHR Parameters for Each Cluster for Truck Drivers

Cluster	c	m	l	τ (s)	RMSE (m/s)
1	-68.84	-2.33	-0.25	2.86	0.34
2	-53.46	-0.87	1.07	3.00	0.50
3	0.24	1.54	1.19	3.00	0.32
4	61.94	-0.18	1.86	3.00	0.34
5	0.17	0.94	1.58	0.01	0.13
6	-7.84	-0.59	0.91	3.00	0.64
7	-2.95	-0.07	0.74	0.01	1.31
8	-0.22	0.79	1.29	3.00	0.75
9	-1.28	1.17	2.30	1.86	0.63
10	3.32	4.61	5.09	2.16	0.07
11	-8.55	-8.53	-0.81	2.94	0.23
13	-2.99	1.03	2.04	3.00	0.65
14	-44.50	-5.01	-1.83	2.81	0.48
15	-5.07	-0.56	0.51	1.31	1.26
17	25.94	-14.02	-4.12	0.53	0.18
19	-5.93	0.37	1.79	1.48	0.77
20	0.00	-2.88	-3.53	0.00	1.27
22	4.99	-0.10	1.17	1.23	0.55
23	1.11	-8.77	-3.83	0.15	0.71
24	-4.93	-0.39	1.25	2.55	0.52
25	0.01	1.26	0.42	0.34	1.91
26	-2.06	0.77	1.53	3.00	0.83
27	-1.07	0.91	1.01	0.73	1.82
29	0.00	2.11	-0.87	0.02	1.48

3.5 Discussion and Implications

The methodology included in this paper revealed some very interesting characteristics of driver behavior. The results showed heterogeneity among car drivers and homogeneity among truck drivers. The implications behind those statements are that car drivers should be modeled independently and truck drivers can be modeled as a group. The homogeneity among truck drivers could be due in part to the very similar operating conditions for trucks and the fact that truck drivers are trained drivers. The heterogeneity among car drivers could be due to different vehicle characteristics, but these drivers will continue to drive their own vehicles which means that the heterogeneous behaviors will still be present in their driving under normal conditions. For truck drivers, consideration of two clusters captured most of the behavior, but there were other clusters that need consideration. While a behavior or cluster might appear sporadically, that cluster can represent cases where the driver posed a threat to others or disrupted the flow of traffic.

In contrast to truck drivers, car drivers showed much more complicated behavior that consisted of a distribution of multiple clusters for each driver. Proper consideration of this finding would implicate a shift from a global distribution of behavior, as currently used in simulation software, to individual distributions of behavior. The fact that car drivers showed the same

behaviors or clusters in different frequencies allows the creation of a global set of behaviors that car drivers then represent in varying frequencies.

This study analyzed data from a variety of driving conditions in order to capture the full range of behavior of drivers. Shifting research towards the study of the driver behavior instead of the driving conditions can result in new findings. This way one behavior could be potentially linked to multiple driving conditions and the cross comparison of these driving conditions could reveal the true reasons behind that one behavior. The issue is that this cross comparison would require detailed information from multiple sources in order to properly attribute a behavior to a certain driving condition characteristic.

Another point for discussion from the results of this study, is the sample sizes needed for the study of naturalistic data. There are two main considerations in terms of sample size: the number of samples for each driver and the number of drivers to include. The number of samples needed for each driver should be based on statistical significance analysis. The number of drivers to include should be based on the need to study further patterns in more details. For cars, studying more drivers could potentially result in finding more unique driving patterns. For trucks, however, the results of this study showed homogeneity and thus future research into truck drivers can consist of a small sample of drivers and still accurately capture truck driver behavior. On the other hand, there are many different types of trucks in use with varying characteristics. Thus, future research would be to examine the impacts of different truck characteristics on the distribution of truck driver behavior.

Results of calibration of the GHR model showed that the accuracy of modeling can be improved when the model is calibrated to each cluster. Future research should focus on modeling the relationship between different clusters in each car-following episode, leading to a Markov chain modeling that can result in a higher simulation fidelity. The key for moving this research forward is the inclusion of new information that can be used to decode the meaning of the clusters and eventually find the factors that contribute to certain driving behaviors or patterns.

Drivers are humans and human behavior has multiple influencing factors that are difficult to consider like emotions, personality, medical conditions, hunger, thirst, etc. [1]. These influencing factors can cause a driver to behave differently when the same situation repeats. The approach presented in this paper would theoretically account for unobserved influencing factors. Discovery of the characteristics of the influencing factors will require data from each cluster that includes additional variables. This can be accomplished by matching data from other sources to each cluster or conducting studies that are focused on examining certain clusters. The segmentation and clustering technique is not limited to eight variables nor to naturalistic data, which can be useful in the processing of other datasets. By focusing future research efforts on certain clusters, the influencing factors behind certain behaviors can be found and accounted for, which would lead to drastically more realistic traffic simulations.

3.6 Conclusions

This research effort aimed to accomplish three objectives: (1) to find a prescriptive set of state-action clusters that can be used to characterize possible driving patterns of car and truck drivers, (2) to compare the different patterns associated with multiple drivers, and examine the similarities and differences between different drivers, and (3) to show that the performance of car-following models can be improved by consideration of the proposed segmentation and clustering technique. The methodology used in this effort was a 2-step algorithm that would first segment car-following periods and then cluster these individual segments based upon longitudinal

acceleration, lateral acceleration, yaw rate, speed, lane offset, yaw angle, range, and range rate. The sample for this study was twenty drivers—ten car and ten truck drivers—with over 100 car-following periods each, totaling 3000 car-following periods.

The research findings show that the driving behavior in the examined naturalistic data can be characterized using 30 unique clusters. While we feel 30 clusters are probably representative enough for a wide range of driving behavior (covered by the vast dataset used in this research), using other datasets might reveal that a different number of clusters is appropriate. Car drivers tend to exhibit very specific distributions of behavior (heterogeneity) while truck drivers tend to exhibit a common distribution of behavior (homogeneity). This supports the notion that trucks drivers can be grouped and modeled as a unit (homogeneous), but car drivers cannot be grouped and have to be modeled individually (heterogeneous). Also, truck drivers exhibit two main clusters of behavior that appear in near equal frequency, suggesting that car-following models must take their dual nature into account in order to be accurate. As for car drivers, each driver should be modeled individually in order to be accurate. Future work would be to expand on these findings by investigating cluster sequencing within and between car-following periods.

References

- [1] S. K. Balogun, N. A. Shenge, and S. E. Oladipo, "Psychosocial factors influencing aggressive driving among commercial and private automobile drivers in Lagos metropolis," *The Social Science Journal*, vol. 49, pp. 83-89, March 2012.
- [2] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
- [3] P. S.-T. Chen and H. S. Mahmassani, "Dynamic interactive simulator for studying commuter behavior under real-time traffic information supply strategies," *Transportation Research Record*, pp. 12-21, 1993.
- [4] R. Hoogendoorn, S. Hoogendoorn, K. Brookhuis, and Winnie Daamen, "Mental Workload, Longitudinal Driving Behavior and Adequacy Of Car Following Models In Case of Incidents In The Other Driving Lane," *TRB 2010 Annual Meeting CD-ROM*, p. 21, 2010.
- [5] (2013, February 20, 2014). *Traffic Analysis Tools: Next Generation Simulation (NGSIM)*.
- [6] S. Ossen, S. Hoogendoorn, and B. Gorte, "Interdriver Differences in Car-Following: A Vehicle Trajectory-Based Study," *Transportation Research Record*, vol. 1965, pp. 121-129, 2006.
- [7] R. Wiedemann, "Simulation des Strassenverkehrsflusses," Instituts für Verkehrswesen der Universität Karlsruhe Universität Karlsruhe, 1974.
- [8] M. Fellendorf, "VISSIM: A Microscopic Simulation Tool to Evaluate Actuated Signal Control Including Bus Priority," in *64th Institute of Transportation Engineers Annual Meeting*, Dallas, TX, 1994, pp. 1-9.
- [9] H.-T. Fritzsche, "A Model for Traffic Simulation," *Transportation Engineering Contribution*, vol. 5, pp. 317-321, May 1994.
- [10] G. D. B. Cameron and G. I. D. Duncan, "PARAMICS—Parallel microscopic simulation of road traffic," *The Journal of Supercomputing*, vol. 10, pp. 25-53, March 1 1996.
- [11] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [12] B. Higgs, M. Abbas, A. Medina, and C. Y. D. Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," in *TRB Traffic Flow Theory Summer Meeting and Conference*, 2010.
- [13] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear Follow-The-Leader Models of Traffic Flow," *Operations Research*, vol. 9, pp. 545-567, 1961.
- [14] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 62, pp. 1805-1824, 2000.
- [15] R. Jiang, Q. Wu, and Z. Zhu, "Full Velocity Difference Model for a Car-Following Theory," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, 2001.
- [16] A. Kesting and M. Treiber, "Calibrating Car-Following Models using Trajectory Data: Methodological Study," *Transportation Research Record*, vol. 2088, pp. 148-156, 2008.

- [17] J. Kim and H. Mahmassani, "Correlated Parameters in Driving Behavior Models," *Transportation Research Record*, vol. 2249, pp. 62-77, December 2011.
- [18] V. Punzo, B. Ciuffo, and M. Montanino, "Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted?," *Transportation Research Record*, vol. 2315, pp. 11-24, December 2012.
- [19] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board 91st Annual Meeting, Washington D.C., 2012.
- [20] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. Boca Raton, FL: CRC Press, 2013.
- [21] Z.-H. Fan, D.-K. Huang, J.-Z. Li, and K.-H. Wang, "Morphing Cluster Dynamics: Time series clustering with multipartite graph," in *2004 International Conference on Machine Learning and Cybernetics*, Shanghai, China, 2004, pp. 1749-1754.
- [22] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in *7th IEEE International Conference on Data Mining*, Omaha, NE, United states, 2007, pp. 351-360.
- [23] B. Higgs and M. Abbas, "A Two-Step Segmentation Algorithm for Behavioral Clustering of Naturalistic Driving Styles," presented at the 16th International IEEE Conference on Intelligent Transportation Systems, The Hague, The Netherlands, 2013.
- [24] "MATLAB," 2012b ed. Natick, MA: The MathWorks Inc., 2012.
- [25] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087-1092, 1953.
- [26] "JMP," Pro 10 ed. Cary, NC: SAS Institute Inc., 1989-2013.

4. Identification and Classification of State-Action Clusters of Car-Following Behavior³

Abstract

This research effort aimed to identify and classify state-action clusters of driver behavior. The methodology first segmented and clustered car-following periods into clusters that identify a specific combination of state variables (speed, lane offset, yaw angle, range and range rate) and action variables (longitudinal acceleration, lateral acceleration, and yaw rate). The state-action clusters were then analyzed using discriminant analysis to reveal the clusters that could be identified using: (1) only state variables, (2) only action variables, and (3) both state and action variables. The sample used in this paper included ten different drivers with over 100 car-following periods each, totaling over 1500 car-following periods. In summary, the results revealed that: (1) 60% of the state-action clusters could be identified using only state variables, (2) 30% of the state-action clusters could be identified using only action variables, and 100% of the state-action clusters could be identified using both state and action variables. Also, 20% of the state-action clusters require the use of both state and action variables to be identified.

³ Paper has been submitted for publication in the proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC 2014)

4.1 Introduction

There are many variations of car-following models that have been developed over the years. These models are structured to relate driving states to driving actions. The key issue that most of these models share is that they are structured for a one to one relationship of driving states to driving actions. This structure is sound in theory, but car-following models can only consider a limited number of state variables. Reduction of driving states to a small number of variables can cause different driving states to appear to be the same state. Errors would occur if the different driving states and linked to different actions. For example, a car-following period during the day with heavy rain and a car-following period at night could show similar values on some state variables, and thus be modeled with the same action when the conditions actually resulted in different driver actions. Consideration of all potentially influential state variables would be costly and inefficient, which highlights the need for a solution that can capture the effects of variables that are not included in the modeling process.

The analysis in this paper addresses that need by creating a structure that allows multiple connections to be formed between driving states and driving actions. The methodology proposed in this paper first identified all of the connections between driving states and driving actions through the use of state-action clusters. The methodology then used discriminant analysis to find the clusters that can be identified using: (1) only state variables, (2) only action variables, and (3) both state and action variables.

In the following sections, we first present a synthesis of past effort, then we describe the naturalistic driving data that were previously collected by Virginia Tech Transportation Institute (VTTI), followed by a description of our methodology, and finally we present our findings and conclusion.

4.1.1 Synthesis of Past Efforts

The past efforts that influenced this work are summarized by previous work [1-12] that focused on four main areas of study: (1) car-following data collection methods, (2) comparison of car-following models, (3) development of car-following models, and (4) car-following data segmentation and clustering. Previous research [1, 5, 7, 10, 11, 13] examined naturalistic data that was collected from drivers during their normal daily routines. The comparison of car-following models [7, 10] examined the properties, strengths, and weaknesses of various car-following models. The development of car-following models [1, 3, 4] used the findings of the strengths and weaknesses of existing models to develop new models that would: (1) remove the weaknesses of existing models or (2) combine the strengths of two different models. The findings of both the comparison of car-following models and the development of car-following models showed a large amount of variability within and between drivers that, upon further investigation, revealed that driving behavior changes: (1) between drivers, (2) between car-following periods, and (3) within a car-following period [7]. These findings lead to the development of a car-following data segmentation and clustering method that divides car-following periods into segments of similar behavior [8]. The method then clusters the segments to find unique state-action clusters of behavior. The analysis included in this paper explores these clusters to find their defining characteristics.

4.2 Methodology

The methodology used in this study consisted of three main parts: extraction of car-following periods from the naturalistic database, a segmentation and clustering technique, and an analysis of

the clusters. The method for extraction of car-following periods from the naturalistic database has been described in detail in previous research. The segmentation and clustering technique, from previous research [8], segmented and clustered car-following periods into state-action clusters based upon five state variables (speed, lane offset, yaw angle, range and range rate) and three action variables (longitudinal acceleration, lateral acceleration, and yaw rate). The analysis of the clusters examined the similarities and differences between the state-action clusters based upon the five state variables and three action variables. First, we present a brief overview of the Naturalistic Database used in this methodology, followed by the method used to extract car-following periods from the exorbitant amount of data. Next, we present the segmentation method followed by the clustering method. Lastly, we present the discriminant analysis of the clusters identified by the clustering method.

The clustering step was more constricting than the segmentation step, thus some adjacent segments were placed in the same cluster. The algorithms for segmentation and clustering are data-driven with the maximum number of segments and clusters as the only input. This segmentation and clustering technique was designed to not be unique to the dataset of this study explained in the next subsection.

4.2.1 Naturalistic Driving Data

Naturalistic driving data refers to data collected from drivers in their natural environment. This was accomplished by equipping vehicles with specialized sensors and “vehicle network” recording equipment and then allowing the participants to drive the vehicles as they see fit. The equipment recorded a large number of variables (e.g., speed, acceleration, steering wheel positions), and, of specific interest to car-following, a radar positioned at the front of the vehicle recorded the differences in position and speed between the subject vehicle and the lead vehicle. The equipment also included cameras that recorded: what the driver saw from the front, what the driver saw from the two side mirrors, the drivers’ face, and what the driver was doing inside the vehicle. The naturalistic data used in this research effort were collected by VTTI [14]. This data were collected from nine trucks and 100 cars that were used by multiple drivers and was recorded at a rate of 10-hertz or 10 samples per second.

4.2.2 Extraction of Car-Following Periods

Car-following periods were extracted automatically according to specific conditions that fall into two categories. The first set of conditions served as filters that extracted data relevant to car-following periods. The second set of conditions served two purposes: to extract car-following periods that are long enough for analysis and to filter out false-positives. The automatic extraction process was then used to extract car-following periods from 10 car drivers resulting in over 1500 total car-following periods used in the analysis of this paper.

4.2.3 Segmentation Optimization Equation

Time-series segmentation was accomplished by the use of segment lengths and segment centroids. Segment lengths were variables that define the length in time of each segment. The segment lengths were optimized such that the variance within each segment was minimized. The segment centroid is defined as the average of the data points contained within the segment.

4.2.4 Clustering

The second part of this methodology clustered the segments, found by the segmentation algorithm, in order to find similar segments or behaviors in the car-following data. Clustering the segments showed how certain behaviors repeat throughout the data within and between drivers. K-means clustering was used in this step with the segment centroids as summary statistics that represent each segment. The results of previous research found 30 unique state-action clusters that were used for this study.

4.2.5 Discriminant Analysis

The last part of the methodology analyzed the state-action clusters identified by the previous step using discriminant analysis. The analysis had three parts: (1) discriminant analysis using only state variables, (2) discriminant analysis using only action variables, and (3) discriminant analysis using both state and action variables. The first two parts revealed the overlaps of different driving states and different driving actions respectively. The last part served as a verification of the clustering method and revealed the clusters that required both state and action variables to be identified. The statistical software, JMP [15], was used for this step, due to the fact that JMP has a set of tools and techniques for discriminant analysis.

4.3 Results

Figure 4-1 shows an example segmented and clustered car-following period in the same format as the Wiedemann model. The right side of the graph represents differences in speed, between the lead and following vehicles, which result in the following vehicle getting closer to the lead vehicle while the left side of the graph represents cases where the headway is increasing.

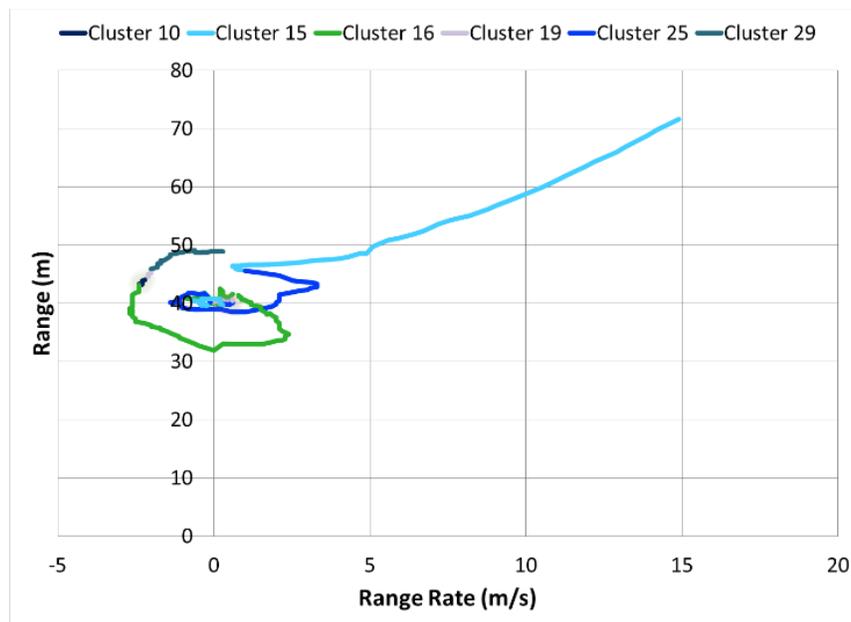


Figure 4-1: Example Segmented and Clustered Car-following Period

The figure shows that the behavior of the subject driver changed for different states and different actions. The first cluster, 15 (light blue), shows that the subject driver was approaching the lead vehicle and decelerated to follow. Cluster 25 (dark blue) showed a situation where the

lead vehicle started to decelerate, thus changing the driving state. The subject driver then began to accelerate along with the lead vehicle maintaining the headway distance (light blue at 40 meters). The behavior shifted to a new driving state (cluster 16 in green) as the lead vehicle stopped accelerating, but the subject vehicle continued accelerating. The car-following period then shifted between various driving states (clusters 10, 19, and 29) as the subject vehicle was attempting to pass the lead vehicle while accelerating.

Figure 4-2 shows that the results of the discriminant analysis revealed that some clusters could be identified using only one variable. The figure also shows that there was a large amount of misclassification of clusters and the misclassification appeared to be dependent upon the variable. For each variable, the clusters that were the focus of the misclassification can be distinguished by a light color that forms a vertical line. This pattern would indicate that multiple clusters were being misclassified into the same cluster. For example, considering yaw rate, the misclassifications were driven towards clusters 8, 12, 16, and 19.

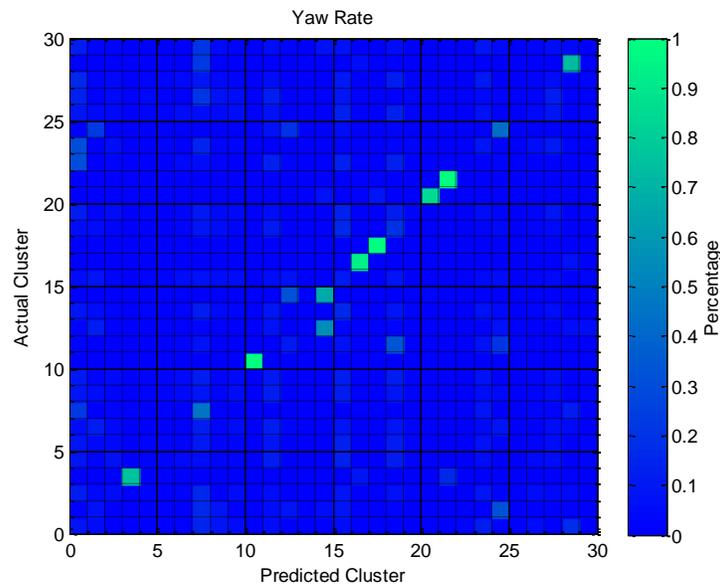


Figure 4-2a: Discriminant Analysis Classification Matrix for Yaw Rate

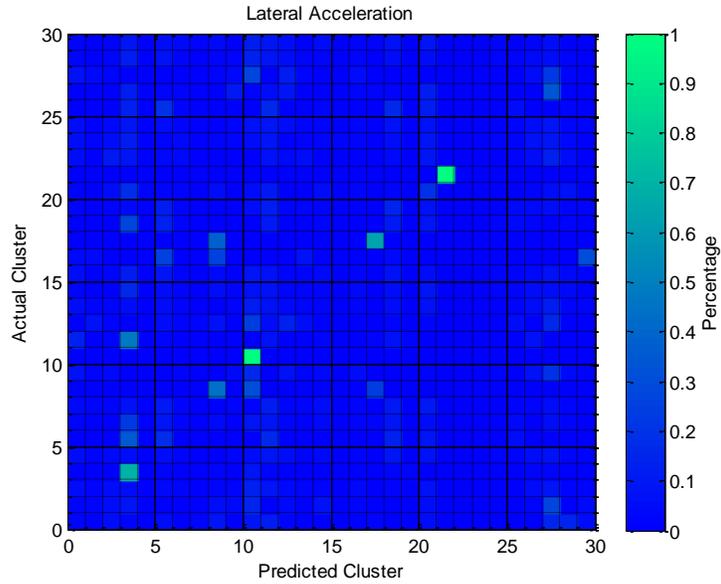


Figure 4-2b: Discriminant Analysis Classification Matrix for Lateral Acceleration

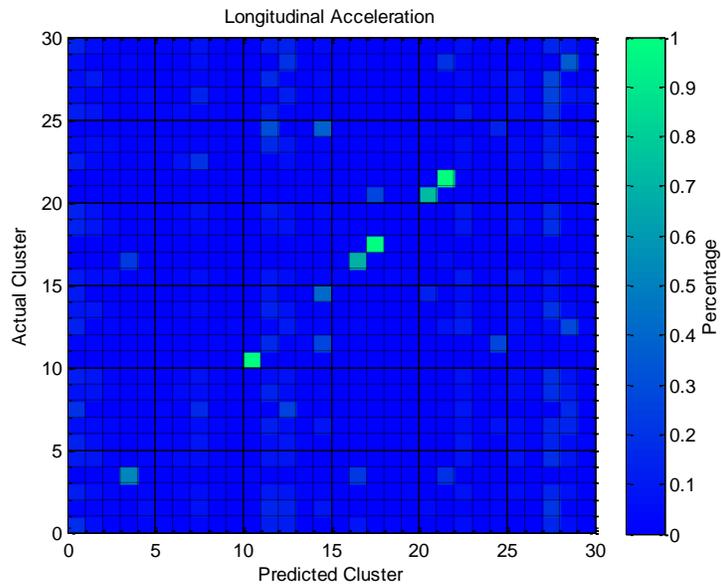


Figure 4-2c: Discriminant Analysis Classification Matrix for Longitudinal Acceleration

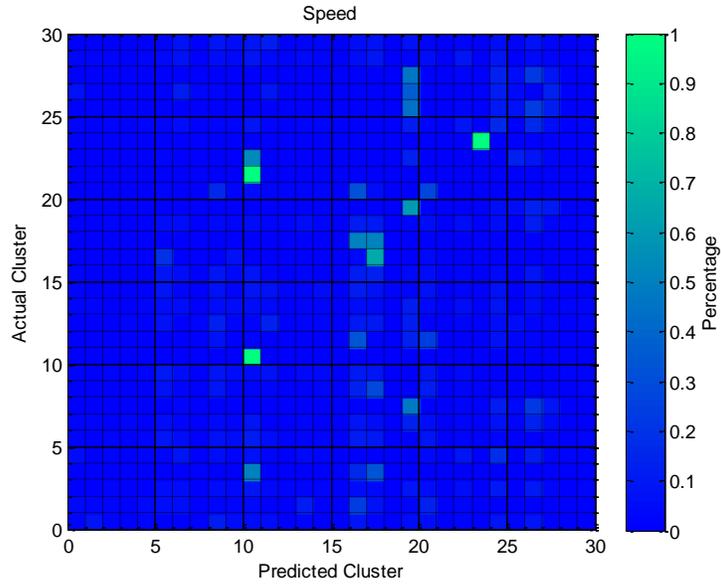


Figure 4-2d: Discriminant Analysis Classification Matrix for Speed

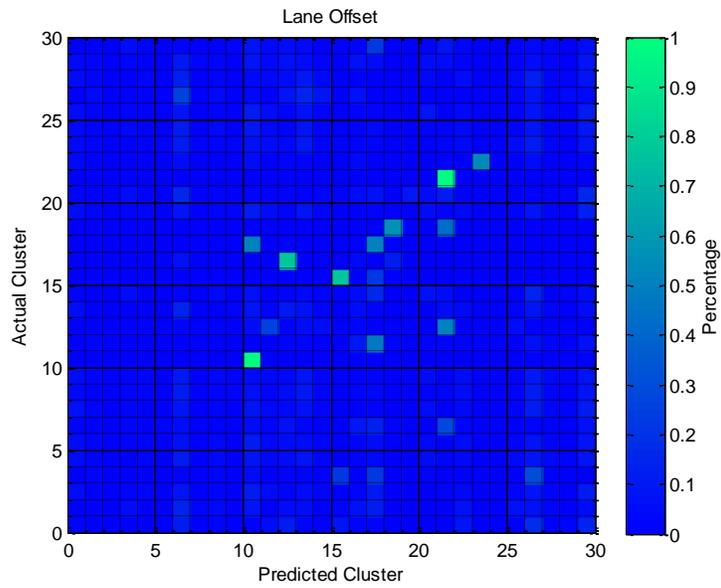


Figure 4-2e: Discriminant Analysis Classification Matrix for Lane Offset

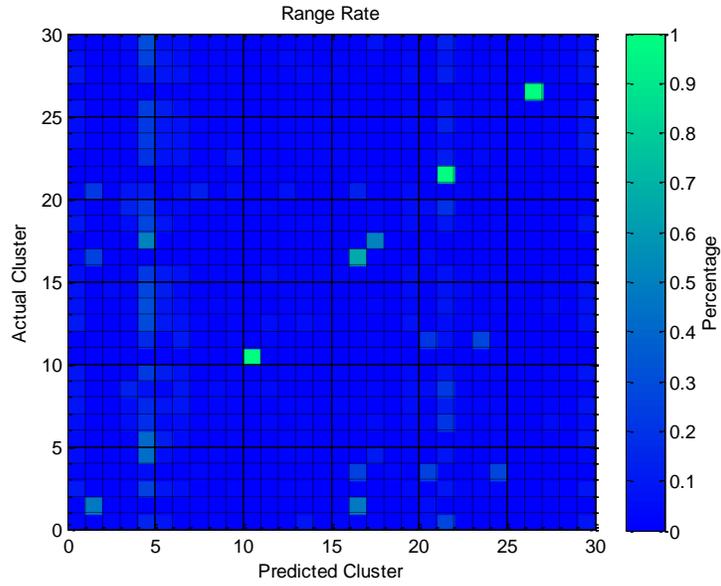


Figure 4-2f: Discriminant Analysis Classification Matrix for Range Rate

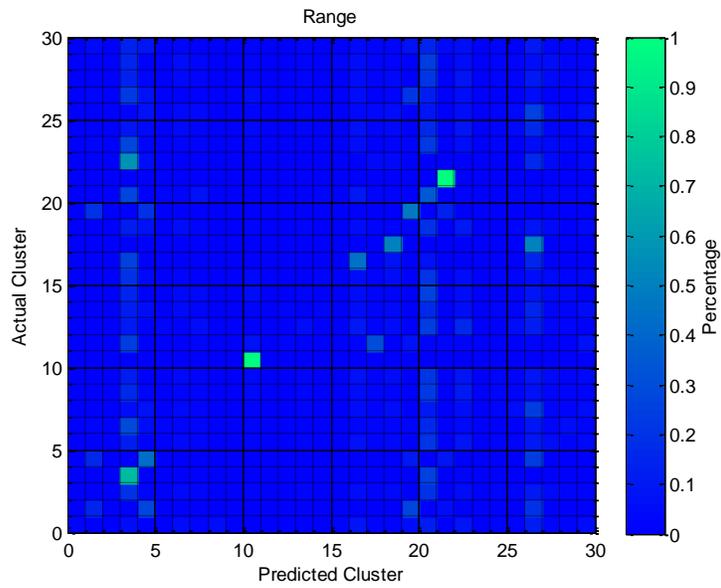


Figure 4-2g: Discriminant Analysis Classification Matrix for Range

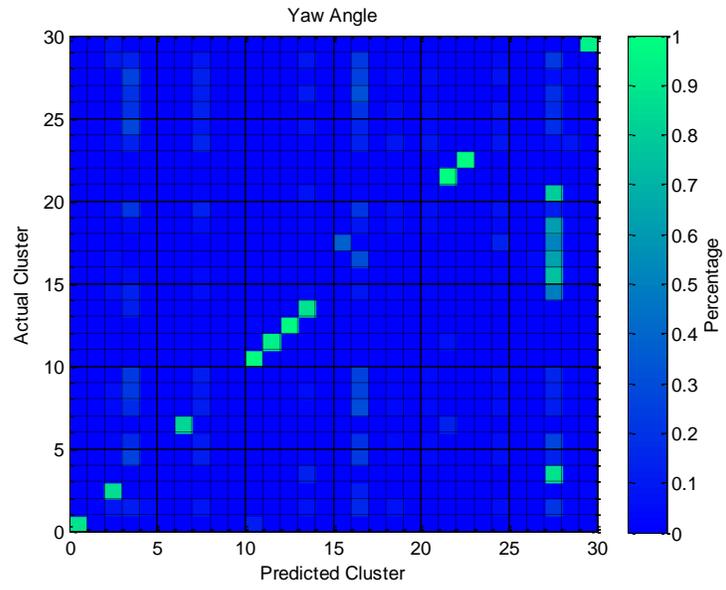


Figure 4-2h: Discriminant Analysis Classification Matrix for Yaw Angle

Table 4-1 shows the results of a discriminant analysis that used: only state variables, only action variables, and both state and action variables. The table shows that all of the clusters could be identified using both state and action variables. Some of the clusters showed values less than one due to the fact that the clusters were created using five state variables, three action variables, and time as another variable and the discriminant analysis was only analyzing the five state variables and the three action variables. This difference revealed that time does not have a significant impact on the classification of the clusters. It is interesting to note that some of the clusters could be identified using only action variables (highlighted in yellow). It is also interesting to note that some of the clusters could only be identified using both state and action variables (highlighted in red).

Table 4-1: Classification of Clusters

Cluster	Total Occurrences	Rank	State	Action	State-Action
1	33	24	0.91	0.12	1.00
2	151	19	0.95	0.34	1.00
3	1228	9	0.93	0.01	0.98
4	32	25	0.41	0.78	1.00
5	1929	5	0.86	0.00	0.97
6	3399	3	0.32	0.42	0.97
7	418	15	0.93	0.00	0.99
8	1708	7	0.26	0.51	0.95
9	1144	10	0.43	0.81	0.97
10	4608	1	0.15	0.17	0.96
11	4	29	1.00	1.00	1.00
12	29	26	0.93	0.38	1.00
13	42	23	0.98	0.45	1.00
14	1109	11	0.91	0.04	0.97
15	349	17	0.11	0.79	0.92
16	309	18	0.95	0.04	0.99
17	27	27	0.96	0.93	1.00
18	8	28	0.50	1.00	1.00
19	439	14	0.90	0.14	0.97
20	1873	6	0.97	0.02	0.98
21	60	22	0.02	0.93	1.00
22	1	30	1.00	1.00	1.00
23	70	21	1.00	0.04	1.00
24	508	13	1.00	0.19	1.00
25	1375	8	0.25	0.60	0.93
26	3682	2	0.37	0.18	0.98
27	82	20	1.00	0.01	1.00
28	2805	4	0.51	0.51	0.97
29	663	12	0.15	0.70	0.96
30	381	16	0.98	0.00	0.99

4.4 Discussion and Implications

The discriminant analysis could only distinguish all of the clusters using all of the five state variables and three action variables. Each cluster of driving behavior was a unique nexus of state and action variables, but the connection of driving states to driving actions was not straightforward. When a cluster cannot be identified using only state variables (highlighted rows in

Table 4-1), it means that there were multiple actions associated with the same driving state. When a cluster cannot be identified using only action variables, it means that there were multiple driving states that were associated with the same action. When clusters cannot be identified using either state variables or action variables but requires both (highlighted rows in red in

Table 4-1), it means that the state-action cluster was a specific combination of a driving state shared by multiple clusters and a driving action shared by multiple clusters.

The association of multiple driving actions to the same driving state can indicate two possibilities: (1) the driver had variability in their driving process or (2) there were missing variables that significantly influenced driver behavior. Both of these possibilities could be investigated by including additional variables in future studies. If the first possibility was true, then the driver would show a similar amount of variability in their driving process. If the second possibility was true, then the driver behavior would change according to differences in an additional variable.

4.5 Conclusions

This research effort identified and classified state-action clusters of car-following behavior. The methodology used in this effort first segmented and clustered car-following periods based upon five state variables (speed, lane offset, yaw angle, range and range rate) and three action variables (longitudinal acceleration, lateral acceleration, and yaw rate). The sample for this study was ten car drivers with over 100 car-following periods each, totaling 1500 car-following periods. The findings were that 18 of the 30 state-action clusters could be identified using only state variables. It was also found that 9 of the thirty state-action clusters could be identified using only action variables. Six of the 30 state-action clusters required both state and action variables to be identified. This indicated that there were overlaps between the state clusters and action clusters. Further research should include more car-following periods and analyze the difference between the car-following periods in depth in order to identify key variables that affect the behavior distributions of drivers.

References

- [1] M. Abbas, B. Higgs, A. Medina, and C. Y. D. Yang, "A Hybrid Wiedemann-GHR Model Calibration using Naturalistic Driving Data," in *13th International IEEE Conference on Intelligent Transportation Systems*, 2010.
- [2] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data," presented at the Transportation Research Board 90th Annual Meeting, Washington D.C., 2011.
- [3] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board 91st Annual Meeting, Washington D.C., 2012.
- [4] Bryan Higgs, Montasir Abbas, Alejandra Medina, and C. Y. D. Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," presented at the TRB Traffic Flow Theory Committee (TFTC) Summer Meeting and Greenshields Symposium, Annecy, France, 2010.
- [5] B. Higgs and M. Abbas, "A Cross-Validation of Car and Truck Agents Using Naturalistic Data: One-Parameter Car-Following Model for Reducing Computational Burden of Multi-Agent Systems," presented at the Conference on Agent-Based Modeling in Transportation Planning and Operations, Blacksburg, VA, 2013.
- [6] B. Higgs and M. Abbas, "A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, 2013, pp. 857-862.
- [7] B. Higgs and M. Abbas, "Multi-Resolution Comparison of Car-Following Models using Naturalistic Data," presented at the Transportation Research Board 93rd Annual Meeting, Washington D.C., 2014.
- [8] B. Higgs and M. Abbas, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns," *IEEE Intelligent Transportation Systems Transactions: Special Issue on Uncertainty in Computational Traffic Models*, 2014.

- [9] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [10] B. Higgs, M. M. Abbas, and A. Medina, "Naturalistic Car Driver Behavior: A Comparison of Car-Following Models," in *19th ITS World Congress*, 2012.
- [11] B. J. Higgs, "Application of Naturalistic Truck Driving Data to Analyze and Improve Car Following Models," Master's, Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, 2011.
- [12] B. Higgs and M. Abbas, "Experimental Design for a Psychophysiological Driving Simulator Study," presented at the 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 2014.
- [13] M. Abbas, A. Medina, B. Higgs, L. Chong, A. Leonhardt, C. McGhee, *et al.*, "Driver Behavior In Traffic, Final Report," Federal Highway Administration December 2012 2012.
- [14] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
- [15] "JMP," Pro 10 ed. Cary, NC: SAS Institute Inc., 1989-2013.

5. Multi-Resolution Comparison of Car-Following Models using Naturalistic Data⁴

Abstract

This research effort aimed to compare car-following models at different levels of analysis. The levels were: driver, car-following period, and cluster. The driver level compared models that were calibrated to multiple car-following periods of individual drivers. The car-following period level compared models that were calibrated to a specific car-following period and the cluster level compared models that were calibrated to clusters of data within a car-following period. Dividing a car-following period into clusters was accomplished through the use of a two-step algorithm for segmentation and clustering. The sample used in this paper included ten different car drivers and over 3000 car-following periods. The methodology of this paper compared the Gazis-Herman-Rothery, Intelligent Driver, Velocity Difference, Wiedemann, and Hybrid Wiedemann-GHR models along with segmented versions of the Gazis-Herman-Rothery, Intelligent Driver, and Velocity Difference Models. The results showed that the Hybrid Wiedemann-GHR and Velocity Difference models exhibited excellent performance when driver specific parameters are used for multiple car-following periods per driver. On the other hand, the Hybrid Wiedemann-GHR and the segmented Gazis-Herman-Rothery (cluster specific parameters) models exhibited excellent performance on a car-following period specific basis.

⁴ Paper has been published in the proceedings of the Transportation Research Board 93rd Annual Meeting

5.1 Introduction

Over the past 50 years there have been many advances and innovations made in the modeling of car-following behaviors of drivers. Multiple models have been created to explain and recreate the behaviors of drivers. Comparison of the models depends upon the data used in the analysis and this effort aims to analyze multiple models at different levels in order to shed light upon the strengths and weaknesses of each model. Different levels of analysis are required due to the simple fact that human behavior is variable. This fact means that variability cannot be excluded from car-following models, but models should embrace variability. Human behavior can be influenced by many factors; in reference to the modeling of car-following behaviors of drivers, these factors can be numerous. Great care must be taken to make valid and reasonable assumptions to simplify, yet accurately model driver behavior. The biggest of these assumptions is homogeneity. Assuming homogeneity greatly simplifies the modeling task, but attributes changes in human behavior to erratic deviation of the norm. Calibration to multiple car-following periods at once creates a distribution of performance where some car-following periods have low error while others have high error. The objective of this research effort was to compare various car-following models at different levels of homogeneity including: driver level, car-following period level, and within car-following period level. This analysis revealed the top performing models at each level and examined the distribution of performance for each model in order to find the level of analysis and model that produced the best performance.

In the following sections, we first present a synopsis of the Naturalistic Driving Data that were previously collected by Virginia Tech Transportation Institute (VTTI) and were mined in this research. Next, we present a synthesis of past efforts that summarizes previous work that influenced and guided this effort. Then, we describe in detail our methodology for comparison of the models. Lastly, we present the implications of our findings and our conclusions.

5.1.1 Naturalistic Driving Data

Naturalistic driving data refers to data collected from drivers in their natural environment. This means that the data were collected while the drivers were completing their normal daily routines. The main characteristics of naturalistic driving data are the following: unobtrusive observation, participation in real traffic streams, no intervention by the observers, data collected by multiple sensors, and videos record the driver and surrounding environment. This was accomplished by equipping vehicles with specialized sensors and “vehicle network” recording equipment and then allowing the participants to drive the vehicles as they see fit. The equipment recorded a large number of variables (e.g., speed, acceleration, steering wheel positions), and, of specific interest to car-following, a radar positioned at the front of the vehicle recorded the differences in position and speed between the subject vehicle and the lead vehicle. The equipment also included cameras that unobtrusively recorded: what the driver saw from the front, what the driver saw from the two side mirrors, the drivers’ face, and what the driver was doing inside the vehicle.

This data offered new insights into the nature of driving behavior, but the sheer amount of data collected can be overwhelming. For example, the naturalistic truck driving study (NTDS) conducted by VTTI [1], collected approximately 14,500 driving-data hours covering 735,000 miles traveled and the naturalistic car driving study collected approximately 43,000 driving-data hours covering 2,000,000 miles traveled. This data were collected from nine trucks and 100 cars that were used by multiple drivers.

5.1.2 Synthesis of Past Efforts

The past efforts that influenced this work are divided into two categories: (1) modeling car-following behaviors and (2) data clustering. The modeling of car-following behaviors focuses on the numerous car-following models that have been developed and improved over the years. Data clustering focuses on the techniques employed in clustering data that have been applied in multiple fields.

Ossen and Hoogendoorn [2] studied the car-following behavior of individual drivers using vehicle trajectory data that were extracted from high-resolution digital images collected at a high frequency from a helicopter. The analysis was performed by estimating the parameters of different specifications of the Gazis-Herman-Rothery (GHR) car-following model car-following rule for individual drivers. In 80 % of the cases, a statistical relation between stimuli and response could be established. The Gipps (a safe distance model) and Tampere (stimulus-response model) models and a synthetic data based approach were used for assessing the impact of measurement errors on calibration results. According to the authors, the main contribution of their study was that considerable differences between the car-following behaviors of individual drivers were identified that can be expressed in terms of different optimal parameters and also as different car-following models that appear to be optimal based on the individual driver data. This is an important result taking into account that in most models a single car-following rule is used. The authors also proposed for future research to apply more advanced statistical methods and to use larger databases. Brackstone et. al. [3] used data collected with an instrumented vehicle that was assembled at TRG Southampton to parameterize the Wiedemann's threshold for a typical following spiral. As a result they represent the action points as a function of a probability distribution based on ground speed.

Micro-simulation software packages use a variety of car-following models including Gipps' (AIMSUN, SISTM, and DRACULA), Wiedemann's (VISSIM), Pipe's (CORSIM), and Fritzsche's (PARAMICS). And different automated calibration parameters such as genetic algorithms have been used to calibrate the distribution of car-following sensitivity parameters [4]. For example, Panwai and Dia [5] compared the car-following models between different simulation software, including AIMSUN, PARAMICS and VISSIM using an instrumented vehicle to record differences in speed and headway (Leading speed, relative distance, relative speed, follower acceleration were recorded). The error metric (EM) shows similar values for psychophysical models in VISSIM and PARAMICS and lower values in AIMSUN. The RMS error and qualitative drift and goal-seeking analyses also showed a substantially different car-following behavior for PARAMICS. Siuhi and Kaseko [6] demonstrated the need for separate models for acceleration and deceleration responses by developing a family of car-following models and addressing the shortcomings of the General Motors (GM) model. Previous work from Osaki [7] and Subranmanian [8] modified the GM model separating the acceleration and deceleration responses. Ahmed [9], following some work from Subranmanian assumed non linearity in the stimulus term and introduced traffic density. Results from Ahmed [9] and Toledo [10] showed that acceleration increases with speed but decreases with vehicle separation. Due to statistical insignificance, Ahmed and Toledo also removed speed from their deceleration models. Siuhi and Kasvo [6] addressed some of these shortcomings by developing separate models, not only for acceleration and deceleration, but also for steady-state responses. Nonlinear regression with robust standard errors was used to estimate the model parameters and obtain the distributions across drivers. The stimulus response thresholds that delimit the acceleration and deceleration responses were determined based on signal detection theory.

Using two models of similar complexity (number of parameters): the “Intelligent Driver Model” (IDM) and the “Velocity Difference Model” (VDIFF), Kesting and Treiber [11] researched car-following behaviors on individual drivers using publicly available trajectory data for a straight one-lane road in Stuttgart, Germany. They used a nonlinear optimization procedure based on a genetic algorithm to minimize the deviations between the observed driving dynamics and the simulated trajectory. One of the major findings of the study was that a significant part of the deviations between measured and simulated trajectories can be attributed to the inter-driver variability and the intra-driver variability (stipulating that human drivers do not drive constantly over time, and their behavioral driving parameters change)—the latter accounts for a large part of the deviations between simulations and empirical observations.

Menneni et al [12] presented a calibration methodology of the VISSIM Wiedemann car-following model based on integrated use of microscopic and macroscopic data using NGSIM Relative distance vs. relative speed graphs were used for the microscopic calibration, specifically to determine the action points (it is important to note that action points were not identical to perception threshold). Scatter and distribution of action points on relative distance versus relative velocity graphs also showed similarity in driver behavior between the two freeways.

Higgs et al. [13] presented a calibration of the Wiedemann model over different speed categories. The findings were that each speed category resulted in different parameters and the differences were driver specific. Higgs et al. [14] also worked on adding new thresholds to the Wiedemann model. The new thresholds were: a hook following threshold and a pass threshold. Hook following is defined as when a driver is passed by another vehicle and then maintains a car-following period with that vehicle. The pass threshold is created when a driver is catching up to another vehicle and sometimes the driver decides to follow the lead vehicle and other times the driver decides to immediately pass the lead vehicle. The pass threshold also aids in discovering times when the driver made a decision to pass but was forced to follow due to other factors like being trapped by a vehicle to the side.

Hoogendoorn and Hoogendoorn [15] proposed a generic calibration framework for joint estimation of car-following models. The method employed relies on the generic form of most models and weights each model based on its complexity. This new approach can cross-compare models of varying complexity and even use multiple trajectories when individual trajectory data is scarce. Prior information can also be used to realistically estimate parameter values.

Abbas et al. [16] presented a hybrid car-following model (W-GHR) that combines the strengths of the Wiedemann model and the GHR model. The model uses the thresholds of the Wiedemann model to define the driving tasks such as decelerating to follow lead vehicle, accelerating to maintain distance to lead vehicle, decelerating to maintain distance to lead vehicle, etc. The model then uses specific calibrated GHR parameters for each task, due to the difference between the tasks, to calculate the acceleration of the following vehicle.

The research into clustering data is divided into two main groups: univariate clustering and multivariate clustering. There are many facets within each group, but for the application to modeling car-following behavior, the facet of time-series clustering is the most important. Most of the research into clustering has occurred outside of the field of transportation engineering. Univariate clustering is easier than multivariate clustering, thus there is more extensive research into using univariate clustering methods. Fan et al. presented a methodology for clustering time-series data. His methodology has two main parts: the segmentation of the data and then the clustering of the segments. The paper suggests various algorithms for both segmentation and

clustering [17]. Kremer et al. also worked with time-series clustering, but the focus of the effort was placed upon the changes or evolution of different clusters over time [18].

The research into multivariate clustering is limited, but the impacts of the researched methods have a much broader range of applications than univariate clustering methods. Plant et al. worked with the clustering of multivariate time-series. The researchers emphasized clustering based upon the interactions of the different variables [19]. Wang et al. conducted research on the clustering of multivariate time-series data. Their research was based upon the reduction of various segments of the data to statistical measures which could then be clustered by normal k-means methods [20]. In the area of transportation, Higgs et al.[21] have proposed a segmentation and clustering technique for car-following data. The technique is multivariate, considering 8 variables. The technique first divides car-following periods into segments using an optimization equation and then clusters those segments using a k-means method. This work highlights that drivers show distributions of behavior and that each driver has a unique distribution.

5.2 Methodology

The methodology used in this study expanded upon previous work [21] that first extracted car-following periods from the exorbitant amount of data contained within the Naturalistic Database using the method from previous work [13, 14, 22, 23]. This analysis had multiple levels of resolution for comparison of the GHR, IDM, VDIFF, Wiedemann, and W-GHR models. The first level was created by grouping car-following periods according to driver. The second level was individual car-following periods and the third level used the segmentation and clustering method mentioned in previous work [21] to create a level of analysis within a car-following period. The formulation of the segmentation is shown in Equations 1 and 2. Equation 1 converts each variable to a Z-scale, borrowed from statistics, accounting for the standard deviation as shown in Equation 2. The Z-scale converts all the variables to the same scale and puts each data point on terms of deviation from the mean which allows this equation to be used to try to minimize each data point's deviation from the mean on multiple variables.

$$\text{Min } Z = \sum_{i=1}^l \sum_{j=1}^m \sum_{k=1}^n \frac{|x_{ijk} - \bar{x}_{ij}|}{s_{ij}} \quad (1)$$

Subject to:

$$\begin{aligned} t_i &= [0 \sim T] \forall_i \\ \sum t_i &= T \end{aligned}$$

Where:

x_{ijk} is the kth observation of variable j in segment i

\bar{x}_{ij} is the centroid of segment i for variable j

l is the number of segments in a car-following period

m is the number of variables

n is the number of observations in segment i

t_i is the length, in time, of segment i

$$s_{ij} = \sqrt{\frac{\sum_{k=1}^n (x_{ijk} - \bar{x}_{ij})^2}{n-1}} \quad (2)$$

The segmentation and clustering method first divided a car-following period into continuous segments of similar data. For example, a period could be divided into 4 parts where the driver maintained speed for 4 seconds, accelerated for 2 seconds, braked for 1 second and then maintained speed for 3 seconds. The segments were then clustered together because some behaviors can repeat over time. For example, using the example given earlier, both segments where the driver maintained speed would be considered one cluster while the other two segments would remain in separate clusters. The analysis of all drivers together was excluded from the car-following model comparison due to the high amount of error each model produced in any attempt to use data from multiple drivers in one calibration. Table 5-1 describes each level of analysis through the data for each level and the number of calibrated parameters used for each level of the analysis.

Table 5-1: Multiple Resolutions of Analysis.

Level of Analysis	Data	Calibration
Driver Specific	Data from all car-following periods of one driver	One set of calibrated parameters is used for each driver
Car-Following Period Specific	Data from one car-following period	One set of calibrated parameters is used for each car-following period
Cluster Specific	Data from one car-following period that is divided into clusters of data	One set of calibrated parameters is used for each cluster of data

In this paper, the car-following behaviors of ten different drivers were compared. Over 3000 car-following periods of data were used in total. This study included five different car-following models: Gazis-Herman-Rothery (GHR), Intelligent Driver Model (IDM), Velocity Difference Model (VDIFF), Wiedemann, Hybrid Wiedemann-GHR model (W-GHR). These models were chosen because they represented both existing models and new innovations. The IDM, VDIFF and Wiedemann models were chosen for being resilient and accurate in previous studies [13, 24]. The GHR model was chosen for comparison to the W-GHR and segmented models that both use the GHR model in their formulation.

Table 5-2 lists the formulations for the GHR, Wiedemann, IDM, and VDIF models used in this paper.

Table 5-2: Table of Car Following models and parameters

GHR	$a_f(t + T_r) = cv_f^m(t + T_r) \frac{\Delta v(t)}{\Delta x^l(t)}$	<p>T_r =time between the observation of a certain stimulus and the reaction on that stimulus $a_f(t + T_r)$ =acceleration of the following vehicle at time $t + T_r$ $v_f(t + T_r)$ = speed of following vehicle at time $t + T_r$ $\Delta v(t)$= relative speed between the following car and the car immediately in front ($v_{leader} - v_{follower}$) $\Delta x(t)$= relative distance between following car and car immediately in front ($x_{leader} - x_{follower}$) m, l, c= parameters describing car following behavior</p>
Wiedemann Model	$AX = L_{n-1} + AXadd + RND1_n \times AXmult$ $ABX = AX + BX$ $BX = (BXadd + BXmult \times RND1_n) \times \sqrt{v}$ $v = \begin{cases} v_{n-1} & \text{for } v_n > v_{n-1} \\ v_n & \text{for } v_n \leq v_{n-1} \end{cases}$ $SDX = AX + EX \times BX$ $EX = (EXadd + EXmult \times (NRND - RND2_n))$ $SDV = \left(\frac{\Delta x - L_{n-1} - AX}{CX} \right)^2$ $CX = CXconst \times (CXadd + CXmult \times (NRND1_n + RND2_n))$ $OPDV = CLDV \times (-OPDVadd - OPDVMult \times NRND)$ $AD = S_{n-1} + T_D \times v_n$ $AR = S_{n-1} + T_r \times v_{n-1}$ $AS = S_{n-1} + T_s \times v_n$ $\Delta b_m = b_{min} + a_{n-1}$ $AB = AR + \frac{\Delta v^2}{\Delta b_m}$	<p>AX: the desired distance between stationary vehicles ABX: the desired minimum following distance at low speed differences SDX: the threshold for maximum following distance SDV: the point which the driver notice that he approach a slower vehicle. OPDV : Increasing speed difference, L_{n-1} : The leading vehicle length AXadd and AXmult are calibration parameters RND1_n is a normal distributed random number for vehicle (n) v: Vehicle speed CXconst, CXadd and CXmulti are calibration parameters NRDV is a normally distributed random number OPDVadd and OPDVMult are calibration parameters</p>
Intelligent Driver Model (IDM)	$a_n(t + \tau) = a \left[1 - \left(\frac{v_n(t)}{v^*} \right)^\delta - \left(\frac{\Delta x_{min}^*(v_n(t), \Delta v_{n,n-1}(t))}{\Delta x_{n-1,n}(t)} \right)^2 \right]$ $\Delta x_{min}^*(v_n(t), \Delta v_{n,n-1}(t)) = d + T_{safe} * v_n(t) + \frac{v_n(t) \Delta v_{n,n-1}(t)}{2\sqrt{a^{max} * b^{max abs}}}$	<p>$\Delta v_{n,n-1}(t)$=approaching rate of the following vehicle v^*=desired speed T_{safe}=safe time headway a^{max}=maximum desired acceleration of following vehicle $b^{max abs}$=absolute maximum desired deceleration of following vehicle δ acceleration component $\Delta x_{n-1,n}(t)$=distance headway d=vehicle length</p>
Velocity Difference Model (VDIFF)	$a(s, v, \Delta v) = \frac{v_{opt}(s) - v}{\tau} - \lambda \Delta v$ $v_{opt}(s) = \frac{v_0}{2} \left[\tan \left(\frac{s}{l_{int}} - \beta \right) - \tan(-\beta) \right]$	<p>s=distance headway Δv=relative velocity τ=relaxation time λ=sensitivity parameter v_0 =desired velocity l_{int}=interaction length β=form factor</p>

The GHR, VDIFF, and IDM models were also used in this paper in a manner that followed the hybrid Wiedemann-GHR car-following model [16]. In the hybrid model concept, the GHR model was calibrated to small regimes of behavior that were within certain thresholds. In the original hybrid model, the thresholds were defined by the Wiedemann model, but in this paper the thresholds were found using the segmentation and clustering technique from previous work [21]. This allowed the third level of analysis within a car-following period, where parameters for each model were calibrated to each cluster of data points or regimes. The clusters were then recombined into one continuous car-following period in order to show a comparison to the other models at the car-following period level.

5.3 Results

The validation of the methodology consisted of a cross-comparison of the root mean squared error of the GHR car-following model at the different levels of inquiry. The levels of inquiry were as follows: (1) data from ten car drivers, (2) data from one car driver, (3) data from one car-following period, and (4) data from a segmented and clustered car-following period. For each level, the GHR model was calibrated to the data to find the root mean squared error for the difference in speed between the model and the data. For the last level, each cluster was calibrated to the GHR model in order to find parameters that apply to each cluster individually. Figure 5-1 shows the results of the validation and how different parameters of the GHR model performed for an example car-following period. The figure shows different data series that represent: the GHR model parameters calibrated to all ten car drivers, the GHR model parameters calibrated to all car-following periods of one driver, the GHR model parameters calibrated to one car-following period, and the GHR model parameters calibrated to each segment of one car-following period. The trend showed that as the resolution level changes to more specific data, the accuracy of the model increased which indicates that the performance of a model was not consistent at the different levels of analysis and thus a model could be superior at one level while being inferior at another level.

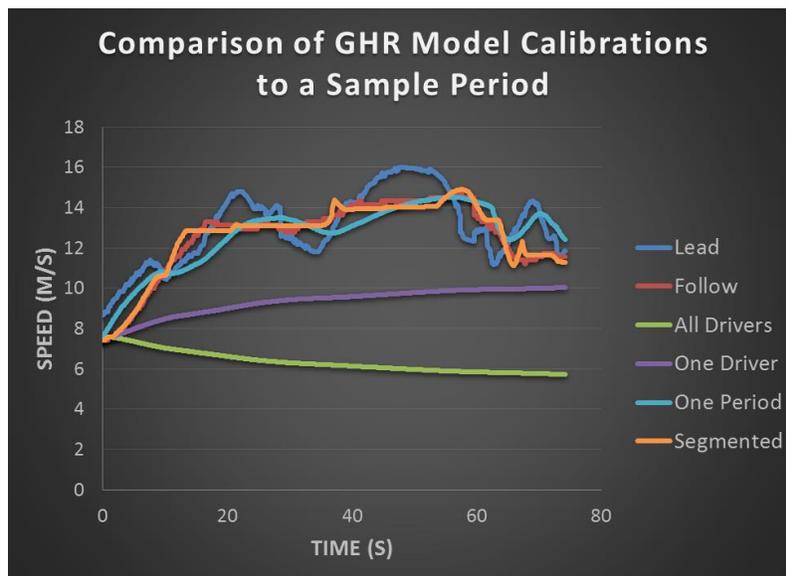


Figure 5-1: Comparison of GHR Model Calibrations to Different Levels of Analysis

5.3.1 Driver Level of Analysis

The driver level of analysis used all car-following periods of each driver to obtain driver specific calibrated parameters, but this assumed that all the car-following periods could be grouped together. Figure 5-2 shows that only using one set of parameters for a driver resulted in high variability in error between different car-following periods. For example, the Wiedemann model had five car-following periods with a RMSE less than 0.5 and three periods with RMSE greater than 5.5. This highlighted that the calibration does not accurately represent every car-following period, but instead accurately represented a fraction of the periods and accepted a large error for the rest of the periods. The figure also shows that the distributions were different for each model.

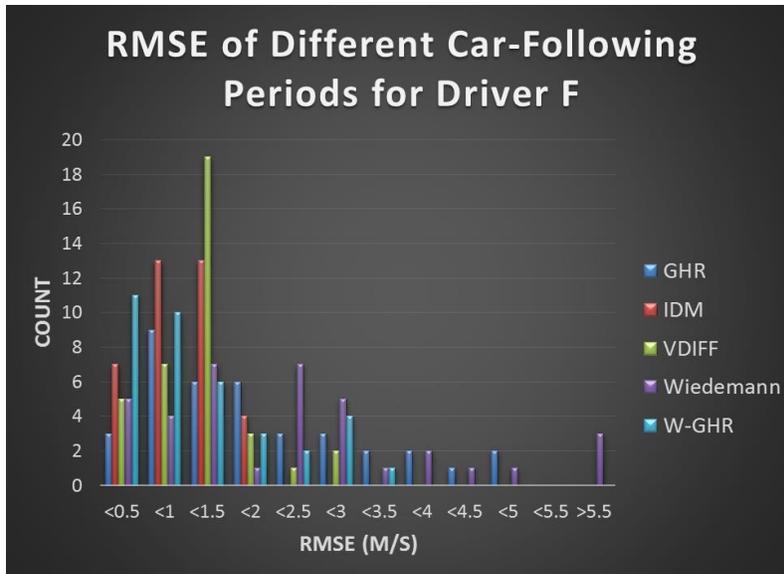


Figure 5-2: Histogram of RMSE of Car-Following Periods using Parameters Calibrated to Driver F

To summarize the distributions of RMSE for each model, the average and standard deviation were calculated and presented in Table 5-3. The table shows that for multiple car-following periods per driver, the W-GHR model offered the lowest average root mean squared error along with the lowest standard deviation. Also of interest, is the fact that the W-GHR model is a combination of the Wiedemann and GHR models which showed the two highest errors.

Table 5-3: RMSE by Model for All Car-Following Periods of Each Driver

	GHR	IDM	VDIFF	Wiedemann	W-GHR
Average	2.15	1.56	1.30	29.90	1.21
Std Dev	0.36	1.03	0.26	33.65	0.14

5.3.2 Car-Following Period Level of Analysis

The car-following period level of analysis showed both the results at the car-following period level and cluster level. This was due to the fact that the clusters combine to make a continuous car-following period which was considered as a basis for comparison. Figure 5-3 shows a comparison of the GHR, IDM, VDIFF, Wiedemann, W-GHR, segmented GHR (S. GHR), segmented IDM (S. IDM), and segmented VDIFF (S. VDIFF) models for an example car-following period.

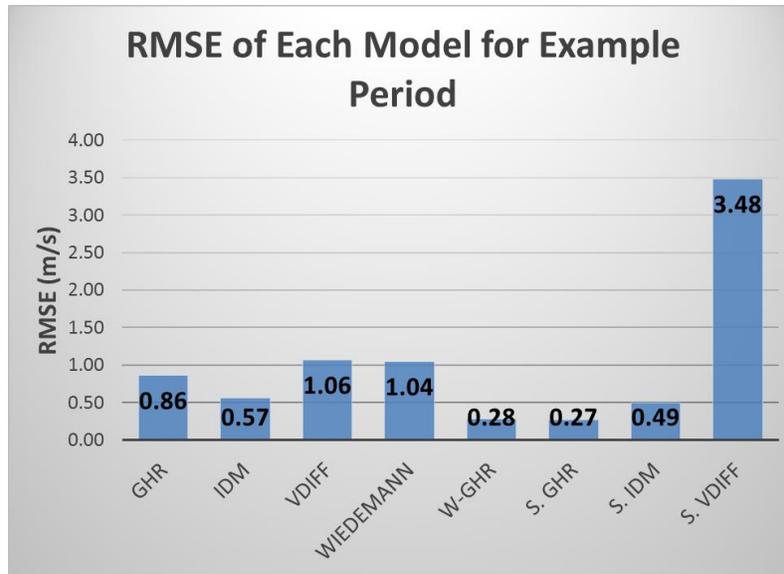


Figure 5-3: Comparison of Models for Example Car-Following Period

Figure 5-4 shows the speed vs. time plot for the top three models identified in Figure 5-3 for the example car-following period. All three models showed a general adherence to the data, but each model showed different deviations from the data. For example, the segmented IDM model showed a slow acceleration at the start of the period followed by a high acceleration where the data showed a constant medium acceleration for 0 to 15 seconds. The similar performances of the segmented GHR and W-GHR models was not surprising due to the similarities of the models. Each model divided the car-following period into segments and then applied the GHR model to each segment. The difference was that the W-GHR model defined the divisions between segments using the Wiedemann model thresholds while the segmented GHR model defined the divisions using the segmentation and clustering method.

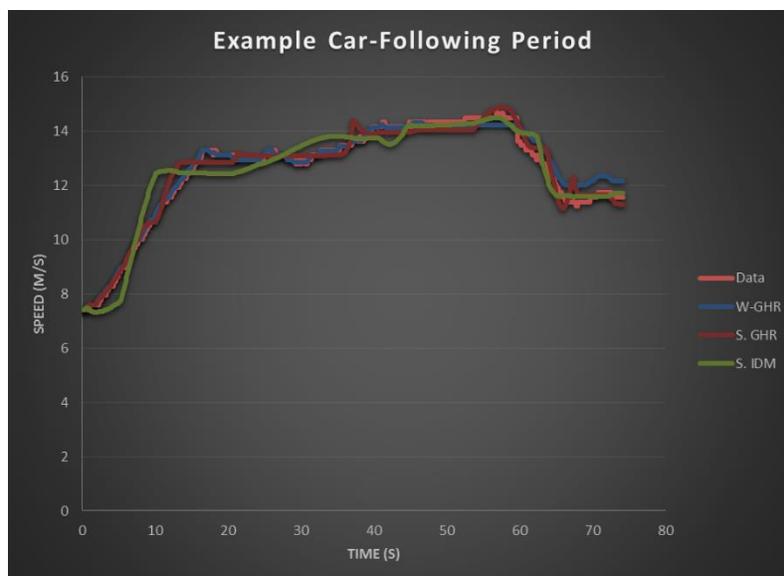


Figure 5-4: Comparison of Top 3 Models for Example Car-Following Period

5.3.3 Cluster Level of Analysis

For comparison on the cluster level, the GHR, IDM, VDIFF, Wiedemann, and W-GHR models used the same speed trajectories from the car-following period level of analysis. The root mean squared error was calculated by cluster using the clusters given by the segmentation and clustering method which allowed the models without cluster specific parameters to be compared to models with cluster specific parameters on the cluster level. Table 5-4 shows the root mean squared error for each segmented model by cluster which showed that the segmented GHR and W-GHR models had the best performance in most of the clusters. In cluster 4, the segmented IDM model outperformed the segmented GHR model this showed that the S. GHR model was not the best model to use for every cluster. The segmenting method divided a car-following period into different segments of behavior much like the Wiedemann and W-GHR thresholds, thus the similar performance of the W-GHR and S. GHR models was to be expected. The table also shows that the performance of each model was variable depending upon the cluster. This suggested that the data presented some scenarios that were conducive to each model and some scenarios that conflict with each model. Also of interest, was that the S. GHR and S. IDM models showed reduced error in most clusters as compared to the original models (GHR and IDM, respectively).

Table 5-4: RMSE of Segmented Models by Cluster for Example Car-Following Period

Cluster	1	2	3	4	5	6	7	8
GHR	0.86	1.34	0.69	0.63	0.98	0.43	0.35	1.40
IDM	0.26	0.64	0.54	0.38	0.62	0.46	0.29	1.13
VDIFF	0.74	1.16	0.51	1.38	1.10	1.01	0.56	1.57
Wiedemann	0.75	1.48	0.68	1.28	0.58	1.13	0.80	0.91
W-GHR	0.09	0.40	0.09	0.43	0.38	0.05	0.06	0.11
S. GHR	0.13	0.29	0.20	0.39	0.44	0.20	0.14	0.39
S. IDM	0.56	0.43	0.75	0.33	0.82	0.43	0.21	0.44
S. VDIFF	0.94	3.89	2.76	4.51	5.21	2.67	2.80	4.94

5.4 Implications

The findings of this paper have multiple implications on the field of car-following models. The errors shown in Figure 5-1 support that “overloading” a car-following model with too much data can result in large errors. This does not mean that the model is wrong or insufficient; it merely means that the model is not suited to handle data from multiple different situations at the same time. To avoid this, the data for future research needs to be filtered and clustered such that each scenario is calibrated separately. Failing to do so, will result in any parameter values obtained through calibration being representative of a fraction of the data.

Car-following models need data that is specific. Data from different drivers should be considered incompatible as there are numerous factors that influence drivers to behave differently (personalities, demographics, age, etc.). Also, data from one driver needs to be categorized as some car-following periods are harmonious while others are not harmonious as shown by the distributions in Figure 5-2. Even the data from a single car-following period needs to be checked for continuity as shown by the results in Table 5-4. Comparison of different clusters of data within a car-following period can reveal data points or behaviors that are difficult for a model to recreate which can then become a focal point of effort to make car-following models more robust.

Simulation of multiple clusters of behavior can be accomplished through the use a Markov chain that relates the transitions between clusters.

5.5 Conclusions

This research effort aimed to compare various car-following models at different levels of analysis. The methodology used in this effort examined three different levels of analysis: driver, car-following period, and within a car-following period levels. At the driver level, data from ten car drivers was used to calibrate the GHR, IDM, VDIFF, Wiedemann, and W-GHR models. At the car-following period level data from a car-following period was used for calibration. Within a car-following period, the period was divided into segments and clustered using a method from previous work and each cluster was calibrated separately. The sample for this study was ten car drivers, with over 3000 car-following periods in total.

Analyzing multiple car-following periods at once can result in high variability of error. This means that a calibration is forced to sacrifice accuracy in one scenario in order to accurately represent another scenario. Examining drivers individually improved the accuracy of a model, but the resulting parameters accurately represent only a fraction of the car-following periods. Analyzing individual car-following periods improved the accuracy as well, but still showed signs of trying to represent too many scenarios at once. The use of the segmentation and clustering algorithm allowed another level of analysis within a car-following period by dividing that period into smaller parts. Calibration of the segmented periods required more parameters, but the error was reduced significantly as compared to a model calibrated to the full period.

At the driver specific level, the W-GHR and VDIFF models showed the lowest error. At the car-following period specific level, the W-GHR model showed the lowest error. At the segmented car-following period level, the S. GHR model showed the lowest error. This highlighted that each model has advantages and disadvantages that depend upon the level of analysis and that current models can be improved by using the segmentation and clustering method. Future research would be to focus on the identification of sorting techniques that can properly group and separate whole car-following periods. Also, future research would be to identify methods for grouping drivers in order to reduce the computational load required for analysis of car-following behaviors.

References

- [1] R. Olson, R. J. Bocanegra, R. Hanowski, and J. Hickman, "Driver Distraction in Commercial Vehicle Operations," Center for Truck and Bus Safety, Virginia Tech Transportation Institute, Blacksburg, VA FMCSA-RRR-09-042, 2009.
- [2] S. Ossen and S. P. Hoogendoorn, "Validity of trajectory-based calibration approach of car-following models in presence of measurement errors," *Transportation Research Record*, pp. 117-125, 2008.
- [3] M. Brackstone, "Driver Psychological Types and Car Following: Is There a Correlation? Results of a Pilot Study," in *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2003, p. 6.
- [4] G. G. Schultz and L. R. Rilett, "Analysis of Distribution and Calibration of Car-Following Sensitivity Parameters in Microscopic Traffic Simulation Models," p. 11, 2004.
- [5] S. Panwai and H. Dia, "Comparative evaluation of microscopic car-following behavior," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 6, pp. 314-325, 2005.
- [6] S. Siuhi and M. Kaseko, "Parametric Study of Stimulus-Response Behavior for Car-Following Models," in *89th Transportation Research Board Annual Meeting*, Washington, D.C., 2010.
- [7] H. Osaki, "Reaction and anticipation in the Car-Following Behavior," in *12th International Symposium on the Theory of Traffic Flow and Transportation*, 1993.
- [8] H. Subranmanian, "Estimation of Car-Following Models," Master of Science in Transportation Master Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1996.
- [9] K. I. Ahmed, "Modeling Drivers' Acceleration and Lane Changing Behavior," Doctor of Science in Transportation Systems and Decision Sciences PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1999.
- [10] T. Toledo, "Integrating Driving Behavior," Doctor of Philosophy PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2003.
- [11] A. Kesting and M. Treiber, "Calibrating Car-Following Models using Trajectory Data: Methodological Study," *Transportation Research Record*, vol. 2088, pp. 148-156, 2008.
- [12] S. Menneni, C. P. D. Sun, P.E., and P. Vortisch, "An Integrated Microscopic and Macroscopic Calibration for Psycho-Physical Car Following Models " in *Transportation Research Board Annual Meeting*, 2009.
- [13] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [14] M. A. Bryan Higgs, Alejandra Medina, C. Y. David Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," presented at the TRB Traffic Flow Theory Committee (TFTC) Summer Meeting and Greenshields Symposium, Annecy, France, 2010.
- [15] S. Hoogendoorn and R. Hoogendoorn, "A Generic Calibration Framework for Joint Estimation of Car Following Models using Microscopic Data," *TRB 2010 Annual Meeting CD-ROM*, p. 17, 2010.
- [16] M. M. H. Abbas, Bryan; Medina Flintsch, Alejandra;, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board Annual Meeting, Washington D.C., 2012.
- [17] Z.-H. Fan, D.-K. Huang, J.-Z. Li, and K.-H. Wang, "Morphing Cluster Dynamics: Time series clustering with multipartite graph," in *2004 International Conference on Machine Learning and Cybernetics*, Shanghai, China, 2004, pp. 1749-1754.
- [18] H. Kremer, S. Gunnemann, and T. Seidl, "Detecting Climate Change in Multivariate Time Series Data by Novel Clustering and Cluster Tracing Techniques," in *10th IEEE International Conference on Data Mining Workshops*, Sydney, NSW, Australia, 2010, pp. 96-97.
- [19] C. Plant, A. M. Wohlschlagel, and A. Zherdin, "Interaction-Based Clustering of Multivariate Time Series," in *9th IEEE International Conference on Data Mining*, 2009, pp. 914-919.
- [20] X. Wang, A. Wirth, and L. Wang, "Structure-based statistical features and multivariate time series clustering," in *7th IEEE International Conference on Data Mining*, Omaha, NE, United states, 2007, pp. 351-360.
- [21] B. Higgs and M. Abbas, "A Two-Step Segmentation Algorithm for Behavioral Clustering of Naturalistic Driving Styles," presented at the 16th International IEEE Conference on Intelligent Transportation Systems, The Hague, The Netherlands, 2013.
- [22] M. Abbas, B. Higgs, A. Medina, and C. Y. D. Yang, "Discriminant analysis of driver behavior before safety critical events," *Advances in Transportation Studies*, pp. 77-86, 2011.

- [23] B. Higgs, "Application of Naturalistic Truck Driving Data to Analyze and Improve Car Following Models," Master of Science, Civil Engineering, Virginia Tech, 2011.
- [24] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data," presented at the Transportation Research Board 90th Annual Meeting, Washington D.C., 2011.

6. Experimental Design for a Psychophysiological Driving Simulator Study⁵

Abstract

This research effort described a methodology to investigate the effects of personality and emotion on driver behavior. The methodology detailed in this paper had four main components: development of psychological personality surveys, development of psychological emotion surveys, physiological data collection, and development of simulator scenarios. There were three simulator scenarios used in this study: a base scenario, an anxiety scenario, and an anger scenario. The base scenario allowed the participants to interact without any pressure being applied to them. The anxiety scenario applied pressure to the participants through a time limit and police vehicles patrolling the roadway. The anger scenario applied pressure to the participants through aggressive actions by the surrounding traffic. The results showed that there can be drastic differences in driver behavior in different emotional states. The tendencies were that drivers become more aggressive and show smaller headways when they were angered. The results also showed that drivers have different emotional responses to the same stimuli.

⁵ Paper has been submitted for publication in the proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC 2014)

6.1 Introduction

There are many factors that can affect driving behavior that are not considered in modeling. The main issue to consideration of most of these factors is that they are difficult to collect and measure. Emotions and personality traits are among these factors because they are internal to the driver. The influence of personality and emotions are seen in driver behavior through the existence and tendencies of people to have road rage. Road rage is a combination of personality traits and an emotional state. People have different ways to express their anger including: verbally by yelling and cursing, physically by swerving and accelerating, and in some cases vengefully by targeting and attacking the vehicle that angered them.

The methodology in this paper addressed the need to collect personality and emotion data during driving studies. The proposed methodology had four main components: development of psychological personality surveys, development of psychological emotion surveys, physiological data collection, and development of simulator scenarios. In the following sections, we first present a synthesis of past effort, followed by a description of our methodology, and finally we present our findings and conclusion.

6.1.1 Synthesis of Past Efforts

The past efforts that influenced this work are divided into three categories: (1) car-following data collection methods, (2) psychological inventories, and (3) psychophysiology. The data collection methods focus on the strengths and weaknesses of various car-following data collection methods. The psychological inventories focuses on various inventories that have been developed to measure a person's psychological state and their personality. Finally, psychophysiology focuses on the measurement of physiological changes in the body that occur in response to changes in psychological states.

6.1.1.1 *Car-Following Data Collection Methods*

Car-following studies typically collect vehicle trajectory data through various means, including Naturalistic, simulator, and video data collection methods.

TABLE 6-1 highlights the strengths and weaknesses of each data collection method, which ultimately affect the accuracy of the modeling results obtained through calibration of car-following models. For example, naturalistic and simulator data are useful for observing the range of behaviors of one driver, but fixed camera data would be useful for observing multiple drivers in the same traffic stream.

Table 6-1: Types of Vehicle Trajectory Data Collection

Data Type	Description	Strengths	Weaknesses
Naturalistic [1]	-An instrumented vehicle is driven in normal driving routines	-Driver is in natural environment -Multiple trajectories are observed for each driver	-Drivers know they are being observed
Simulator [2, 3]	-Drivers drive in a simulated environment	-Driving environment in controlled -Multiple trajectories are observed for each driver	-Drivers know they are being observed -Drivers might perform differently in a simulated environment
Video [4, 5]	-A video camera is used to collect data within a certain area	-Drivers do not know they are being observed	-Observation period and area are fixed -Only one trajectory is observed per vehicle

6.1.1.2 Psychological Inventories

Psychological inventories measure either personality traits or emotional states. Personality traits are the tendencies of a person in everyday situations. Emotional states are the emotions that are being felt in the moment. For example, when a person rates high in trait anger (personality), they are more likely to become angry (state) than others in situations that are irritating. TABLE 6-2 shows a list of psychological inventories and gives a description of what they measure.

Table 6-2: Psychological State and Trait Inventories

Inventory	Description	State/Trait
Cook-Medley Hostility Scale [6]	Measures a person's negative views of social interaction (distrust, aggression, etc.)	Trait
Semantic Differential Scale [7-9]	Measures a person's current emotional state in terms of: pleasure, arousal, and dominance.	State
Taylor Manifest Anxiety Scale [10]	Measures how anxiety manifests in a person's daily life.	Trait
State-Trait Anger Expression Inventory [11]	Measures five aspects of anger: State anger, trait anger, anger in, anger out, and anger control.	State and Trait
Driving Anger Scale [12]	Measures how much anger different driving situations induce in a person.	Trait

6.1.1.3 Psychophysiology

The human body reacts to stimuli by various changes (increased heart rate, pupil dilation, etc.) which all stem from the nervous system, particularly the autonomic nervous system. The autonomic nervous system is divided into two parts the parasympathetic and sympathetic nervous systems. The parasympathetic nervous system is responsible for regulation of the body's organs at rest while the sympathetic nervous system is responsible for the fight or flight response. When the sympathetic nervous system activates the following occurs [13]:

- Increased mental activity
- Increased secretion of adrenaline (epinephrine), noradrenalin (norepinephrine) and cortisol
- Increased heart rate
- Increased breathing rate

- Increased muscle contraction which leads to increased strength
- Pupils of the eyes dilate
- Sweat glands increase secretion

The stimuli that can cause the nervous system to react also includes the emotional state of the driver. For example, an angry driver will have an increased heart rate and respiration rate and their attention will narrow onto a target. The physiological consequences of the emotional state can affect driving performance, as stated in the example by a narrowing of the driver’s attention. Physiological changes that occur in the body in response to changes in psychological state can be captured and represented using psychophysiological measures. Psychophysiological measures have five focus areas which are brain activity, skin conductance, cardiovascular, muscle activity, and eye measures. Each area has multiple methods for collecting those measures as shown in TABLE 6-3.

Table 6-3: Methods for Physiological Data Collection [14]

Measurement	Measurement Method
Brain Activity	electroencephalography (EEG)
	functional magnetic resonance imaging (fMRI)
Skin Conductance	skin conductance response (SCR)
	Galvanic skin response (GSR)
Cardiovascular measures	Heart rate (HR)
	Heart rate variability (HRV)
	Vasomotor activity
Muscle Activity	Electromyography (EMG)
	Electrogastrogram (EGG)
Eye Measures	Changes in pupil diameter (pupillometry)
	Electro-oculogram (EOG)

6.2 Methodology

The methodology described in this paper consisted of six main parts: development of a personality survey, development of before and after emotional state surveys, design of physiological data collection, design of driving simulator scenarios, classification of psychological data, and the study procedure. The personality survey measured each participants’ traits and tendencies pertaining to driving and everyday life. The before and after emotion surveys measured the participants’ current emotional state. Physiological data collection consisted of monitoring the body’s responses to the situations that are presented to the participants. The driving simulator scenarios presented the participants with various driving situations that were designed to induce different emotional states. The classification of psychological data presents the process used to extract the participant’s emotion from the emotional state surveys. The study procedure describes the steps of the experiment used to collect data from the participants.

6.2.1 Personality Survey

The personality survey aimed to capture a person’s proneness to be affected during different situations. The survey was constructed by combining the driving anger scale, State Trait Anger Expression Inventory, hostility index, and manifest anxiety scale. These four inventories

were chosen because they capture and quantify the personality traits that result in high amounts of anger and anxiety.

6.2.1.1 *Driving Anger Scale*

The Driving Anger Scale consists of 14 questions that present real driving situations that are likely to cause anger. The participants responded to these questions with a rating between 1 and 5 for how angry they would get in that situation. For example, a participant was given the question “A slow vehicle on a mountain road will not pull over and let people by.” and they responded with a 5 if that makes them very angry or a 1 if that would not affect them at all. The responses to all 14 questions were summed together to rate the participants on the scale with higher scores representing angrier drivers.

6.2.1.2 *State-Trait Anger Expression Inventory*

The State Trait Anger Expression Inventory measures the participant on multiple aspects of anger. State anger refers to anger that occurs in the moment, much like the situations presented in the Driving Anger Scale. Trait anger is a measure of a person’s proneness to anger. Anger control is a measure of a person’s ability to control their anger. Anger out is a measure of how a person chooses to vent their anger. Anger in is a measure of how a person chooses to hold their anger inside. This inventory consists of a series of statements that the participants answered as true if the statement described them or false if the statement did not describe them. For example, a participant was given the statement “I am quick tempered” and responded with true if they were quick tempered.

6.2.1.3 *Cook-Medley Hostility Scale*

The Cook-Medley Hostility Scale measures a person’s distrust of others. For example, a participant was given the statement of “Someone has it in for me” and they responded with true if the statement described them.

6.2.1.4 *Taylor Manifest Anxiety Scale*

The manifest anxiety scale measures anxiety in a person as a personality trait. For example, a participant was given the statement “At times I have lost sleep over worry” and responded with true if the statement described them.

6.2.2 *Emotion Surveys*

The emotion survey was based upon the Semantic Differential Scale. The scale measures the emotional state of the participants through 18 questions that assess pleasure, arousal, and dominance. Pleasure is the positive or negative attitude of the emotion. Arousal is the degree to which the emotion is felt. Dominance is whether the emotion makes a person feel empowered or belittled. Anger is an emotion that has: negative pleasure, high arousal, and high dominance. Anxiety is an emotion that has: negative pleasure, high arousal, and low dominance.

6.2.3 *Physiological Data Collection*

Physiological data were collected using a Psychlab system, shown in Figure 6-1. This system can accommodate a wide variety of sensors, but the three sensors used for this study were: ekg (placed on the chest), skin conductance (placed on the hands), and respiration (placed around the torso). The EKG measures the electrical activity of the heart. When a physiological reaction

occurs the heart rate will either increase or decrease according to the situation. The skin conductance sensor measures the electrical conductance of the skin. During a fight or flight response, the sweat glands become more active and this would register as an increase in skin conductivity. The respiration sensor measures the expansion and compression of the chest while breathing in and out. A typical calming mechanism for dealing with emotions is to take deep breaths while anxiety is known to cause short, shallow breaths.

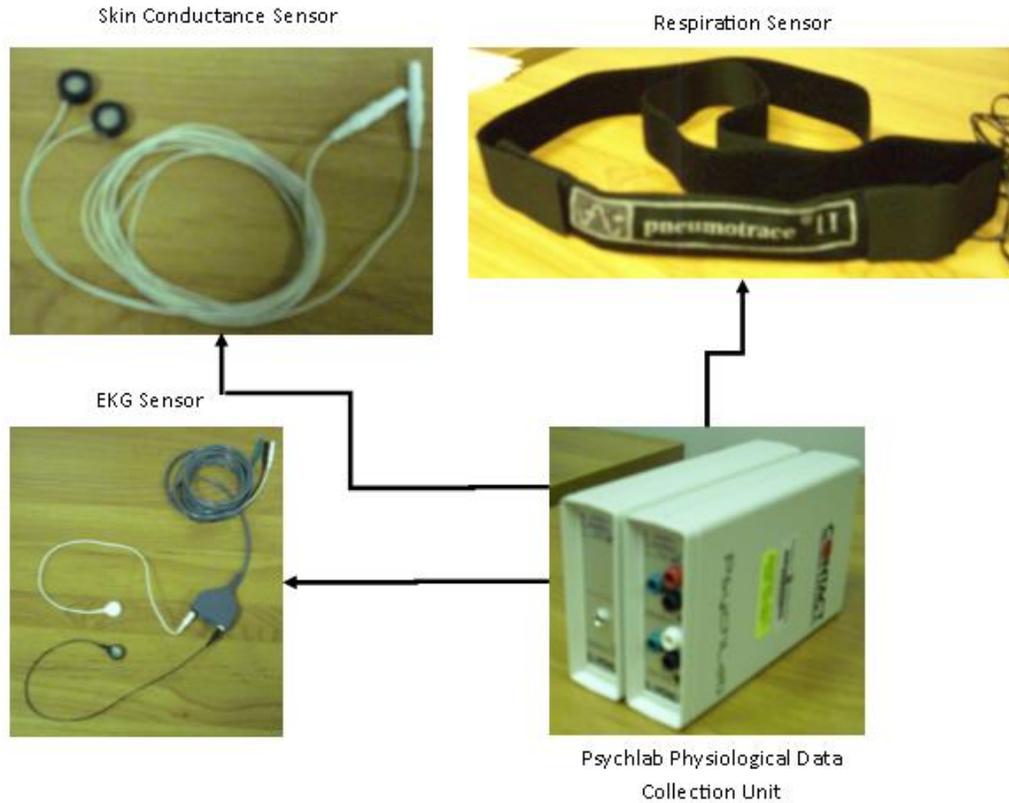


Figure 6-1: Psychlab Physiological Data Collection System

The data were collected using the Psychlab Acquire software [15], shown in Figure 6-2. The top line on the left of the figure (red) is the electrical signal being received by the ekg sensor. The line below that (green) is the measured electric conductivity of the skin. The third line (blue) is an event marker that can be changed during data collection by pressing the keys F1 through F9. The fourth line (second red line) is the measurement of the rise and fall of the chest from respiration. The fifth line (second green) is another event marker.

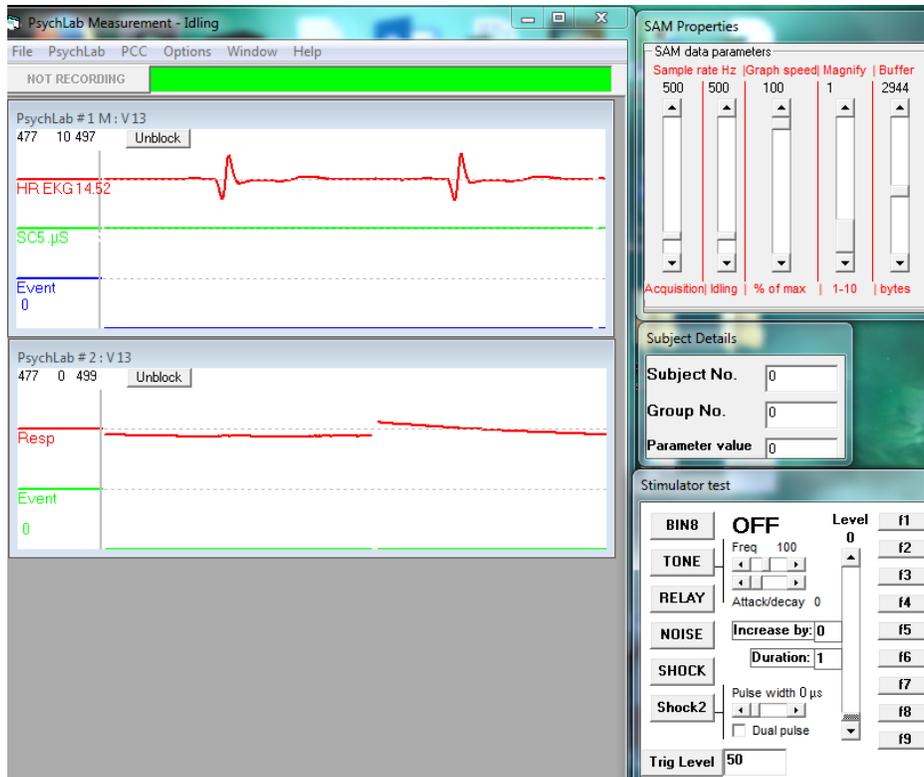


Figure 6-2: PsychLab Acquire Software

6.2.4 Driving Simulator Scenarios

The driving simulator used in this framework was the Drive Safety DS-250, shown in Figure 6-3. This simulator contained a driver's seat with a dashboard to better immerse drivers in the simulated environment. The simulated environment was presented to the drivers on three computer screens that were located on top of the dashboard. The three screens presented a wide view of environment in front of the driver along with both side mirrors (bottom left and right of left and right screens in Figure 6-3) and a rear view mirror (top right of middle screen in Figure 6-3).



Figure 6-3: Picture of Drive Safety Driving Simulator

6.2.4.1 Scenario Design

There were three scenarios designed for this study: (1) a base scenario, (2) an anxiety scenario, and (3) an anger scenario. All of the scenarios were designed such that the participant started in park at the top on an on-ramp, drove on a freeway for 4-5 minutes, and then exited the freeway using an off-ramp. While driving on the freeway, the participant interacted with a moderate amount of traffic that was coded to exhibit certain behaviors according to the scenario. For example, in the anger scenario the traffic was coded to try to impede the participant and act aggressively towards the participant. Figure 6-4 shows an aerial view of the simulated driving environment.

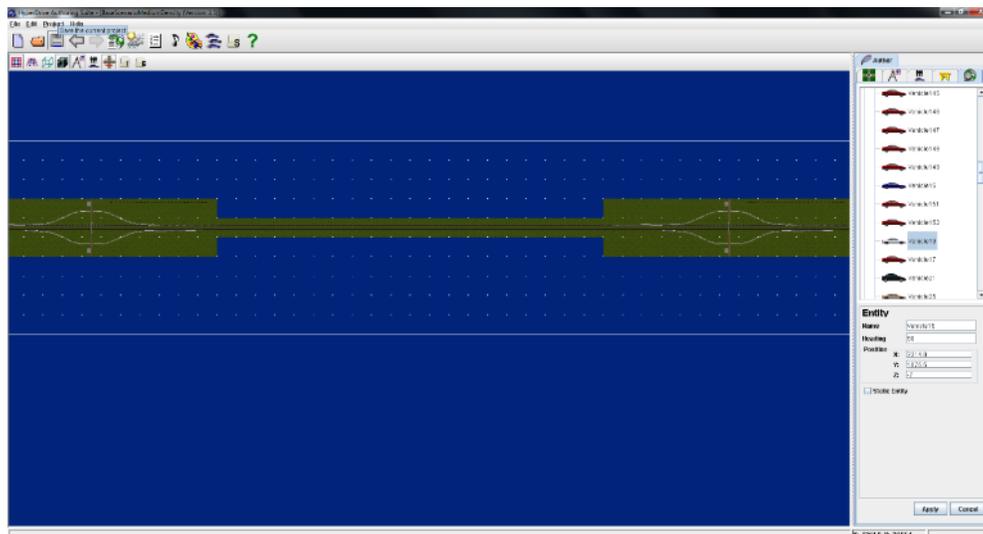


Figure 6-4: HyperDrive Screenshot

6.2.4.1.1 Base Scenario

The base scenario was coded such that the participant interacted with a moderate amount of traffic without the application of external pressure. This scenario also collected the baseline

physiological data for each participant. The baseline data was collected while the participant was driving to include the potential effects that the physical act of driving would have the physiology of the participants.

6.2.4.1.2 Anxiety Scenario

The anxiety scenario was designed to apply pressure on the participant. This pressure was applied by first giving them a time limit that will display in white if they are on time or in red if they are going to be late. Switching between different colors gave the participant immediate feedback about their performance rather than them having to estimate if they were on time. This also applied pressure to the participants any time that they were running late throughout the scenario. Additional pressure was applied to the participants by placing police vehicles in the scenario that would give chase to the participant if they were speeding. The real trick to the scenario to fully induce anxiety in the participants was that in order to not be late, they had to risk being caught speeding by the police vehicles.

6.2.4.1.3 Anger Scenario

The anger scenario was designed to capitalize on the increased arousal from the anxiety scenario by coding the surrounding traffic to exhibit annoying behaviors. The participants were told that the difference between the anxiety and anger scenarios were that there are no more police, but the other vehicles will be more aggressive. This scenario included coding that caused the traffic surrounding the participant reactive to the participant's actions. The vehicle in front of the participant was coded to slow down and travel below the speed limit. At the same time, the vehicle behind the participant was coded to speed up, tail-gate the participant, and periodically honk the horn at the participant. Also, if the gap between the participant and the vehicle in front of them was sufficient, then vehicles would periodically weave into the gap effectively cutting the participant off. Some of the actions of the surrounding traffic were coded to occur periodically to avoid the participant anticipating the events.

6.2.5 Classification of Psychological Data

The psychological data collection method observed the pleasure, arousal, and dominance of a participant's emotional state. This data were classified into discrete emotions using a model developed by Lang [16]. This model, shown in Figure 7-1, divided the Pleasure and Arousal space into discrete emotions. In this model: anger is the combination of low valence and high arousal, stressed or anxiety is the combination of neutral valence and high arousal, happy is high valence, and neutral is neutral valence and medium to low arousal.

6.2.6 Study Procedure

The participants were given the personality survey at the beginning of the study. Next, they would drive through a few adaptation scenarios that are designed to adapt the driver to the driving simulator. This was required so that the data collected at the beginning of the driving scenarios did not reflect the participants adjusting to the simulator which would have skewed the results. The next step was preparation for physiological data collection which included placing the ekg, skin conductance, and respiration sensors on the participant. The participants were then given the first psychological survey to capture their starting emotional state. The participants were then asked to complete one run of the base scenario. During the scenario, multiple variables, shown in

TABLE 3-4, were collected that pertain to driver behavior, the environment surrounding the driver, and the driver themselves. After the scenario, the participants were given another psychological survey to capture their emotional state after the scenario. This process of before psychological survey, scenario run, and after psychological survey was then repeated for the rest of the scenarios. Each participant took roughly 90 minutes to complete the study.

The sample for this study included 20 drivers with ages ranging from 20 to 83 and an average age of 39.8. There were eight females (40%) and twelve males (60%) that participated in the study. Most of the participants were originally from a variety of states and cities along the east coast of the United States.

Table 6-4: Simulator Data Variables

Variable	Definition
EntityAccel	The acceleration of the lead vehicle (m/s ²).
EntityVelocity	The speed of the lead vehicle (m/s).
HeadwayDist	The distance from the subject's front bumper to the rear bumper of the vehicle ahead (m).
HeadwayTime	The time to the vehicle ahead (s).
LaneIndex	The number of the lane in which the subject is travelling.
LanePos	The lane offset within the current lane (m).
LatAccel	The lateral acceleration of the subject vehicle (m/s ²).
LongAccel	The longitudinal acceleration of the subject vehicle (m/s ²).
SubjectHeading	The subject vehicle's current heading (deg).
SubjectX	The subject vehicle's current X position (m).
TTC	The time to collision with the vehicle ahead (s).
VehAhead	The unique name of the lead vehicle.
Velocity	The speed of the subject vehicle (m/s).
EKG	The electrical activity of the subject's heart.
Skin Conductance	The electrical conductivity of the subject's skin.
Respiration	The expansion of the subject's chest due to the lungs being filled with air.

6.3 Results

Figure 6-5 shows that there was a drastic difference in car-following behavior between two different emotional states of the same driver. The blue line is data collected from the first base scenario and the red line is data collected from the second anger scenario. It is interesting to note that even when this driver was happy, the following distance was low with a minimum distance close to 10 meters. Also, when the driver was angry, they engaged in tailgating with most of the car-following period occurring with a headway of less than 5 meters. The angry car-following period showed signs that the driver felt trapped behind the lead vehicle.

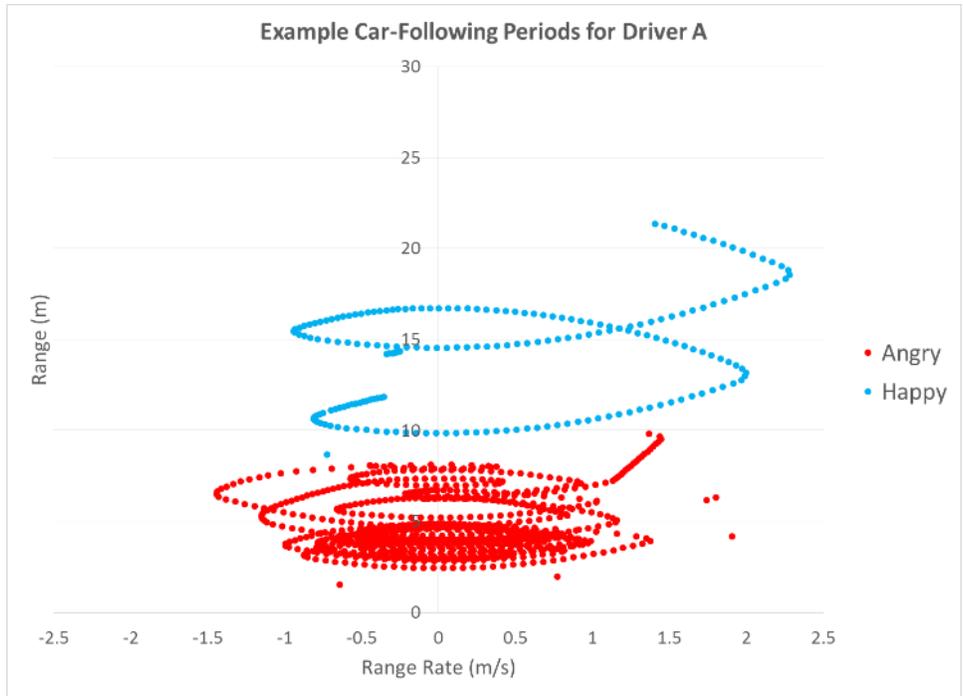


Figure 6-5: Comparison of Example Car-following Periods for Driver A

Figure 6-6 shows that the behavior of Driver B is different between different emotional states and different from Driver A. The blue line is data collected from the first base scenario and the red line is data collected from the second anger scenario. The largest difference was that Driver A tended to have short car-following periods when they were happy while Driver B tended to have long car-following periods when they were happy. Another big difference was that the two drivers responded differently to the anger scenario. Driver A became angry and then engaged in tailgating while Driver B transferred to an anxious emotional state.

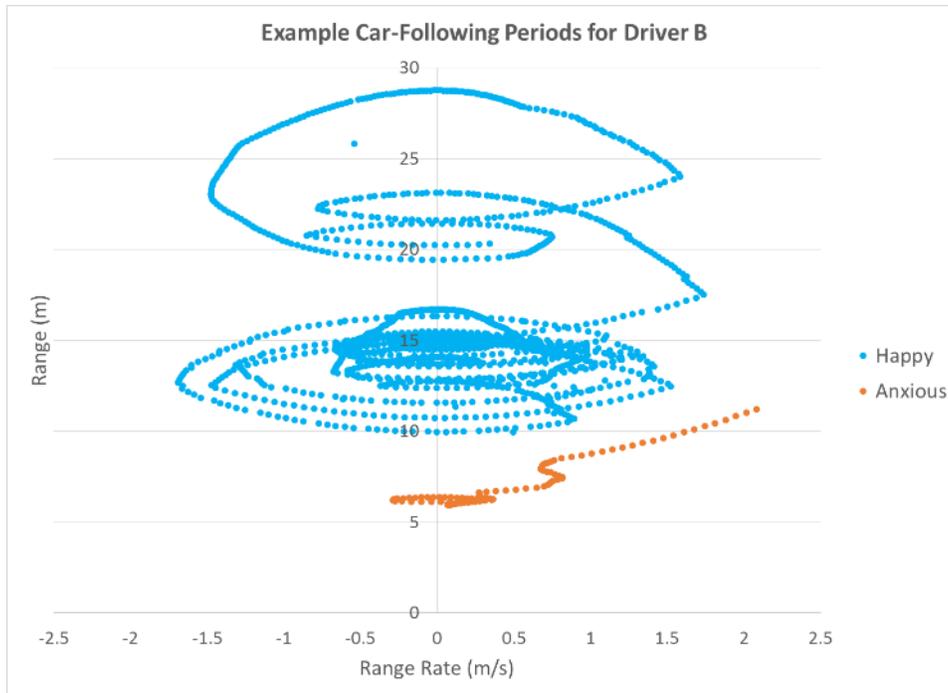


Figure 6-6: Comparison of Example Car-following Periods for Driver B

6.4 Discussion and Implications

The results of the two drivers showed that different drivers had different emotional responses to the same stimuli. This could be attributed to personality differences between the drivers that in this case served as a resistance to becoming angry. It is also interesting to note that while Driver B did not become angry, the minimum headway distance still decreased.

The implications to these findings are that it is necessary to consider both personality traits and emotional states when studying driver behavior. Personality traits can be measured using various psychological inventories, but emotional states have to be induced before their effects can be measured. This poses a problem because some emotions can be difficult to induce like sadness and extreme anger. The other difficulty that emotions present is that participants in a study begin with different emotional states. Thus, some participants may start in an emotional state that makes them very resilient to becoming angry while other participants may start in an emotional state that makes them very susceptible to becoming angry. Emotions and personality traits pose many difficulties to the study of driver behavior, but they also are the key to understanding many driving behaviors.

6.5 Conclusions

This research effort investigated the effects of personality and emotion on driver behavior. The methodology had four main components: development of psychological personality surveys, development of psychological emotion surveys, physiological data collection, and development of simulator scenarios. The personality surveys measured the tendency of each participant towards anger or anxiety. The emotional surveys measured the emotional states of the participants in terms of pleasure, arousal, and dominance. Pleasure refers to the positive or negative attitude of an emotional state. Arousal is how strongly that the emotional state is felt. Dominance refers to whether the emotion make the participant feel dominant or submissive. There were three simulator

scenarios used in this study: a base scenario, an anxiety scenario, and an anger scenario. The base scenario allowed the participants to interact without any external pressure being applied to them. The anxiety scenario applied pressure to the participants through a time limit and police vehicles patrolling the roadway. The anger scenario applied pressure to the participants through aggressive actions by the surrounding traffic.

The results showed that there can be drastic differences in driver behavior in different emotional states. The tendencies were that drivers became more aggressive and showed smaller headways when they were angered or anxious. The results also showed that drivers had different emotional responses to the same stimuli. Future research should investigate the similarities and differences in driver behavior that stem from personality traits. These results showed that this methodology was very effective in inducing and measuring the emotional responses of drivers to various driving situations.

References

- [1] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
- [2] P. S.-T. Chen and H. S. Mahmassani, "Dynamic interactive simulator for studying commuter behavior under real-time traffic information supply strategies," *Transportation Research Record*, pp. 12-21, 1993.
- [3] R. Hoogendoorn, S. Hoogendoorn, K. Brookhuis, and Winnie Daamen, "Mental Workload, Longitudinal Driving Behavior and Adequacy Of Car Following Models In Case of Incidents In The Other Driving Lane," *TRB 2010 Annual Meeting CD-ROM*, p. 21, 2010.
- [4] (2013, February 20, 2014). *Traffic Analysis Tools: Next Generation Simulation (NGSIM)*.
- [5] S. Ossen, S. Hoogendoorn, and B. Gorte, "Interdriver Differences in Car-Following: A Vehicle Trajectory-Based Study," *Transportation Research Record*, vol. 1965, pp. 121-129, 2006.
- [6] W. W. Cook and D. M. Medley, "Proposed hostility and Pharisaic-virtue scales for the MMPI," *Journal of Applied Psychology*, vol. 38, pp. 414-418, 1954.
- [7] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, pp. 49-59, 1994.
- [8] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*. Cambridge, MA, US: The MIT Press, 1974.
- [9] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273-294, 9// 1977.
- [10] J. A. Taylor, "A personality scale of manifest anxiety," *The Journal of Abnormal and Social Psychology*, vol. 48, p. 285, 1953.
- [11] C. D. Spielberger and S. J. Sydeman, "State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory," in *The use of psychological testing for treatment planning and outcome assessment*, ed Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 1994, pp. 292-321.
- [12] J. L. Deffenbacher, E. R. Getting, and R. S. Lynch, "DEVELOPMENT OF A DRIVING ANGER SCALE," *Psychological Reports*, vol. 74, pp. 83-91, 1994/02/01 1994.
- [13] *Nordic Road and Transport Research*, vol. 16, 2004.
- [14] J. L. Andreassi, *Psychophysiology: Human behavior and physiological response*: Psychology Press, 2000.
- [15] "PsychLabAcquire," ed. Cambridge, UK: Contact Precision Instruments, 2009.
- [16] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, attention, and the startle reflex," *Psychological Review*, vol. 97, pp. 377-395, 1990.

7. Development of an Emotional Car-Following Model⁶

Abstract

This research effort proposed a new car-following model that included the effects of emotion on driver behavior. The methodology used to investigate the effects of emotion on driver behavior had three main components: a psychological personality survey, a psychological emotion survey, and simulator scenarios. There were three simulator scenarios used in this study: a base scenario, an anxiety scenario, and an anger scenario. The design of the study was such that the scenarios induced certain emotions in the participants and those emotions were measured using psychological emotion surveys. The base scenario allowed the participants to interact without any pressure being applied to them. The anxiety scenario applied pressure to the participants through a time limit and police vehicles patrolling the roadway. The anger scenario applied pressure to the participants through aggressive actions by the surrounding traffic. The findings showed that emotions caused drivers to change the distribution of their driving behaviors. The new car-following model used a markovian process to account for the different distributions of behavior for each emotional state.

⁶ Paper has been submitted for publication in the proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC 2014)

7.1 Introduction

The variability of driver behavior is one of the most difficult aspects to consider in modeling. The real challenge to conquering this difficulty is that the variability can stem from numerous sources including: emotion, hunger, sensitivity to light conditions, personality, thirst, etc. [1]. Emotion is potentially the most influential source of variability as evidenced by the existence of road rage and driving anger. Road rage is very intriguing as it is a combination of personality traits and emotional state. Road rage occurs more frequently in some people versus others which indicates that some personalities are more resistant to road rage than others. Also, different personalities have different ways of venting their anger. For example, some people express their anger verbally (yelling and cursing) while others express it physically (erratic accelerations), and in some cases people focus their anger in a vengeful manner (attacking a certain vehicle).

This effort proposed a new car-following model that captured the effects of emotion on driver behavior. The data for this effort were collected by driving simulator scenarios that were designed to induce emotions in the participants. In the following sections, we first present a synthesis of past effort, followed by a description of our methodology, and finally we present a discussion of the results and our conclusions.

7.1.1 Synthesis of Past Efforts

The past efforts that influenced this work are summarized by previous work [2-13] that used two different types of data: naturalistic [14] and driving simulator [13, 15]. Naturalistic data refers to data collected from drivers during their normal daily routines. This type of data offers the opportunity to study individual drivers in depth over the course of multiple days, weeks, or months. The disadvantage to this type of data is that it focuses on the individual driver and does not collect data to describe the overall traffic environment (macroscopic data). Driving simulator studies, like naturalistic data, are focused on the individual driver, but the driver can be presented with dangerous driving situation without endangering the driver. Previous research [3, 8, 11] also compared various car-following models identifying the properties, strengths, and weaknesses of the models. The development of car-following models [2, 4-6] used the findings of the strengths and weaknesses of existing models to develop new models that would: (1) remove the weaknesses of existing models or (2) combine the strengths of two different models. The findings of both the comparison of car-following models and the development of car-following models showed a large amount of variability within and between drivers that, upon further investigation, revealed that driving behavior changes: (1) between drivers, (2) between car-following periods, and (3) within a car-following period [7]. These findings lead to two efforts: the development of a car-following data segmentation and clustering method [7, 9], and the development of a psychophysiological driving simulator study [13]. The segmentation and clustering method divides car-following periods into segments of similar behavior then clusters the segments to find unique state-action clusters of behavior. The psychophysiological driving simulator study was developed on the hypothesis that emotions explain the variability that was observed in driver behavior. The simulator study used psychological inventories to collect and quantify the participants' personality and emotional state.

7.1.1.1 *Psychological Inventories*

Psychological inventories measure either personality traits or emotional states. Personality traits are the tendencies of a person in everyday situations. Emotional states are the emotions that are being felt in the moment. For example, when a person rates high in trait anger (personality),

they are more likely to become angry (state) than others in situations that are irritating. Table 7-1 shows a list of psychological inventories and gives a description of what they measure.

Table 7-1: Psychological State and Trait Inventories

Inventory	Description	State/Trait
Cook-Medley Hostility Scale [16]	Measures a person’s negative views of social interaction (distrust, aggression, etc.)	Trait
Semantic Differential Scale [17, 18]	Measures a person’s current emotional state in terms of: pleasure, arousal, and dominance.	State
Taylor Manifest Anxiety Scale [19]	Measures how anxiety manifests in a person’s daily life.	Trait
State-Trait Anger Expression Inventory [20]	Measures five aspects of anger: State anger, trait anger, anger in, anger out, and anger control.	State and Trait
Driving Anger Scale [21]	Measures how much anger different driving situations induce in a person.	Trait

7.2 Methodology

The methodology used in this research effort consisted of four main parts: driving simulator experimental design, car-following data segmentation and clustering, and psychological data classification. The driving simulator experimental design describes the various data collection methods used and the reasoning behind the experimental design. The segmentation and clustering of car-following data describes the process that was used to divide the car-following data into clusters of similar behavior. Psychological data classification describes the process used to classify the collected psychological data into discrete emotions.

7.2.1 Driving Simulator Experimental Design

The following section gives a brief overview of the driving simulator experimental design that was developed in a parallel research effort [13]. The driving simulator experimental design consisted of four main parts: development of a personality survey, development of before and after emotional state surveys, design of driving simulator scenarios, and the study procedure.

7.2.1.1 Personality Survey

The personality survey aimed to capture a person’s proneness to be affected during different situations. The survey was constructed by combining the Driving Anger Scale, State-Trait Anger Expression Inventory, Cook-Medley Hostility Scale, and the Taylor Manifest Anxiety Scale. These four inventories were chosen because they capture and quantify the personality traits that result in high amounts of anger and anxiety.

7.2.1.2 Emotional State Survey

The emotion survey was based upon the Semantic Differential Scale. The scale measures the emotional state of the participants through 18 questions that assess pleasure, arousal, and dominance. Pleasure is the positive or negative attitude of the emotion. Arousal is the degree to which the emotion is felt. Dominance is whether the emotion makes a person feels empowered or belittled.

7.2.1.3 *Driving Simulator Scenarios*

The driving simulator used in this framework was the Drive Safety DS-250. This simulator contained a driver's seat with a dashboard to better immerse drivers in the simulated environment. The simulated environment was presented to the drivers on three computer screens that were located on top of the dashboard. The three screens presented a wide view of environment in front of the driver along with both side mirrors and a rear view mirror.

7.2.1.3.1 *Scenario Design*

There were three scenarios designed for this study: (1) a base scenario, (2) an anxiety scenario, and (3) an anger scenario. The base scenario was coded such that the participant can interact with a moderate amount of traffic without the application of external pressure. The anxiety scenario was designed to apply pressure on the participant through a time limit and the threat of being chased by a police car. The real trick to fully induce anxiety in the participants was that in order to finish within the time limit, they had to risk being caught speeding by the police vehicles. The anger scenario was designed to irritate the participants through aggressive behaviors exhibited by the surrounding traffic. The aggressive behaviors included: trapping the participant behind a slow vehicle, tailgating the participant, periodically honking the horn at the participant, and cutting off the participant.

7.2.1.4 *Study Procedure*

The experiment began with each participant completing the personality survey. Then, the participants completed adaptation scenarios to make them more comfortable with the simulator. Next, the participants completed the first emotional state survey before beginning to drive the developed scenarios. The scenarios were presented to the participants in the following order: two base scenarios, two anxiety scenarios, and two anger scenarios. The participants completed another emotional state survey after each scenario. The scenarios were arranged in a specific way in order to better induce emotions in the drivers. For example, anxiety makes drivers tense and uneasy which causes them to be more likely to become angry. Each participant took roughly 90 minutes to complete the study. The sample for this study included 20 drivers with ages ranging from 20 to 83 and an average age of 39.8. There were eight females (40%) and twelve males (60%) that participated in the study.

7.2.2 *Car-Following Data Segmentation and Clustering*

The data collected during the scenarios were processed using the findings of previous research [9] that segmented and clustered car-following periods into clusters of similar behavior. This process was used because it enables analysis of the distribution of behavior for each emotional state. Comparison of the distributions was needed because the effects of emotions on driver behavior could have manifested as a shift in the percentage of occurrence of certain behaviors. For example, a driver known for being aggressive may sometimes tailgate the vehicle in front of them, but when this driver becomes angry, they tailgate every vehicle that is in front of them.

7.2.3 *Psychological Data Classification*

The psychological data collection method observed the pleasure, arousal, and dominance of a participant's emotional state. This data were classified into discrete emotions using a model developed by Lang [22]. This model, shown in Figure 7-1, divided the Pleasure and Arousal space into discrete emotions. In this model: anger was the combination of low valence and high arousal,

stressed or anxiety was the combination of neutral valence and high arousal, happy was high valence, and neutral was neutral valence and medium to low arousal.

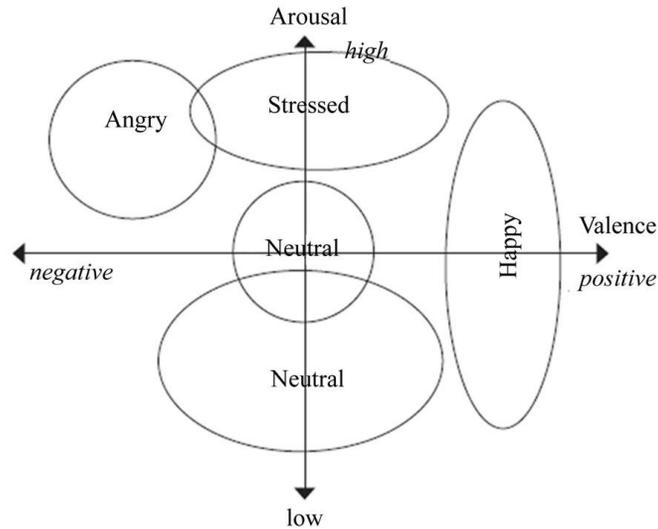


Figure 7-1: Locations of Various Emotional States in Valence (Pleasure) and Arousal Space [23]

7.3 Results

Figure 7-2 shows that the three main clusters of behavior were 20, 26, and 28. The figure also shows that anger had a very high concentration of cluster 28. The interesting point to note about Figure 7-2 is that the different emotional states appeared to have the same clusters, but with different distributions.

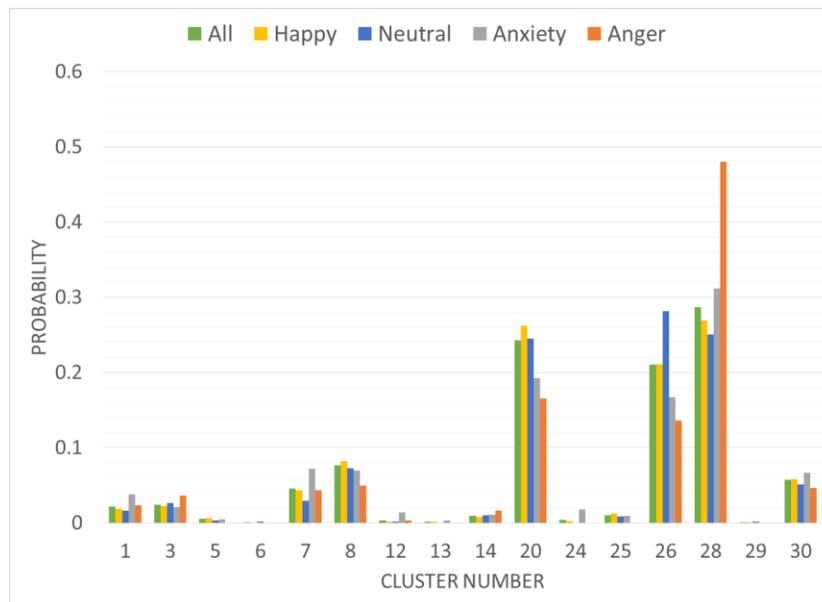


Figure 7-2: Cluster Distributions by Emotion

Figure 7-3 shows that clusters 20, 26, and 28 were distinguishable by the distance between the subject vehicle and lead vehicle. It is interesting to note that cluster 28 consisted of ranges that are mostly below 50 feet.

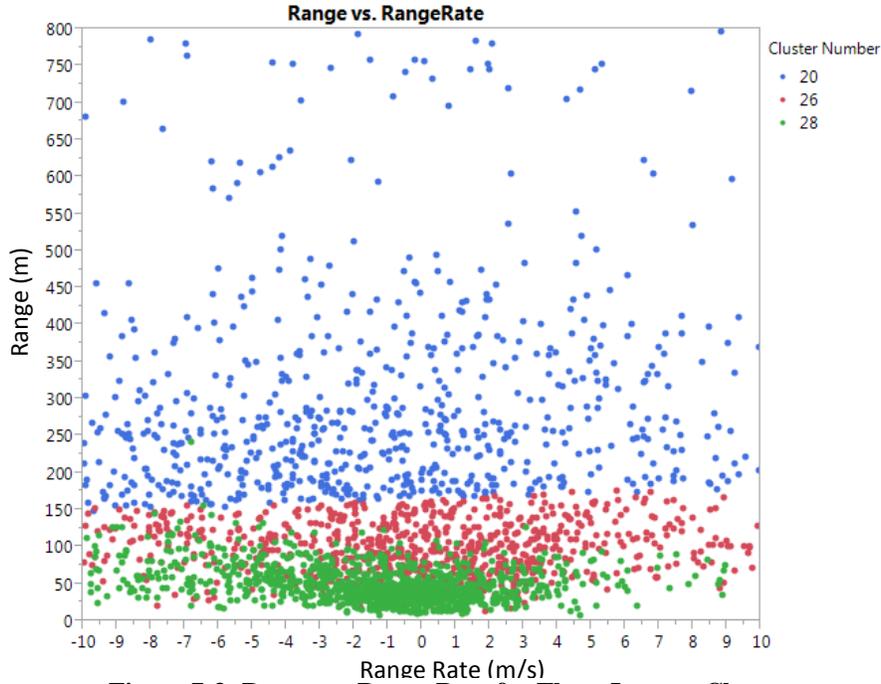
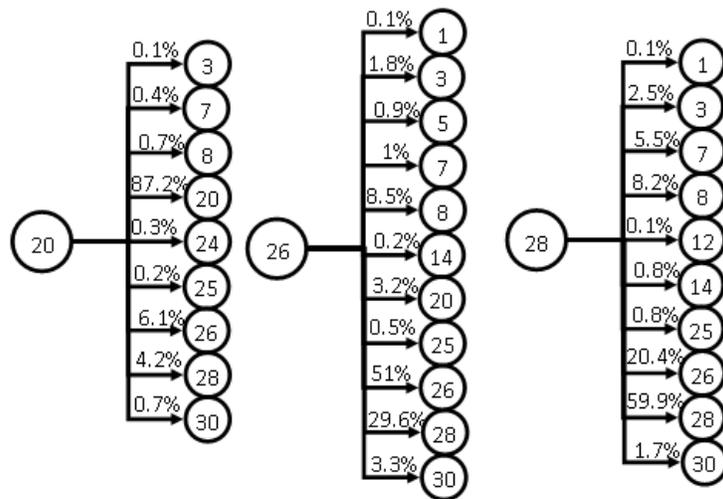


Figure 7-3: Range vs. Range Rate for Three Largest Clusters

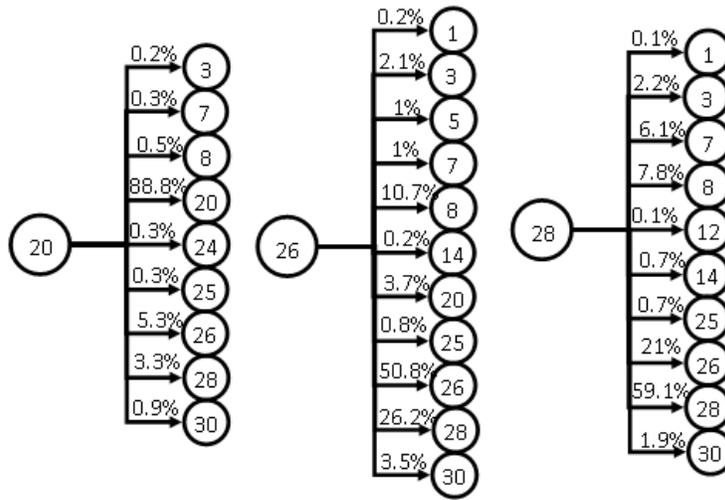
The next step was to translate the findings into a car-following model framework. A Markovian framework was chosen for this model because of the need to consider multiple clusters of behavior with varying probabilities. Utilizing a Markovian framework also added a dynamic that considered each driving state or cluster and the probabilities to transition into or out of that cluster. Figure 7-4 shows the transitions for the three largest clusters (20, 26, and 28) for different emotional states. It is interesting to note that the neutral and happy emotional states had a very high probability for remaining in cluster 26 while the angry and anxious emotional states had a very high probability to transition from cluster 26 to cluster 28.

All Emotions



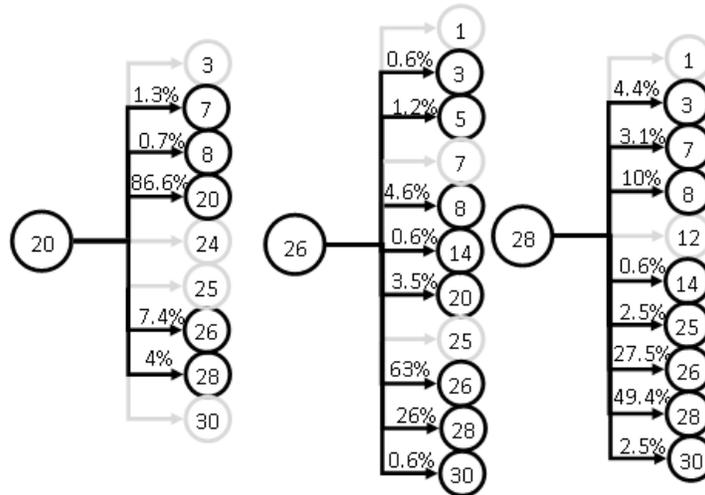
(a)

Happy



(b)

Neutral



(c)

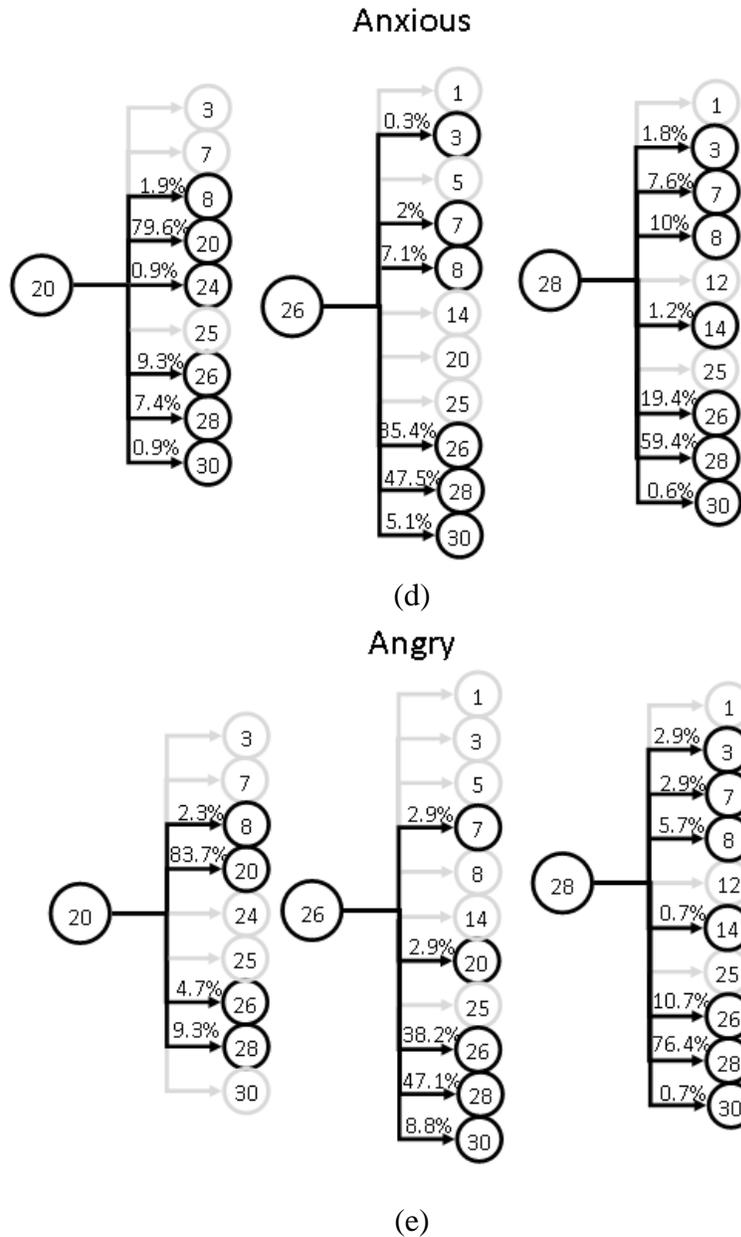


Figure 7-4: Transition Probability Trees for Multiple Emotions

7.4 Discussion and Implications

The results of this study showed that emotions cause drivers to change their behavior and the changes were dependent upon the emotion. It is also important to note that the change in behavior was not that emotions lead to new behaviors, but that emotions changed the distribution of behaviors. The implications to this finding are that modeling efforts can use a global set of behaviors to represent different drivers in different emotional states.

There are many different emotional states with the potential to affect driving behavior (boredom, sadness, excitement, etc.). The difficulty to considering these additional emotional states is that these emotions are difficult to induce and to measure. Induction of some emotional states can be dependent upon factors that are outside of the driving environment. It can also be difficult to distinguish the difference between emotional states.

Another implication is that emotions can be induced in drivers by other drivers. For example, anger was induced in the participants of the study by coding the car behind them to behave aggressively. This also indicates a troubling notion that aggressive driving can be contagious. It also appears that drivers have different resistance levels to becoming agitated while driving. There are many different reactions to agitation in drivers including: becoming aggressive towards other vehicles on the road (physical), and using words to express displeasure (verbal). Some drivers that choose to express their aggression physically focus their aggression on the vehicle that angered them (vengeance).

7.5 Conclusions

This research effort investigated the effects of emotion on driver behavior and captured those effects in a car-following model. The methodology used in this effort had three components: a personality survey, an emotional state survey, and driving simulator scenarios. There were three driving simulator scenarios developed for this study: a base scenario, and anxiety scenario, and an anger scenario. These scenarios were designed to induce emotions in the participants. The anxiety scenario forced participants to risk being caught speeding and chased by a police vehicle. The anger scenario used aggressive actions by the surrounding traffic to irritate the participants. The results showed that a variety of emotional states were observed during the study and the distribution of driving behaviors was affected by the emotional state. A Markovian process was used for the car-following model because it captured the effects that emotions made on the distribution of driving behaviors.

Future research should analyze the effects of personality traits on the distribution of behaviors and the distribution of emotional states. Future research should also observe additional emotional states and add those to the new car-following model. The results showed that the car-following model produced through this effort was very effective in capturing the changes in driver behavior that occurred in different emotional states.

References

- [1] S. K. Balogun, N. A. Shenge, and S. E. Oladipo, "Psychosocial factors influencing aggressive driving among commercial and private automobile drivers in Lagos metropolis," *The Social Science Journal*, vol. 49, pp. 83-89, March 2012.
- [2] M. Abbas, B. Higgs, A. Medina, and C. Y. D. Yang, "A Hybrid Wiedemann-GHR Model Calibration using Naturalistic Driving Data," in *13th International IEEE Conference on Intelligent Transportation Systems*, 2010.
- [3] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data," presented at the Transportation Research Board 90th Annual Meeting, Washington D.C., 2011.
- [4] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board 91st Annual Meeting, Washington D.C., 2012.
- [5] Bryan Higgs, Montasir Abbas, Alejandra Medina, and C. Y. D. Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," presented at the TRB Traffic Flow Theory Committee (TFTC) Summer Meeting and Greenshields Symposium, Annecy, France, 2010.
- [6] B. Higgs and M. Abbas, "A Cross-Validation of Car and Truck Agents Using Naturalistic Data: One-Parameter Car-Following Model for Reducing Computational Burden of Multi-Agent Systems," presented at the Conference on Agent-Based Modeling in Transportation Planning and Operations, Blacksburg, VA, 2013.
- [7] B. Higgs and M. Abbas, "A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, 2013, pp. 857-862.

- [8] B. Higgs and M. Abbas, "Multi-Resolution Comparison of Car-Following Models using Naturalistic Data," presented at the Transportation Research Board 93rd Annual Meeting, Washington D.C., 2014.
- [9] B. Higgs and M. Abbas, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns," *IEEE Intelligent Transportation Systems Transactions: Special Issue on Uncertainty in Computational Traffic Models*, 2014.
- [10] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [11] B. Higgs, M. M. Abbas, and A. Medina, "Naturalistic Car Driver Behavior: A Comparison of Car-Following Models," in *19th ITS World Congress*, 2012.
- [12] B. J. Higgs, "Application of Naturalistic Truck Driving Data to Analyze and Improve Car Following Models," Master's, Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, 2011.
- [13] B. Higgs and M. Abbas, "Experimental Design for a Psychophysiological Driving Simulator Study," presented at the 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 2014.
- [14] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
- [15] R. Hoogendoorn, S. Hoogendoorn, K. Brookhuis, and Winnie Daamen, "Mental Workload, Longitudinal Driving Behavior and Adequacy Of Car Following Models In Case of Incidents In The Other Driving Lane," *TRB 2010 Annual Meeting CD-ROM*, p. 21, 2010.
- [16] W. W. Cook and D. M. Medley, "Proposed hostility and Pharisaic-virtue scales for the MMPI," *Journal of Applied Psychology*, vol. 38, pp. 414-418, 1954.
- [17] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, pp. 49-59, 1994.
- [18] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, pp. 273-294, 9// 1977.
- [19] J. A. Taylor, "A personality scale of manifest anxiety," *The Journal of Abnormal and Social Psychology*, vol. 48, p. 285, 1953.
- [20] C. D. Spielberger and S. J. Sydeman, "State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory," in *The use of psychological testing for treatment planning and outcome assessment*, ed Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 1994, pp. 292-321.
- [21] J. L. Deffenbacher, E. R. Getting, and R. S. Lynch, "DEVELOPMENT OF A DRIVING ANGER SCALE," *Psychological Reports*, vol. 74, pp. 83-91, 1994/02/01 1994.
- [22] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, attention, and the startle reflex," *Psychological Review*, vol. 97, pp. 377-395, 1990.
- [23] H. Holzapfel, C. Fuegen, M. Denecke, and A. Waibel, "Integrating emotional cues into a framework for dialogue management," in *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, 2002, pp. 141-146.

8. Investigation of the Effects of Physiological States on Driver Behavior

Abstract

This research effort examined the effects of different physiological states on driver behavior. The methodology presented in this effort had two main parts: (1) clustering of physiological data and (2) comparison of clusters of physiological data. The sample for this study included data for ten drivers that were collected during a psychophysiological driving simulator study. The study used three different scenarios that were designed to invoke emotional and physiological responses in the participants. The clustering of physiological data grouped the data into clusters of similar physiological responses. Car-following data were used to compare the clusters to find the similarities and differences. The results showed that physiological changes in the driver caused differences in behavior that were exhibited as a different distribution of behaviors.

8.1 Introduction

Emotions can have a large impact on the behavior of drivers as evidenced by road rage. The difficulty studying the effects of emotion on driver behavior is that emotions are experienced within a person and only they can describe how they feel. Emotions can be measured using psychological surveys, but this would require halting data collection and it would only be measuring the participant's recollection of past events. The key to moving forward is the consideration of psychophysiological measures of the body. These measures capture the physiological changes that occur in the body due to changes in emotion. For example, when someone becomes angry, the following physiological changes occur: heart rate increases, pupils dilate, sweat glands increase production, etc.

This effort investigates the effects of different physiological states on driver behavior. The data for this effort were collected by a driving simulator study that was designed to induce emotions in the participants. In the following sections, we first present a synthesis of past effort, followed by a description of our methodology, and finally we present a discussion of the results and our conclusions.

8.1.1 Synthesis of Past Efforts

The past efforts that influenced this work are summarized by previous work [1-12] that used two different types of data: naturalistic [13] and driving simulator [12, 14]. Naturalistic data refers to data collected from drivers during their normal daily routines. This type of data offers the opportunity to study individual drivers in depth over the course of multiple days, weeks, or months. The disadvantage to this type of data is that it focuses on the individual driver and does not collect data to describe the overall traffic environment (macroscopic data). Driving simulator studies, like naturalistic data, are focused on the individual driver, but the driver can be presented with dangerous driving situation without endangering the driver. There were two main effort in previous research: the development of a car-following data segmentation and clustering method [6, 8], and the development of a psychophysiological driving simulator study [12]. The segmentation and clustering method divides car-following periods into segments of similar behavior then clusters the segments to find unique state-action clusters of behavior. The psychophysiological driving simulator study was developed on the hypothesis that emotions explain the variability that was observed in driver behavior. The simulator study used psychological inventories to collect and quantify the participants' personality and emotional state.

8.2 Methodology

The methodology presented in this effort consisted of a psychophysiological driving simulator study, clustering of physiological data, and a comparison of physiological clusters. The psychophysiological driving simulator study was conducted in previous research [12]. This study collected physiological data on participants during three different scenarios (base, anxiety, and anger) designed to induce emotions in the drivers. The base scenario was designed as a baseline for comparison to the other two scenarios. The anxiety scenario was designed to cause drivers to become anxious as they had to take risks to complete the scenario within the time limit. The anger scenario was designed to annoy drivers using car horns, being cut-off in traffic, and being tail gated. During these scenarios three measures were collected: EKG, respiration, and skin conductance.

K-means clustering was used to divide the physiological data into clusters of similar responses. Clustering the physiological data revealed how certain physiological states repeated

throughout the data. The statistical software, JMP [15], was used for this step, due to the fact that JMP has a set of tools and techniques for analysis that go beyond simple clustering.

The physiological changes that occurred are in response to different stimuli, thus data on the driving environment was required. The results of previous research [6, 8] found state-action clusters of driving behavior that were a combination of both the driving state and the action taken. The comparison of physiological clusters used the distribution of state-action clusters to reveal any similarities and differences.

8.3 Results

Figure 8-1 shows that there are three prominent clusters of physiological data, but there were also a large number of clusters that occur over 300 times. The figure shows that there were both rare and common physiological clusters.

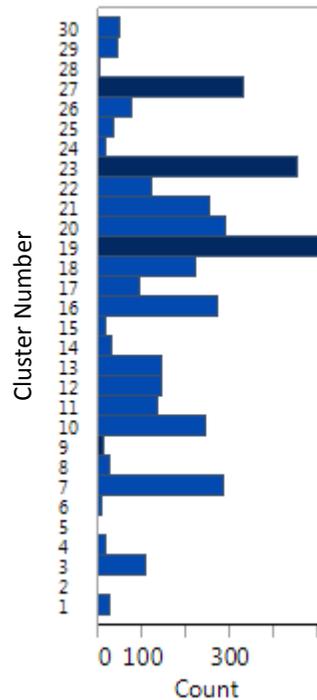


Figure 8-1: Distribution of Physiological Clusters

Figure 8-2 shows the physiological response of an example driver to a car horn being honked at them periodically (vertical blue lines). The observed changes in the R-R interval were in agreement with the findings of past research efforts [16] in that the R-R interval initially lengthens in response to stimuli (orienting response). The R-R interval then lowers indicating a faster heart rate which can be attributed to the driver being irritated or interpreting the stimuli as threatening (defensive response). It is interesting to note that after the heart rate spiked, it began to calm down to normal levels.

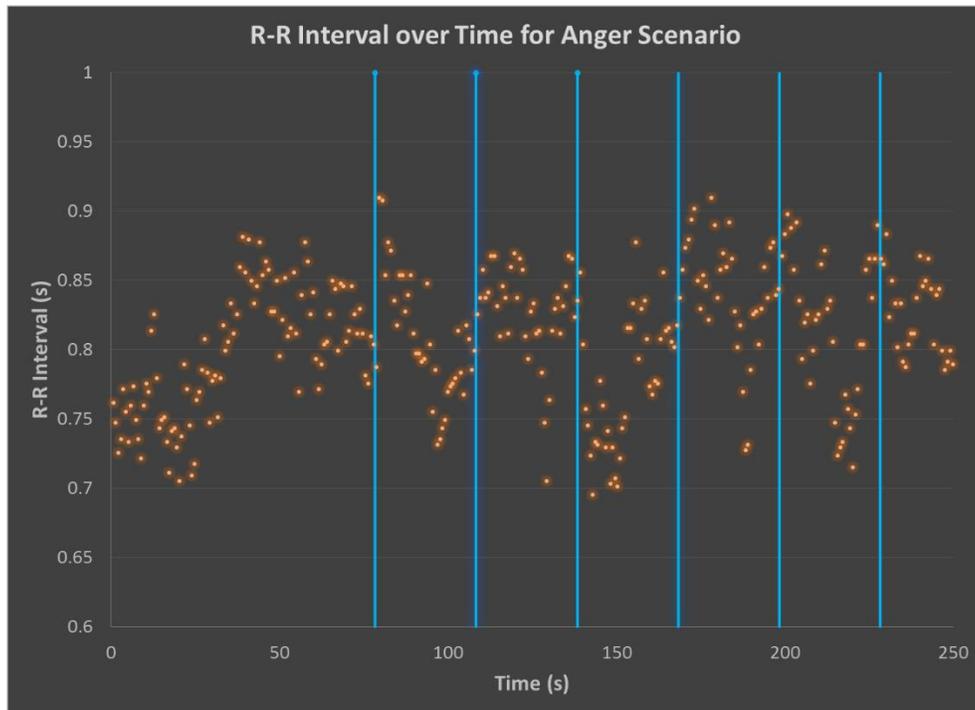


Figure 8-2: R-R Interval over Time for Anger Scenario

Table 8-1 shows that the physiological clusters tended to have high concentrations of state-action clusters 20, 26, and 28. It is interesting to note that the concentrations were different for all of the physiological clusters. Some of the physiological clusters (2, 5, 6, 16, 24, and 26) showed a high degree of partiality toward one state-action cluster.

Table 8-1: Distribution of State-Action Clusters for Each Physiological Cluster

	State-Action Cluster											
	1	3	5	7	8	14	20	25	26	28	30	
Physiological Cluster	1				0.08		0.19		0.19	0.54		
	2									1.00		
	3		0.04		0.02	0.09		0.14		0.27	0.41	0.03
	4		0.05			0.05		0.32		0.26	0.32	
	5									1.00		
	6									1.00		
	7	0.02	0.03		0.07	0.07	0.01	0.16		0.14	0.37	0.11
	8	0.04	0.04		0.08	0.08	0.04	0.08		0.20	0.28	0.16
	9							0.38		0.54	0.08	
	10	0.01	0.02	0.01	0.07	0.05	0.01	0.20		0.14	0.41	0.07
	11	0.01			0.01	0.07		0.24		0.34	0.29	0.03
	12	0.03	0.03		0.05	0.08	0.02	0.16		0.16	0.41	0.07
	13		0.02	0.01	0.03	0.16	0.01	0.10	0.01	0.37	0.22	0.05
	14			0.03		0.09		0.28		0.19	0.41	
	15					0.06		0.17	0.06	0.56	0.17	
	16	0.01	0.00	0.04	0.02	0.02	0.02	0.71	0.01	0.12	0.03	0.01
	17	0.03	0.03			0.02	0.01	0.18		0.14	0.55	0.04
	18	0.04	0.02		0.05	0.05		0.24	0.01	0.17	0.32	0.09
	19	0.03	0.03	0.00	0.08	0.07	0.01	0.21	0.01	0.19	0.27	0.08
	20	0.03	0.02		0.03	0.07	0.01	0.27	0.01	0.19	0.30	0.06
	21	0.01	0.04		0.04	0.09	0.00	0.22	0.00	0.20	0.31	0.06
	22	0.03	0.02		0.04	0.15	0.01	0.10	0.03	0.17	0.41	0.03
	23	0.02	0.02		0.01	0.10	0.01	0.27	0.02	0.34	0.17	0.03
	24		0.05			0.05				0.79	0.11	
	25		0.09	0.03	0.03		0.03	0.40		0.17	0.26	
	26	0.01		0.06	0.01	0.03		0.58	0.01	0.21	0.03	0.05
	27	0.03	0.04		0.05	0.14	0.02	0.14	0.02	0.16	0.34	0.04
	28				0.33		0.33			0.33		
	29		0.02	0.02	0.06	0.02		0.26		0.34	0.26	
	30	0.02	0.02		0.06	0.08	0.04	0.22		0.14	0.30	0.12

8.4 Conclusions

This research effort examined the similarities and differences in car-following behavior in different physiological states. The methodology used in this effort used data that were collected previously in a psychophysiological driving simulator study. The methodology included the clustering of physiological data into clusters of similar physiological states. These clusters were compared using the distributions of state-action clusters that were the result of a previous study. The results showed that driver’s heart rate increased when they were irritated by an aggressive

driver. The results also showed that most of the physiological clusters showed different distributions of driver behavior. Some of the physiological clusters also showed partiality towards a single state-action cluster. The implications to these findings are that physiological changes in the driver cause them to behave differently. The difference in behavior can manifest as exhibiting certain clusters of behavior or as exhibiting a different distribution of behaviors.

Future research should analyze the effects of personality traits on the correlation of physiological states and state-action clusters. Future research should also examine other methods to induce emotions in drivers. Future research should also examine the effects of different driving conditions on the emotional state of the driver.

References

- [1] M. Abbas, B. Higgs, A. Medina, and C. Y. D. Yang, "A Hybrid Wiedemann-GHR Model Calibration using Naturalistic Driving Data," in *13th International IEEE Conference on Intelligent Transportation Systems*, 2010.
- [2] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Comparison of Car-Following Models when Calibrated to Individual Drivers using Naturalistic Data," presented at the Transportation Research Board 90th Annual Meeting, Washington D.C., 2011.
- [3] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board 91st Annual Meeting, Washington D.C., 2012.
- [4] Bryan Higgs, Montasir Abbas, Alejandra Medina, and C. Y. D. Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," presented at the TRB Traffic Flow Theory Committee (TFTC) Summer Meeting and Greenshields Symposium, Annecy, France, 2010.
- [5] B. Higgs and M. Abbas, "A Cross-Validation of Car and Truck Agents Using Naturalistic Data: One-Parameter Car-Following Model for Reducing Computational Burden of Multi-Agent Systems," presented at the Conference on Agent-Based Modeling in Transportation Planning and Operations, Blacksburg, VA, 2013.
- [6] B. Higgs and M. Abbas, "A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, 2013, pp. 857-862.
- [7] B. Higgs and M. Abbas, "Multi-Resolution Comparison of Car-Following Models using Naturalistic Data," presented at the Transportation Research Board 93rd Annual Meeting, Washington D.C., 2014.
- [8] B. Higgs and M. Abbas, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns," *IEEE Intelligent Transportation Systems Transactions: Special Issue on Uncertainty in Computational Traffic Models*, 2014.
- [9] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [10] B. Higgs, M. M. Abbas, and A. Medina, "Naturalistic Car Driver Behavior: A Comparison of Car-Following Models," in *19th ITS World Congress*, 2012.
- [11] B. J. Higgs, "Application of Naturalistic Truck Driving Data to Analyze and Improve Car Following Models," Master's, Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, 2011.
- [12] B. Higgs and M. Abbas, "Experimental Design for a Psychophysiological Driving Simulator Study," presented at the 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 2014.
- [13] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, *et al.*, "The 100-car naturalistic Driving Study; Phase II - Results of the 100-Car Field Experiment," Virginia Tech Transportation Institute, Blacksburg, VA DOT HS 810 593, 2006.
- [14] R. Hoogendoorn, S. Hoogendoorn, K. Brookhuis, and Winnie Daamen, "Mental Workload, Longitudinal Driving Behavior and Adequacy Of Car Following Models In Case of Incidents In The Other Driving Lane," *TRB 2010 Annual Meeting CD-ROM*, p. 21, 2010.
- [15] "JMP," Pro 10 ed. Cary, NC: SAS Institute Inc., 1989-2013.

- [16] F. K. Graham and R. K. Clifton, "Heart-rate change as a component of the orienting response," *Psychol Bull*, vol. 65, pp. 305-20, May 1966.

9. Development of an Emo-Psychophysical Car-Following Model

Abstract

This research effort proposed a new car-following model that incorporated the influence of emotions into an existing psychophysical car-following model. The methodology presented in this effort improved the hybrid Wiedemann-GHR car-following model by the calibration to car-following data obtained in different emotional states. The sample for this effort was ten drivers that participated in a driving simulator study. The participants drove through three different scenarios that were designed to induce an emotional response. The results showed that the thresholds of the Wiedemann-GHR model change depending on the emotional state. The new car-following model was very effective in representing the emotional behavior of drivers.

9.1 Introduction

The accuracy of current car-following models is compromised by one key element that is not fully considered, the human element. There is wide variability within human drivers that goes unexplained by current car-following modeling techniques resulting in high error values and thus inaccurate traffic simulations. The variability of driver behavior is one of the most difficult aspects to consider in modeling. The real challenge to conquering this difficulty is that the variability can stem from numerous sources including: emotion, hunger, sensitivity to light conditions, personality, thirst, etc. Emotion is potentially the most influential source of variability as evidenced by the existence of road rage and driving anger.

This effort proposed a new car-following model that added the ability to account for differences in behavior caused by emotion to an existing car-following model. The data for this effort were collected by a driving simulator study that was designed to induce emotions in the participants. In the following sections, we first present a synthesis of past effort, followed by a description of our methodology, and finally we present a discussion of the results and our conclusions.

9.1.1 Synthesis of Past Efforts

The past research that influenced this effort is divided into two categories: research into car-following models and research into the effects of emotion on car-following behavior. The research into car-following models gives a brief overview of the various car-following models that have been developed. The research into the effects of emotion on car-following behavior gives a brief overview of the recent developments that have been made in the development of car-following models that consider the emotional state of the driver.

9.1.1.1 *Car-Following Models*

Car-following models are designed around stimulus response, where the model processes various stimuli and generates driver actions for those stimuli. The main categories of car following models are: action-point or psychophysical models, linear models, non-linear models, safe distance models, and combination models. Action-point or psychophysical models divide car-following periods into different regimes that represent the drivers intended action. For example, a driver intending to follow a lead vehicle will accelerate and decelerate in an attempt to maintain a desired following distance. Linear and non-linear car-following models are very similar in that they mainly process the headway and difference in speed between the lead and following vehicles with some calibration parameters to create the action of the subject vehicle usually in terms of the vehicle's acceleration. Safe distance models are related to linear and non-linear models except that the driver actions are set by an intention to maintain an acceptable following distance. Combination models are the car-following models that combine characteristics of both psychophysical and non-linear or linear models.

9.1.1.1.1 Action-Point or Psychophysical Models

Research into action-point or psychophysical models has been focused on finding and evaluating the action-points or thresholds of the models. The popular models of this category are the Wiedemann model [1, 2] and Fritzsche [3] model both of which are used in micro-simulation software packages, VISSIM [4] and PARAMICS [5] respectively.

Research into the Wiedemann models is mainly focused on evaluating the thresholds through calibration or comparison to the action points. Higgs et al. [6] presented a calibration of

the Wiedemann model over different speed categories. The findings were that each speed category resulted in different parameters and thus different thresholds. The findings also supported that the differences in the thresholds at different speeds were driver specific. Brackstone et. al. [7] used data collected with an instrumented vehicle that was assembled at TRG Southampton to parameterize the Wiedemann's threshold for a typical following spiral. As a result they represent the action points as a function of a probability distribution based on ground speed. Menneni et al [8] presented a calibration methodology of the VISSIM Wiedemann car-following model based on integrated use of microscopic and macroscopic data using NGSIM. Relative distance vs. relative speed graphs were used for the microscopic calibration, specifically to determine the action points (it is important to note that action points were not identical to perception threshold). Scatter and distribution of action points on relative distance versus relative velocity graphs also showed similarity in driver behavior between the two freeways.

Hoogendoorn and Hoogendoorn [9] analyzed action points based upon a piecewise linear function for speed. The study used vehicle trajectories that were reduced to pieces where the speed was decreasing or increasing. The points where the speed changed from decreasing to increasing or vice versa were then classified as action points. The findings indicate that accelerations are piecewise constant which contrasts with traditional stimulus response where the acceleration changes based upon the stimuli.

One study by Higgs et al. [10] also worked on adding new thresholds to the Wiedemann model. The new thresholds were: a hook following threshold and a pass threshold. Hook following is defined as when a driver is passed by another vehicle and then maintains a car-following period with that vehicle. The pass threshold is created when a driver is catching up to another vehicle and sometimes the driver decides to follow the lead vehicle and other times the driver decides to immediately pass the lead vehicle. The pass threshold also aids in discovering times when the driver made a decision to pass but was forced to follow due to other factors like being trapped by a vehicle to the side.

9.1.1.1.2 Linear and Non-Linear Models

Research into linear and non-linear car-following models is focused on modifying models to make them more robust, creation of new models, and obtaining optimal calibration parameters. The popular models in this category are the General Motors (GM) model [11], Intelligent Driver Model (IDM) [12], and the Velocity Difference model (VDIFF) [13].

The GM model has been used and modified extensively to the point that an entire family of car-following models has been formed based on it which includes various forms of the GM model [14-16], the Gazis-Herman-Rothery (GHR) model [17], the Pipes model [18], and the Forbes model [19]. Siuhi and Kaseko [20] demonstrated the need for separate models for acceleration and deceleration responses by developing a family of car-following models and addressing the shortcomings of the General Motors (GM) model. Previous work from Osaki [21] and Subranmanian [22] modified the GM model separating the acceleration and deceleration responses. Ahmed [23], following some work from Subranmanian assumed non linearity in the stimulus term and introduced traffic density. Results from Ahmed [23] and Toledo [24] showed, against popular belief, that acceleration increases with speed but decreases with vehicle separation. Due to statistical insignificance, Ahmed and Toledo also removed speed from their deceleration models. Siuhi and Kasvo [20] addressed some of these shortcomings by developing separate models, not only for acceleration and deceleration, but also for steady-state responses. Nonlinear regression with robust standard errors was used to estimate the model parameters and obtain the

distributions across drivers. The stimulus response thresholds that delimit the acceleration and deceleration responses were determined based on signal detection theory. Ossen and Hoogendoorn [25] studied the car-following behavior of individual drivers using vehicle trajectory data that were extracted from high-resolution digital images collected at a high frequency from a helicopter. The analysis was performed by estimating the parameters of different specifications of the Gazis-Herman-Rothery (GHR) car-following model car-following rule for individual drivers. According to the authors, the main contribution of their study was that considerable differences between the car-following behaviors of individual drivers were identified that can be expressed in terms of different optimal parameters and also as different car-following models that appear to be optimal based on the individual driver data.

Using two models of similar complexity (number of parameters): the “Intelligent Driver Model” (IDM) and the “Velocity Difference Model” (VDIFF), Kesting and Treiber [26] researched car-following behaviors on individual drivers using publicly available trajectory data for a straight one-lane road in Stuttgart, Germany. They used a nonlinear optimization procedure based on a genetic algorithm to minimize the deviations between the observed driving dynamics and the simulated trajectory. One of the major findings of the study was that a significant part of the deviations between measured and simulated trajectories can be attributed to the inter-driver variability and the intra-driver variability (stipulating that human drivers do not drive constantly over time, and their behavioral driving parameters change)—the later accounts for a large part of the deviations between simulations and empirical observations.

One study, by Kim and Mahmassani [27] analyzed the impacts of correlated parameters in linear and non-linear car-following models. The analysis included the Gipps [28], IDM, and Helly [29] models and used factor analysis to find the correlations between the model parameters. The findings show that ignoring the correlation between parameters produces unreliable and thus inaccurate results. Schultz and Rilett [30] studied the effects of different distributions on the sensitivity parameters of CORSIM [31]. The findings were that using different distributions than the default was very effective at replicating real data from Houston, Texas. Punzo, Ciuffo, and Montanino [32] conducted a study into various calibration techniques used for car-following models. The findings were that the settings used for the calibration method highly affect the results and that most methods have difficulty arriving at the global minimums. Genetic algorithms were able to find the optimal value of the most sensitive parameter consistently, but they were only able to find the global minimum once in 64 runs. The impact of this study is that the results of model calibration are unlikely to be the global minimum and that fact should be taken into account when presenting calibration results.

9.1.1.1.3 Combination Models

Models in this category are attempts to combine the strengths of two different models to create a new stronger more robust car-following model. Abbas et al. [33] presented a hybrid car-following model that combines the strengths of the Wiedemann model and the GHR model. The model uses the thresholds of the Wiedemann model to define the driving tasks such as decelerating to follow lead vehicle, accelerating to maintain distance to lead vehicle, decelerating to maintain distance to lead vehicle, etc. The model then uses specific calibrated GHR parameters for each task, due to the difference between the tasks, to calculate the acceleration of the following vehicle.

9.1.2 Effects of Emotion on Car-Following Behavior

The recent developments into the effects of emotion on car-following behavior fall into two categories: (1) driving simulator studies and (2) development of new car-following models. The driving simulator studies [34-36] focus on various aspects of driving that affect the driver emotionally. The new car-following model [37] that was developed uses a markovian process to recreate the differences that emotions create in the distribution of driving behaviors. The results of this effort also showed that angry and anxious drivers tend to have smaller headways than happy or neutral drivers.

9.2 Methodology

The methodology used in this effort was divided into three main parts: a psychophysiological driving simulator study, calibration of data to Wiedemann car-following model, and calibration of data to Wiedemann-GHR car-following model. First, a brief description of the driving simulator study is given. Then, the Wiedemann model is described in detail followed by the Wiedemann-GHR model. The car following data for ten different drivers from the psychophysiological driving simulator study were used to calibrate the models using a genetic algorithm.

9.2.1 Psychophysiological Driving Simulator Study

The psychophysiological driving simulator study was conducted in previous research [34]. This study used psychological surveys to collect data on the personality traits of drivers. The study presented drivers with three different scenarios (base, anxiety, and anger) designed to induce emotions in the drivers. The base scenario was designed as a baseline for comparison to the other two scenarios. The anxiety scenario was designed to induce anxiety in the driver by forcing them to break the speed limit and risk being chased by police vehicles in order to complete the scenario within the time limit. The anger scenario was designed to induce anger in the drivers by irritating them through the aggressive actions of the surrounding traffic. During these scenarios, data were collected using psychophysiological measures (EKG, respiration, and skin conductance).

9.2.2 Wiedemann Car-Following Model

In 1974, Wiedemann [1] introduced a car-following model that was based on the psychophysical behavior. The concept of thresholds in the Wiedemann model captured the driver's alertness in small space headway and the lack of explicit car-following behavior with large headways. In addition, it allowed the model to explain the oscillation phenomena observed in car-following behaviors.

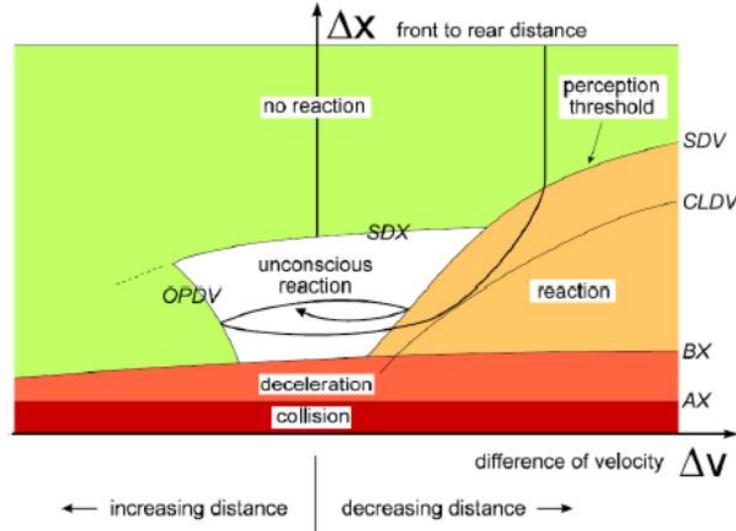


Figure 9-1: Wiedemann 74 Car Following Logic [38]

Figure 9-1 shows the graphical form of the Wiedemann 74 model. The different thresholds are shown with a certain shape that can only be amplified during the calibration procedure. The figure shows the subject vehicle approaching a lead vehicle (ΔX decreasing due to higher subject vehicle's speed shown by a positive ΔV), and entering a perception area (crossing the SDV threshold) where it has to reduce speed. The subject vehicle then crosses another threshold (CLDV) where it reacts and reduces speed even further to enter an unconscious reaction car-following episode. The subject vehicle then continues the unconscious car-following episode as long as it remains bounded by the OPDV, SDX, and SDV thresholds[1].

9.2.3 Wiedemann-GHR Model

The Wiedemann-GHR Model combined the Wiedemann model with the GHR model. The GHR model is shown in Equation 1. The GHR model relates the acceleration to the current speed, relative speed, and space headway. The weakness of the Wiedemann model is that it assumes that the behavior of the driver and the driving environment remain constant within each regime. The weakness of the GHR model is that it assumes that driver behavior is constant throughout a car-following period. The combination of the GHR model with the Wiedemann model used the strengths of both models where: (1) the Wiedemann model defined the driving regimes and (2) the GHR defined the behavior of the driver within each regime.

$$a_n(t) = cv_n^m(t) \frac{\Delta v(t-T)}{\Delta x^l(t-T)} \quad (3)$$

$a_n(t)$ is the acceleration of the subject vehicle at time t

$v_n(t)$ is the speed of the subject vehicle at time t

T is the perception reaction time of the driver

$\Delta v(t - T)$ is the relative speed at time t minus T

$\Delta x(t - T)$ is the space headway at time t minus T

c, l, m are calibration parameters

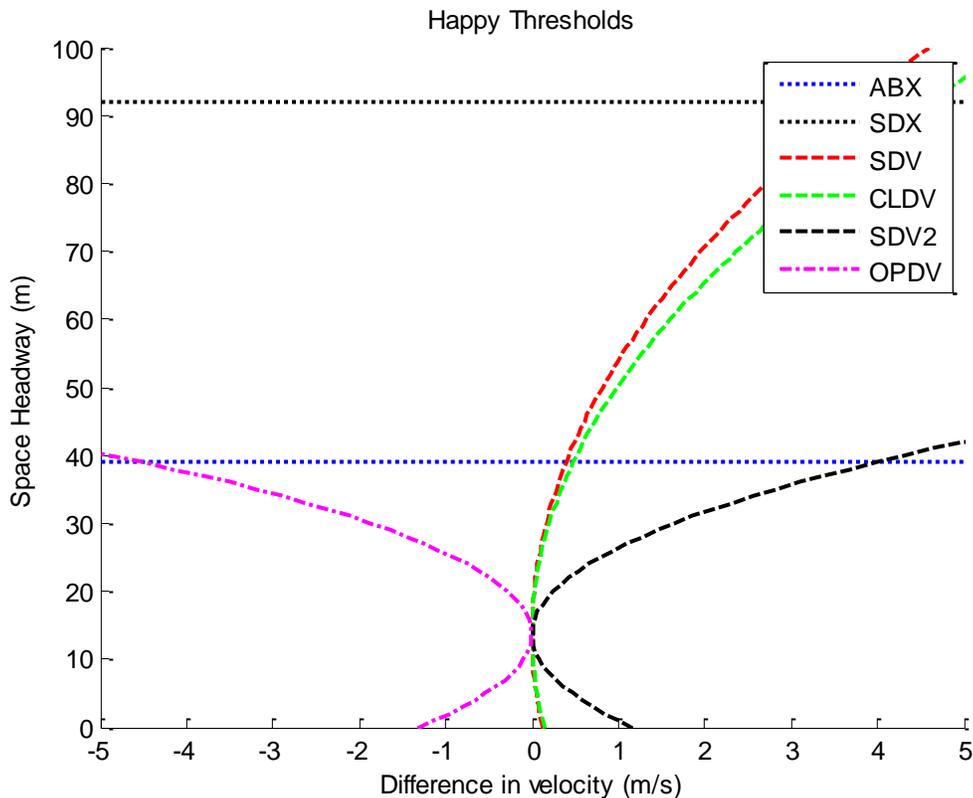
9.3 Results

Table 9-1 shows that the Wiedemann-GHR model outperformed the Wiedemann model in terms of root mean squared error for all of the emotional states included in this study. It is interesting to note that the Wiedemann model showed the lowest error for Angry.

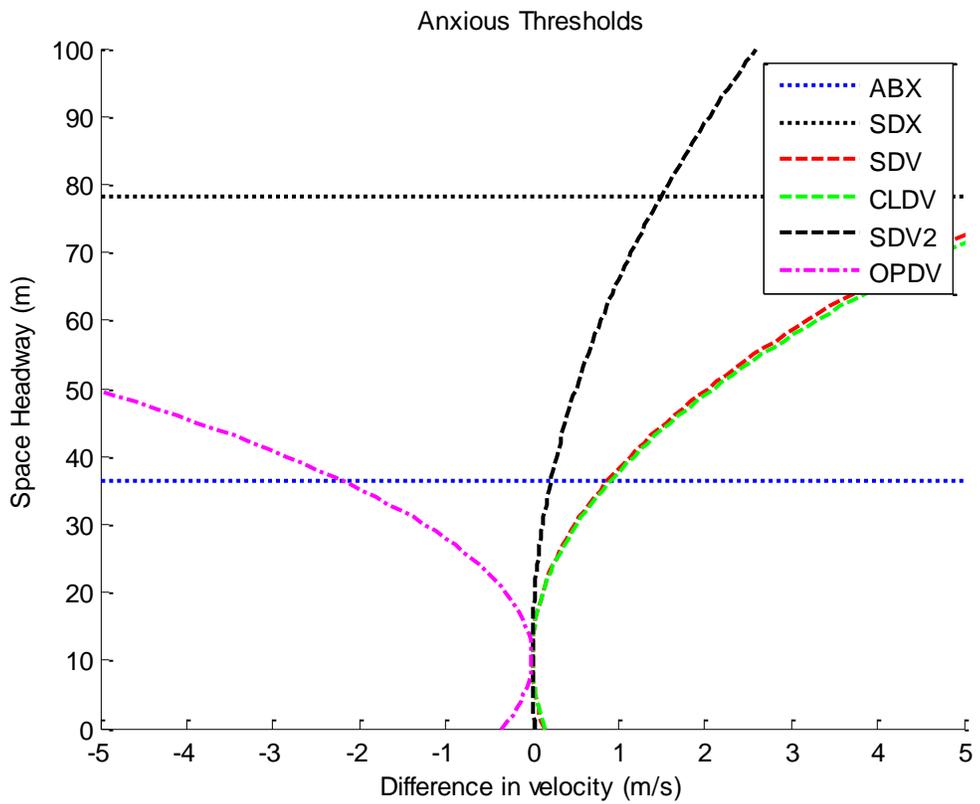
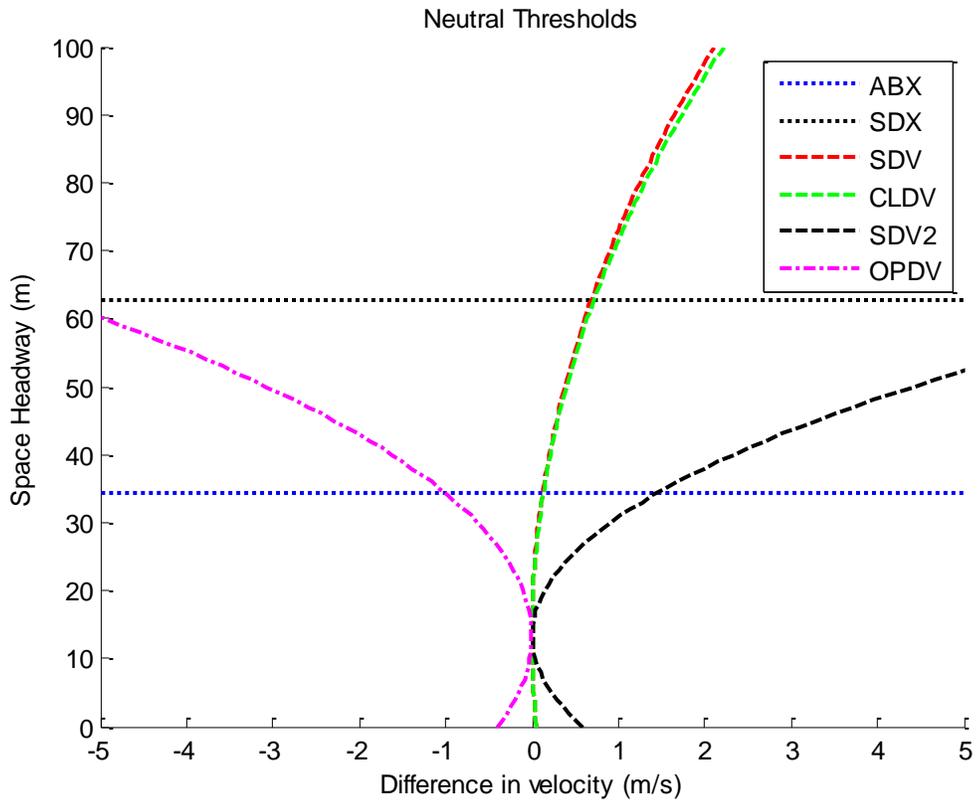
Table 9-1: Root Mean Squared Error of Wiedemann and Wiedemann-GHR Models for Different Emotions

	Happy	Neutral	Anxious	Angry
Wiedemann	39.98	59.73	11.42	3.28
W-GHR	2.16	2.10	1.39	1.52

Figure 9-2 shows that the SDX threshold (upper boundary of following) was highest for the happy emotional state and lowest for the angry emotional state. It is also interesting to note that the SDX threshold for the angry emotional state was lower than the ABX thresholds (lower boundary of following) for all the other emotional states. This indicates that driver exhibited extremely aggressive behavior when they were angry. The SDV2 threshold and the OPDV thresholds were boundaries to the following regime where the driver noticed that they were either falling behind or too close to the lead vehicle and decided to accelerate or decelerate. It is interesting to note that the SDV2 thresholds for the angry and anxious regimes were near vertical indicating that the driver tended to go slower than the lead vehicle. Combining this fact with the low thresholds of the angry emotional state indicated that the driver was forced to follow behind of a slow vehicle and then engaged in tailgating the lead vehicle.



(a)



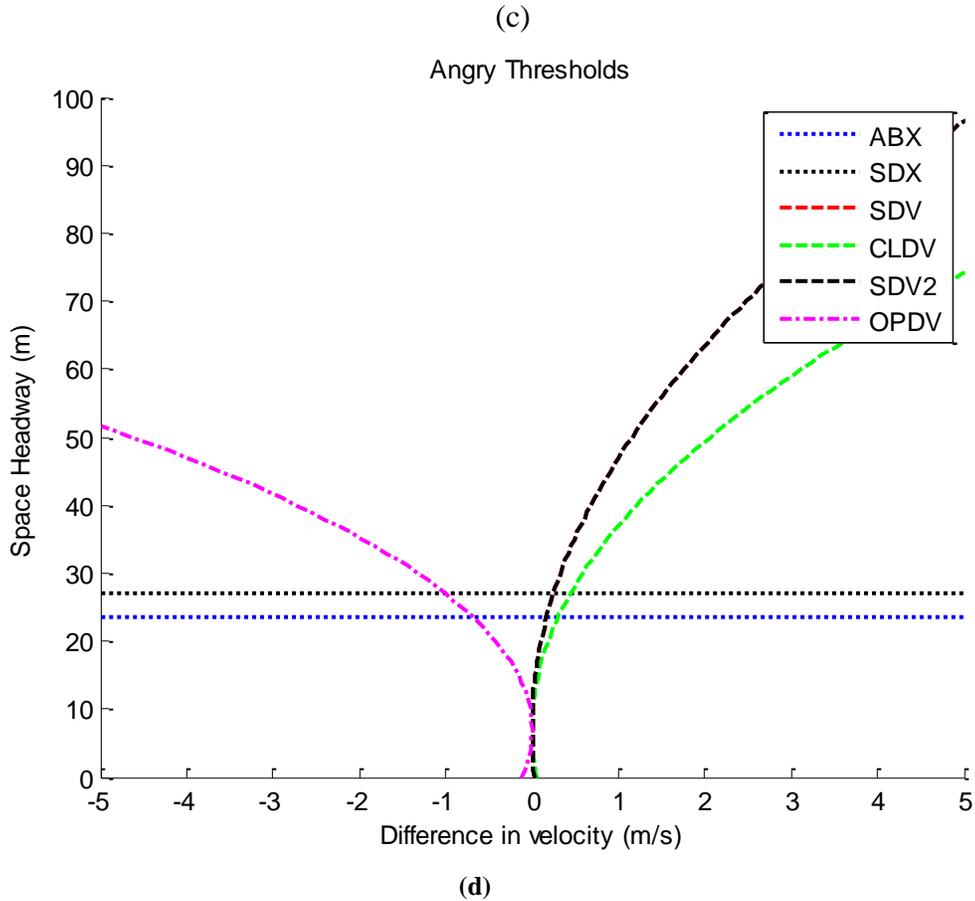


Figure 9-2: Wiedemann Thresholds of the W-GHR Model for Different Emotions

9.4 Conclusions

This research effort incorporated the effects of emotions into a psychophysical car-following model. The methodology used in this effort used data that were collected previously in a psychophysiological driving simulator study. The methodology included calibration of two psychophysical car-following models (Wiedemann and Wiedemann-GHR) to the data that were collected in the simulator study. The results revealed that the Wiedemann-GHR model outperformed the Wiedemann model in terms of root mean squared error. The results also showed that each emotional state has a unique set of optimal thresholds for the Wiedemann-GHR model. The use of these thresholds added the ability to adapt to the emotional state of the drivers to the Wiedemann-GHR model.

Future research should analyze the effects of personality traits on the threshold values. Future research should also examine other methods to induce emotions in drivers. Future research should also examine the effects of different driving conditions on the emotional state of the driver.

References

- [1] R. Wiedemann, "Simulation des Strassenverkehrsflusses," Instituts für Verkehrswesen der Universität Karlsruhe Universität Karlsruhe, 1974.
- [2] R. Wiedemann and U. Reiter, "Microscopic Traffic Simulation: The Simulation System MISSION," CEC Project ICARUS (V1052) final report 1992.

- [3] H.-T. Fritzsche, "A Model for Traffic Simulation," *Transportation Engineering Contribution*, vol. 5, pp. 317-321, May 1994.
- [4] M. Fellendorf, "VISSIM: A Microscopic Simulation Tool to Evaluate Actuated Signal Control Including Bus Priority," in *64th Institute of Transportation Engineers Annual Meeting*, Dallas, TX, 1994, pp. 1-9.
- [5] G. D. Cameron and G. I. Duncan, "PARAMICS—Parallel microscopic simulation of road traffic," *The Journal of Supercomputing*, vol. 10, pp. 25-53, 1996.
- [6] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann Car Following Model over Different Speeds using Naturalistic Data," presented at the Road Safety and Simulation, Indianapolis, Indiana, 2011.
- [7] M. Brackstone, "Driver Psychological Types and Car Following: Is There a Correlation? Results of a Pilot Study," in *Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2003, p. 6.
- [8] S. Menneni, C. P. D. Sun, P.E., and P. Vortisch, "An Integrated Microscopic and Macroscopic Calibration for Psycho-Physical Car Following Models " in *Transportation Research Board Annual Meeting*, 2009.
- [9] S. Hoogendoorn, R. Hoogendoorn, and W. Daamen, "Wiedemann Revisited," *Transportation Research Record*, vol. 2260, pp. 152-162, December 2011.
- [10] Bryan Higgs, Montasir Abbas, Alejandra Medina, and C. Y. D. Yang, "Reconstructing the Wiedemann Model using Naturalistic Driving Data," presented at the TRB Traffic Flow Theory Committee (TFTC) Summer Meeting and Greenshields Symposium, Annecy, France, 2010.
- [11] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic Dynamics: Studies in Car Following," *Operations Research*, vol. 6, pp. 165-184, March-April 1958.
- [12] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 62, pp. 1805-1824, 2000.
- [13] R. Jiang, Q. Wu, and Z. Zhu, "Full Velocity Difference Model for a Car-Following Theory," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, 2001.
- [14] D. C. Gazis, R. Herman, and R. B. Potts, "Car-Following Theory of Steady-State Traffic Flow," *Operations Research*, vol. 7, pp. 499-505, July 1 1959.
- [15] R. Herman, E. W. Montroll, R. B. Potts, and R. W. Rothery, "Traffic Dynamics: Analysis of Stability in Car Following," *Operations research*, vol. 7, pp. 86-106, 1959.
- [16] R. Herman and R. B. Potts, "Single Lane Traffic Theory and Experiment," in *International Symposium on the Theory of Traffic Flow*, Warren, MI, 1959, pp. 120-146.
- [17] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear Follow-The-Leader Models of Traffic Flow," *Operations Research*, vol. 9, pp. 545-567, 1961.
- [18] L. A. Pipes, "Operational analysis of traffic dynamics," *Journal of Applied Physics*, vol. 24, pp. 274-281, 1953.
- [19] T. W. Forbes, "Human Factor Consideration in Traffic Flow Theory," *Highway Research Board Record*, vol. 15, pp. 60-66, 1963.
- [20] S. Siuhi and M. Kaseko, "Parametric Study of Stimulus-Response Behavior for Car-Following Models," in *89th Transportation Research Board Annual Meeting*, Washington, D.C., 2010.
- [21] H. Osaki, "Reaction and anticipation in the Car-Following Behavior," in *12th International Symposium on the Theory of Traffic Flow and Transportation*, 1993.
- [22] H. Subranmanian, "Estimation of Car-Following Models," Master of Science in Transportation Master Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1996.
- [23] K. I. Ahmed, "Modeling Drivers' Acceleration and Lane Changing Behavior," Doctor of Science in Transportation Systems and Decision Sciences PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1999.
- [24] T. Toledo, "Integrating Driving Behavior," Doctor of Philosophy PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2003.
- [25] S. Ossen and S. P. Hoogendoorn, "Validity of trajectory-based calibration approach of car-following models in presence of measurement errors," *Transportation Research Record*, pp. 117-125, 2008.
- [26] A. Kesting and M. Treiber, "Calibrating Car-Following Models using Trajectory Data: Methodological Study," *Transportation Research Record*, vol. 2088, pp. 148-156, 2008.
- [27] J. Kim and H. Mahmassani, "Correlated Parameters in Driving Behavior Models," *Transportation Research Record*, vol. 2249, pp. 62-77, December 2011.
- [28] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, pp. 105-111, 1981.

- [29] W. Helly, "Simulation of Bottlenecks in Single-Lane Traffic Flow," presented at the International Symposium on Transportation and Traffic Theory, Warren, MI, 1959.
- [30] G. G. Schultz and L. Rilett, "Analysis of distribution and calibration of car-following sensitivity parameters in microscopic traffic simulation models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1876, pp. 41-51, 2004.
- [31] A. Halati, H. Lieu, and S. Walker, "CORSIM-corridor traffic simulation model," in *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities*, 1997, pp. 570-576.
- [32] V. Punzo, B. Ciuffo, and M. Montanino, "Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted?," *Transportation Research Record*, vol. 2315, pp. 11-24, December 2012.
- [33] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic Car Driver Behavior: Hybrid Wiedemann and Gazis-Herman-Rothery Model Calibration," presented at the Transportation Research Board 91st Annual Meeting, Washington D.C., 2012.
- [34] B. Higgs and M. Abbas, "Experimental Design for a Psychophysiological Driving Simulator Study," presented at the 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 2014.
- [35] C. Pecher, C. Lemerrier, and J.-M. Cellier, "Emotions drive attention: Effects on driver's behaviour," *Safety Science*, vol. 47, pp. 1254-1259, 2009.
- [36] J.-Y. Kim, M.-Y. Park, and C.-S. Park, "Psychophysiological responses reflecting driver's emotional reaction to interior noise during simulated driving," in *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium', July 29, 2000 - August 4, 2000*, San Diego, CA, United states, 2000, pp. 196-199.
- [37] B. Higgs and M. Abbas, "Development of an Emotional Car-Following Model," presented at the 17th International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 2014.
- [38] PTV-AG, "VISSIM 5.10 User Manual," 2008.

10. Summary of Findings

10.1 Findings

Chapter 2 of this dissertation focuses on the investigation of the heterogeneity of car-following behavior. The results revealed behavior that changed: (1) between drivers, (2) between car-following periods, and (3) within a car-following period. This indicates heterogeneity that extends further than was previously considered as existing car-following models were developed under the assumption that behavior is constant during a car-following period. The exception would be psychophysical models, but these models assume that behavior is constant between car-following periods. The implications to these findings are that existing car-following models require re-evaluation.

Chapter 3 of this dissertation focuses on using a segmentation and clustering technique to examine the patterns in behavior amongst both car and truck drivers. The results revealed that car drivers show a unique distribution of behavior for each driver and that there are some common behaviors shared by drivers. The results also revealed that truck drivers show similar distributions of behavior and this distribution has two main clusters. The differences in the distributions of behavior of car drivers indicates that there are some significant variables that are not being collected and analyzed.

Chapter 4 of this dissertation focuses on the identification and classification of state-action clusters. The results revealed that: (1) 60% of the state-action clusters can be identified using only state variables, (2) 30% of the state-action clusters can be identified using only action variables, and 100% of the state-action clusters can be identified using both state and action variables. Also, 20% of the state-action clusters require the use of both state and action variables to be identified. The implications to these findings are that there are overlaps where multiple actions for the same driving state and where there are multiple driving states for the same action. This further indicates that there are important variables missing from the analysis of driver behavior.

Chapter 5 of this dissertation focuses on a multi-resolution comparison of car-following models. The results show that there is little difference in performance between calibrating the GHR car-following model to state-action clusters and calibration of the Wiedemann-GHR model. The similar performance could be due to the fact that both models effectively divide car-following periods into regimes and calibrates the GHR model to each regime. The difference between the two models is that using state-action clusters allows for more versatility in the separation of driving regimes. The Wiedemann-GHR model has predetermined threshold formulas form the barriers between regimes. The results also revealed variability in the performance of the models between car-following periods.

Chapter 6 of this dissertation focuses on the design of a psychophysiological driving simulator study. The results show that drivers exhibit different distributions of behavior in different emotional states. The results also show that drivers can be influenced emotionally by interactions with the surrounding traffic. The implications to these findings are that car-following models need to be revised to include the ability to model drivers in different emotional states. Also, the findings implicate a slight contagious nature to aggressive driving as drivers can be angered by the actions of other drivers.

Chapter 7 of this dissertation focuses on the development of an emotional car-following model. The resulting model uses a Markovian process because it captures the effects that emotions made on the distribution of driving behaviors. The model's transition probabilities revealed the following: (1) as drivers become angry they use closer following distances that indicate tailgating

the lead vehicle, (2) as drivers become anxious or stressed they also use closer following distances but not to the degree of tailgating, and (3) drivers use larger following distances in happy or neutral emotional states.

Chapter 8 of this dissertation focuses on the effect of physiological changes on driver behavior. The results showed that changes in the physiology of the driver (increased heart rate, sweating, etc.) caused differences in the behavior of drivers that exhibited as a different distribution of behaviors.

Chapter 9 of this dissertation focuses on the implementation of new findings about the effects of emotion into a psychophysical car-following model. The results showed that the Wiedemann-GHR model outperformed the Wiedemann model and is a better choice for the implementation of the modeling of different emotional states. The results also showed that each emotional state has a unique set of model parameter values that effectively represent the behavior of drivers in each emotional state. The use of these thresholds improve the versatility of the Wiedemann-GHR model by adding the ability to adapt to the emotional state of the drivers.

10.2 Limitations of the Study

The main limitation of this research is the limited number of drivers. This limitation has two implications: (1) insights gained through this research only offer implications towards the whole population of drivers and (2) the generalizability of the results is limited to drivers similar to those that participated in the study. The second point is especially true when examining the effects of emotion on driver behavior because of the potential influence of personality differences.

Tests of statistical significance are not included in this study because the objective of this study was to better understand and model the similarities and differences in driver behavior. The differences could be large (occurring in multiple variables) or small (occurring in one variable). The findings of this study give great insights into the nature of driver behavior that can be used to better choose the correct statistical tests that should be applied in the study of driver behavior.

This study highlights the differences in driver behavior, but the practical implications of these differences are not discussed in detail. The practical implications would be: how different state-action clusters affect an overall trajectory, how car-following models could create new state-action clusters in their attempt to recreate a trajectory, etc.

10.3 Recommendations for Future Research

It is recommended that future research expand upon the findings of this dissertation by examining a larger set of drivers to find the characteristics of a population. These characteristics can include the following: (1) a global distribution of driver behavior clusters, (2) identification of groups or classes of drivers using cluster distributions, (3) identification of high-crash risk clusters of behavior, and (4) examination of the effects of more emotional states. To accomplish these tasks, it is recommended that future research use a larger number of drivers that includes homogenous groups of drivers.