Building Tomorrow's Society
Bâtir la Société de Demain

Fredericton, Canada
June 13 – June 16, 2018/ *Juin 13 – Juin 16, 2018*

USING NATURALISTIC DRIVING DATA TO QUANTIFY DRIVER FOLLOWING BEHAVIOR DURING BRAKING

Tawfeek, Mostafa H.[1,3], and El-Basyouny, Karim[2]
[1] Ph.D. Candidate, University of Alberta, Canada
[2] Associate Professor and City of Edmonton Urban Traffic Safety Research Chair, University of Alberta, Canada
[3] mostafa.h.tawfeek@ualberta.ca

Abstract: This study aims at quantifying and comparing drivers' following behavior during braking at intersections and midblock road segments (i.e., non-intersection) using Naturalistic Driving Data (NDS) data. The quantification of driver behavior was based on exploring the Probability Density Functions (PDF) of two parameters (i.e., minimum following distance and time-to-collision) which were extracted from NDS data. To classify drivers' following behavior at intersection and segments, all events were mapped onto the road network using ArcGIS. The intersection events were defined based on the driver's stopping sight distance to the stop line of an intersection. The results showed that there was a significant statistical difference between the PDFs of the parameters at intersection and segment events. Also, the results showed that drivers tend to keep less following distances and high deceleration rates (i.e., more aggressive) during intersection events when compared to segment events. Generally, the results of this study highlight the importance of considering the driver location before judging or classifying his/her behavior.

INTRODUCTION

Driver assistance or in-vehicle warning systems (e.g., Red Light Violation Warning, Forward Collision Warning (FCW), etc.) are expected to have a significant role in crash avoidance. The expected benefits from such systems will be mainly realized by enhancing the driver's awareness regarding their surrounding environment (Farah et al., 2012; Mohammed et al., 2016; Olia et al., 2015). These systems can enhance the awareness of drivers by alerting them of safety-critical events through the use of in-vehicle messaging (e.g., audible, visual, etc.). For instance, FCW provides a warning to the driver when a rear-end collision is expected with a leading vehicle in the same lane and the same direction of travel (Harding et al. 2014). In other words, such a warning is warranted when the actual following distance is equal or less than a minimum following distance which is predefined by the warning system.

One of the major challenges facing these systems is to accommodate a wide range of driving behaviors (e.g., from cautious through neutral to aggressive driving situations) and satisfy various drivers' needs (e.g., reliable warnings). For a cautious driving situation (i.e., when the driver usually keeps a larger following distance to a leading vehicle), an FCW might be considered as a late warning in this situation. On the other hand, for an aggressive driving situation (i.e., when the driver usually keeps a lesser following distance to a leading vehicle) an FCW, in this situation, might be considered as an early warning. The evaluation of warning systems showed that too early warnings, reduce users' acceptance and decrease the impact of the warning on driver behavior (van Driel et al. 2007, Staubach et al. 2014). Conversely, in the case of late warnings, drivers could start a maneuver before the warning is relayed, and hence, drivers will develop a mistrust due to the conflict between their expectations and the warning system performance (Abe and Richardson 2005). Therefore, it is essential to consider the situations/contexts when the warning will be relayed to the driver and drivers' needs.

Several studies investigated and classified drivers based on their behavior to different categories as a preliminary step of the development of an efficient and safe performing driving assistance systems (Wang et al. 2010, Jensen et al. 2011, Martinez et al. 2017). In these studies, driver behavior was analyzed regardless the location of the driver (i.e., at intersections or on midblock segments). However, it is expected that the driver behavior will vary based on several factors such as the roadway geometry (Ben-Bassat and Shinar 2011), driver mental workload (Messer 1980), in-vehicle advisory messages (Mohammed et al. 2016), etc. Therefore, this study aims at, first, investigating the driver following behavior during braking at different road network locations (i.e., at intersections and on midblock segments). This investigation is carried out by quantifying two behavioral measures (i.e., minimum following distance and time-to-collision) which are extracted from naturalistic driving data. These events are then mapped onto the road network using ArcGIS to differentiate between intersection and non-intersection events. Second, a comparison between these measures at intersections and on midblock segments is conducted. The probability distributions of the behavioral measures are generated for each location for three speed groups (i.e., low, medium, and high). To further explore the differences in driver behavior, the probability distributions are compared using statistical testing.

1 LITERATURE REVIEW

The knowledge of unassisted driving behavior is essential for the development of an efficient warning system (Berndt et al. 2007). Therefore, several studies investigated drivers' nominal behavior to develop and/or enhance warning systems. To achieve this goal, drivers' behavior characteristics in terms of TTC were analyzed in different classes such as age, gender groups, and speed ranges (Montgomery et al. 2014). The purpose of this study was to characterize the differences between age and genders groups during car-following while braking situations (Montgomery et al. 2014). In a different study, an approach for driving behavior classification was proposed using several measurable parameters such as time headway, the inverse of TTC, and the time between accelerator pedal release to brake activation. The drivers were classified according to their aggression, stability, conflict proneness, and skillfulness (Wang et al. 2010).

Feng et al. (2017) assessed the longitudinal vehicle jerk as a measure to classify aggressive drivers using Naturalistic Driving Study (NDS) data. Two jerk-based measures were proposed, namely, the frequent utilization of large positive or negative jerk. First, from the NDS data, the drivers recognized to be aggressive according to their excessive speeding, tailgating, or crash/near-crash frequency in the data. Based on that, three groups of aggressive drivers were generated, and the proposed jerk-based measures were validated using these three groups. The results showed that the proposed measures could successfully identify aggressive drivers in the three groups of aggressive drivers. However, the results indicated that the large negative jerk was more efficient when compared to the large positive jerk in detecting aggressive drivers (Feng et al. 2017).

Using speed and lateral and longitudinal acceleration, Gonzalez et al. (2014) developed a linear filter based on Gaussian Mixture Model to detect driving aggressiveness. This model was validated by monitoring the driving behavioral measures of 10 drivers on a customized test route. This test route, which was part of a private road network, included five different road features such as a roundabout, curves, uphill and downhill segments, and sharp bend. The drivers conducted several driving rounds on the route including only one aggressive round for each driver. It is worth to mention that the drivers were not given any instruction/definition of the aggressive driving (Gonzalez et al. 2014).

Zheng et al. (2017) defined the aggressive driving based on a questioner given to 15 drivers before and after local road tests (i.e., within University of Florida campus area). This study aimed to assess the relation between the driver type and driver behavior in a high vehicle-pedestrian interaction low-speed environment. Such an environment was within campus area with speed limit between 20 mph and 35 mph. The driving aggressiveness, obtained from the questioner, was compared to behavioral measures, namely, driving desired speed and yield to pedestrian behavior. The results revealed that the two measures could efficiently identify the driver aggressiveness (Zheng et al. 2017).

Based on smartphone accelerometer data (i.e., the acceleration in X, Y, and Z axes of the smartphone), Osafune et al. (2017) classified the drivers into safe and risky drivers. This classification was validated based on drivers' crash records in 20 years. The data was collected for 15 months covering almost all areas of Japan. Three thresholds were selected to identify risky drivers including acceleration, deceleration, and left acceleration (Osafune et al. 2017).

Sundbom et al. (2013) defined aggressive and normal driving based on drivers' steering behavior in terms of curve lateral acceleration, yaw rate, and lateral movement in the lane. Only two drivers participated in this study on a test track where each driver had a mixture of aggressive and normal driving. A model was developed to online classify the two driving styles (i.e., aggressive and normal). It is worth to mention that the test track had no traffic and consisted only of curved sections and straight sections. Also, the straight segments were removed when estimating the model performance because of drivers' lack of excitation in these segments which generated classification errors (Sundbom et al. 2013). In a simulation environment, Casucci et al. (2010) differentiate between the prudent and aggressive drivers based on braking behavior and lateral performance. It was found that aggressive drivers completed an overtaking task faster than prudent drivers. In addition, the cruising speed (i.e., the final speed after completing the overtaking maneuver) of aggressive drivers is higher than the prudent drivers (Casucci et al. 2010).

In summary, driver behavior classification is an approach to investigate and understand nominal driver behavior which is used to enhance the development of different warning systems. A wide range of measures were used for driver behavior classification such as drivers' speed (Casucci et al. 2010, Rodriguez Gonzalez et al. 2014, Zheng et al. 2017), TTC (Montgomery et al. 2014), jerk (Feng et al. 2017), and acceleration (Osafune et al., 2017; Gonzalez et al., 2014). In almost all the mentioned studies, these measures were used as an aggregate measure (e.g., averaged over a certain period of time or specific maneuver) regardless the location of the driver. For instance, Gonzalez et al. (2014) classified the driver behavior while driving on a part of a private road network which consisted of several road features (i.e., roundabout, curves, etc.). However, the behavioral measures were aggregated on the entire route without considering the change in the driver behavior from one road feature (e.g., roundabout) to another (e.g., curve). In addition, Osafune et al. (2017) and Feng et al. (2017) extracted behavioral measures from smartphone data and NDS (both are high-frequency data) without referring to the location of the events of the interest.

Therefore, this paper is investigating whether the driver will be relatively more aggressive in a specific location on the network when compared to other locations. In other words, the hypothesis in this paper is that there is a difference in driver following behavior when driving on midblock segments and around intersections. Moreover, this paper will characterize driver behavior during braking on these two locations. The results of this paper could be used as the core of a context-aware application. Such an application adopts automatically to the driver behavior variation due to the change in driver's surrounding environment (i.e., other vehicles, pedestrians, infrastructure, weather, etc.) over time.

2 METHODOLOGY

To characterize driver behavior during braking, the NDS data was used to identify car following events as well as extract various driver behavioral measures (e.g., following distance and TTC). The candidate car following events during braking were split into three groups based on the host vehicle speed and plotted on a map using ArcGIS. The detailed procedure is discussed in the below subsections.

2.1 Data Description

Big data has a wide range of definitions in the literature. For instance, big data is defined as the type of data which is hard to store, manage, process, and visualize (Cox and Ellsworth 1997, Manyika et al. 2011). In addition, big data could be defined as the methods and technologies which enable unveiling hidden values from large, complex, and diverse datasets (Hashem et al. 2015). Moreover, big data could be described by high volume (i.e., large storage), variety (i.e., data format), high velocity (i.e., change in data over time), veracity (i.e., data credibility and certainty), and value (i.e., data usefulness) (Elragal 2014). Big data has a significant role in the automotive and transportation industry transformation. For instance, big data is used to develop intelligent vehicles applications such as Self-Driving cars, Autonomous Vehicles, and Connected Vehicles (Luckow et al. 2015). In addition, big data, in the form of NDS, was used for understanding driver behavior to develop robust intelligent vehicles applications (Wang et al. 2017). NDS data, which is considered as big data, is a rich source of detailed real-world driving behavioral and performance data. NDS was carried out to investigate driver behavior and performance by collecting detailed data on volunteer drivers, vehicles, and the surrounding environment during normal every day driving.

The NDS, which is used in this paper, was collected as part of the Safety Pilot Model Deployment (SPMD). SPMD was a data collection effort under real-life conditions with about 3,000 vehicles equipped with V2V communication devices in Ann Arbor, Michigan, US (Henclewood and Rajiwade 2015). SPMD was part of the "Connected Vehicle (CV) Safety Pilot Program" research initiative which was aimed at evaluating the safety benefits of CV technologies. The dataset comes from the Data Acquisition System (DAS) which was developed by Virginia Tech Transportation Institute. This dataset is available in the ITS Public Data Hub which is operated by the US DOT (USDOT 2018). The dataset was collected from 64 host vehicles equipped with Integrated Safety Devices (ISD) for two months (October 2012 and April 2013). For each second, ISD unit recorded data including position and GPS details (e.g., latitude, longitude, elevation, speed and GPS-based data fidelity measures) and driver behavior data (e.g., braking status, speed, yaw rate, and acceleration). In addition, the dataset contains radar data collected from the ISD unit. The radar data provided information pertaining to the detected vehicle type (i.e. light vehicle, heavy vehicle, bike, pedestrian, or unclassified), status (i.e. moving and in path), the distance from the host vehicle, and the relative speed to the object in front of the host vehicle (Henclewood and Rajiwade 2015). The data was stored in two *.csv* files with a total file size of around 40.7 GB. The first file housed the information related to the vehicle operation and position of the host vehicle. The second file stored the radar data of the detected vehicles around the host vehicle.

2.2 Identifying Drivers' Following Events During Braking

One of the car following identification algorithms in naturalistic data was developed by Kusano et al. (2014) which aimed at identifying host vehicle braking events while following another vehicle. This algorithm was validated and successfully identified approximately 92% of the car following events when compared to manual inspection of video data in a100-car NDS. This algorithm was used to investigate various aspects of driving following behavior during braking. For instance, this algorithm was used to identify the situations during normal car-following situations with brake application to compare between the TTC and the Enhanced TTC as thresholds for triggering FCW (Chen et al. 2016). In addition, the same algorithm was utilized to compare the TTC for different speed bins or different drivers' demographic groups (i.e., age and gender) in car-following scenarios while braking (Montgomery et al. 2014, Kusano et al. 2015).

Consequently, this algorithm will be adapted to serve the purpose of this study. The following steps, which were coded in MATLAB, summarize the NDS data processing and events identification:

1. Preparing and filtering the data:
    1. Remove all object types except the object type which was recognized as a light vehicle (i.e., passenger car).
    2. Remove all instances which were marked with invalid GPS data.
2. Identifying car following events during braking using Kusano et al. algorithm (Kusano et al. 2014). For more details about the algorithm, readers are directed to (Kusano et al. 2014, Tawfeek and El-Basyouny 2018)This algorithm could be summarized in the following steps:

1. Identify all braking events for the host vehicle only when the vehicle's speed is above 10 mph (4.47 m/s) for at least 1 second of the braking time.

2. Identify candidate leading vehicles within 3 seconds headway in front of the host vehicle during the braking events. These candidate leading vehicles should be within 2 meters laterally from the path of the host vehicle.

3. Remove fixed objects (i.e., parked vehicle and slow vehicles in adjacent lanes). This step depends on the speed of the leading vehicle, the gap distance between the leading and the following vehicle, and the azimuth change rate. The azimuth here is defined as the angle between the host vehicle path and the candidate leading vehicle.

4. Select leading vehicle, when multiple candidate vehicles were selected from the previous step, based on the lateral distance between the host and the leading vehicle and proximity to the host vehicle.

3. Selecting incidents of interest:

   1. Scan each identified following event for the minimum following distance.

   2. Extract and calculate driver behavior measures (i.e., TTC) occurring at the minimum following distance while braking.

A total of 43,738 car-following events during braking were identified as candidate events. The incidents, when the following distance is the minimum during the event, were prepared to be plotted on a map to differentiate between intersections and segments events as discussed in the next section.

2.3 Mapping and Defining Intersections and Segments Following Events

ArcGIS was used as a tool to map the identified car following events during braking and to classify each event into either an intersection or segment related event. The candidate events were imported to ArcGIS using the available position data of the host vehicle (i.e., latitude and longitude). Before starting the classification process, intersection and segment (non-intersection) events should be defined. The current literature has various definitions of the intersections influence area. For instance, both American Association of State Highway and Transportation Officials (AASHTO) and Highway Safety Manual (HSM) defines at-grade intersections by physical and functional areas (AASHTO 2010, 2011). The physical area is the area bounded by the stop lines of both intersecting roads. On the other hand, the functional area extends upstream the intersections on the intersecting roads. This area includes driver's decision and maneuver distances and queue storage distance. More specifically, intersection-related crashes are defined as the crashes which occur within 15 to 152 meters from the intersection center point (R.Stollof 2008, FHWA 2009). On the other hand, the 152 meters intersection influence area was considered to be suitable for intersections approaches according to the 40 mph speed limit only (Cottrell and Mu 2005). It is worth to mention that intersection crashes and intersection-related crashes have been defined as the crashes which occurred upstream or inside the physical area of the intersection (Wang et al. 2009). Therefore, similar to the definition of intersection crashes, the candidate events were identified as intersection events when they occurred either upstream or inside the physical area of the intersection.

Since speed is a main contributor to a driver's decision (i.e., perception-reaction) distance and maneuver (i.e., braking) distance, the speed limit should be considered when deciding the influence area (Fitzpatrick et al. 2000). Consequently, the candidate events were split into three categories based on the host vehicle speed, namely; low, medium, and high speeds. The boundaries of low, medium and high speeds were defined as less than 25 mph (40 km/h), between 25 mph and 55 mph (88 km/h), and more than 55 mph, respectively. These boundaries are chosen based on the speed limits in the City of Ann Arbor, Michigan (City of Ann Arbor 2013). Based on these categories, intersections influence area for each category was set and defined as shown in Figure 1.

Figure 1: Intersection influence area (shaded in yellow) and an example of the candidate intersection events (red dots) and segment events (blue dots).

The intersection influence distance (i.e., from the stop line and upstream the intersection) is calculated based on the driver's Stopping Sight Distance (SSD) which include driver's decision and maneuver distances as shown in the following equation.

[1] $SSD\ (meter) = 0.278 \times v \times PRT + \frac{v^2}{254\left(\frac{a}{g} \pm G\right)}$

where v is vehicle speed in km/h, PRT is driver's Perception-Reaction Time in seconds, a is vehicle deceleration in m^2/s, and G is roadway longitudinal slope.

The SSD was estimated for low, medium, and high-speed categories as 30, 80, and 150 meters respectively based on AASHTO recommendations for PRT of 2.5 seconds and deceleration of 3.41 m2/s (AASHTO 2011). It is worth to mention that this value of PRT is also recommended by Alberta's Highway Design Guide (Alberta Transportation 1999). Therefore, the intersection influence distance which starts from the intersection's stop line and upstream the intersection will be 30, 80, and 150 meters for low, medium, and high-speed categories. Based on this distance, the intersection influence area (shaded in yellow in Figure 1) was determined, and any candidate event within this area was classified as an intersection event or, otherwise, considered as a segment event as shown in Figure 1. All 44,383 candidate events were classified based on the position of the host vehicle as mentioned before.

3 RESULTS AND DISCUSSION

The identified events were mapped using ArcGIS to differentiate between intersection and segment events in three-speed groups as shown in Table 1. It is worth to mention that the high-speed group does not include any intersection events, as shown in Table 1.

Table 1: Events breakdown by event type and speed group

Speed Group	Intersection Events	Segment Events	Total Events
Low	11403	14962	26365
Medium	1995	12614	14609
High	0	3409	3409
Total	13398	30985	44383

The results were exported from ArcGIS to be processed in MATLAB. To characterize the driver behavior for each event type, the probability density functions (PDF) for minimum following distance and TTC were developed in each speed group for both intersection and segment events. The actual values of the measures were fitted to 17 continuous distribution types, namely; Beta, Birnbaum-Saunders, Exponential, Extreme value, Gamma, Generalized Extreme Value, Generalized Pareto, Inverse Gaussian, Logistic, Log-logistic, Log-normal, Nakagami, Normal, Rayleigh, Rician, *t* Location-Scale, and Weibull. Table 2 shows the distribution's name and parameters which best fits the actual values of the behavioral measures (i.e., minimum range and TTC) for each event type and speed group. The distribution was considered to best fit the actual values of the behavioral measures since the Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) had the lowest values.

Table 2: The parameters of the fitted distributions to the actual PDFs

	Speed Group	Location (mean value)	Distribution Name	Parameters			*P-value*
				Location	Scale	Shape	
Minimum Range (m)	Low	Int.(9.01)	Generalized Extreme Value	7.127	3.153	0.020	<0.001
	Low	Seg.(11.05)	Generalized Extreme Value	8.496	4.071	0.047	<0.001
	Med.	Int.(19.5)	Gamma	-	4.430	4.403	<0.001
	Med.	Seg.(20.99)	Gamma	-	4.984	4.212	<0.001
	High	Seg.(29.07)	Birnbaum-Saunders	-	25.100	0.563	-
TTC (sec)	Low	Int.(10.48)	Generalized Extreme Value	3.556	3.081	0.739	<0.001
	Low	Seg.(17.23)	Birnbaum-Saunders	-	9.383	1.295	<0.001
	Med.	Int.(21.42)	Log-normal	2.549	1.087	-	<0.001
	Med.	Seg.(30.00)	Gamma	-	20.834	1.440	<0.001
	High	Seg.(31.81)	Birnbaum-Saunders	-	24.162	0.794	-

Since the distributions of all the measures were not normally distributed, a nonparametric statistical test (i.e., Kolmogorov-Smirnov (KS) test) was used to check the significance of the difference between the intersection/segment event distributions in each speed group. The null hypothesis is that the values for each measure of the intersection and segment events are from the same continues distribution. As shown in Table 2, the results of KS test showed that all the measures of intersection events were different from segment events at 95% confidence level. These results indicate that driver behavior at intersections is substantially different from driver behavior on segments. Such results are crucial when developing any driving assistance system because drivers will receive the warning differently in varied contexts. In other words, drivers could accept a warning relayed to them on midblock segments while considering the same warning at intersections as an early warning which will have an impact on the entire system acceptance.

Table 2 also shows the mean values of the selected behavioral measures in the third column between brackets. As shown in the table, measures' mean values for intersection events were less than segment events. For instance, the mean value of the minimum range of low-speed intersection events (i.e., 9.01 m) was less than the minimum range of low-speed segment events (i.e., 11.05 m). This indicated that the drivers tend to have larger following distance during braking when driving on segments than when driving on intersections. This observation was confirmed by the values of TTC where TTC of car-following during braking events at intersections were less than segment events. Moreover, the mean values of the behavioral measures in the medium-speed group were larger than the values of the low-speed group. In summary, drivers' following behavior during braking at intersections is statistically different from their behavior on midblock segments. In addition, drivers are more likely to follow closer to leading vehicles and to have lower TTC at intersections. Therefore, the results shown in Table 1 indicate that the drivers are more aggressive when approaching intersections. It is worth to mention that Morton et al., (2005) defined aggressive driving for teenage drivers based on speed and headway (i.e., high speed and smaller headway indicate an aggressive driving). In addition, drivers were identified according to their TTC and headway to prudent and aggressive drivers (Wang et al. 2010). Hence, in this paper, aggressive driver behavior is defined based on the minimum following distance and TTC. However, this aggressive behavior is defined here relative to driving behavior on midblock segments which means that the aggressive behavior in this paper is not necessary linked to risky driving.

4 SUMMARY AND CONCLUSIONS

This paper quantified four driver behavioral measures (i.e., minimum following distance and TTC) for car-following during braking situations. Using NDS, this quantification is carried out for events occurred at intersections and midblock segments. In addition, the selected behavioral measures for intersection and segment events were compared. To achieve the objectives of this paper, the car-following events during braking were extracted from the NDS. These events were mapped onto the road network using ArcGIS to identify intersection and segment events. The intersection influence area was defined for three speed groups (i.e., low, medium, and high) based on driver's SSD and speed. Then, the probability distributions for the behavioral measures for each speed group for intersections and segments were developed and fitted to continuous probability distributions. In addition, intersection/segment driver following behavior during braking was compared by checking the significance of the difference between the behavioral measures distributions in each speed group.

The results of this study revealed that there was a considerable difference between driver behavior during braking at intersections and on midblock segments. This difference was defined as the differences between the behavioral measures distributions of intersection and segment events. Using KS test, these differences were statistically significant at 95% confidence level. In addition, this study documented the distributions' names and parameters which best fitted the behavioral measures. By investigating the mean values of the behavioral measures, it was found that drivers approaching intersections had smaller minimum following distance and smaller TTC than when driving on segments. This finding implies that the drivers are more likely to exhibit relatively aggressive behavior when approaching intersections. This finding highlights the importance of considering the driver location when classifying the drivers based on their aggressiveness. According to the current literature, the classification procedures are usually based only on some selected behavioral measures (e.g., acceleration, TTC, jerk, etc.) regardless the location. Considering the findings of this study, the proposed procedures in the literature could misclassify a driver as an aggressive driver when most of the data were collected in high intersection density areas (e.g., urban areas or downtown areas). Therefore, as the main contribution of this study, the driving data should be split based on the location before classifying drivers' behavior. In addition, the driving assistance systems should consider the surrounding environment (i.e., proximity to intersections) by modifying these systems to context-aware driving assistance systems. Such context-aware systems are expected to be more robust and reliable since it will accommodate the changing drivers' surrounding environment.

One of the primary goals of understanding driver behavior is to develop innovative traffic safety solutions. Hence, the results of this study could be beneficial for the development of collision avoidance/warning systems such as FCW. Forasmuch as driver behavior at intersections is different from the behavior on segments, it is recommended to develop an FCW algorithm which accounts for the driver location in the network. To develop such FCW algorithm(s), further analysis of the results will be conducted to select events of interest based on the distribution of acceleration, jerk, and TTC.

In addition, the results could enhance the calibration process of microsimulation models by updating the distributions of drivers' behavioral measures in simulation packages. Also, the PDF distributions of the behavioral measures could be used in modeling driver following behavior mathematically as suggested by Gonzalez et al. (2014). The PDF distributions generated from this study is expected to lead to more robust simulation results because of the consideration of the driver location and the fact that these distributions were generated based on NDS. The NDS grants the recording of normal driving behavior without exposing the drivers to a test setting as in test track which will affect their behavior.

Moreover, this study shed light on the importance of integrating the NDS with the road network map to deeply investigate the driver behavior. The pairing of the NDS and the road network, using ArcGIS, LiDAR data, or other means, is an up-and-coming area of research which will allow insightful findings in driver behavior understanding. Therefore, as an extension of this study, an automated approach to differentiate between intersection and midblock segment events could be explored. This automated approach could enhance the definition of the intersection influence area which has a wide range of definitions in the literature (e.g., the suggested definition in this study which was based on the SSD and vehicle speed).

References

AASHTO. 2010. Highway safety manual, first edition, with 2014 supplement.

AASHTO. 2011. A Policy on Geometric Design of Highways and Streets. : 942.

Abe, G., and Richardson, J. 2005. The influence of alarm timing on braking response and driver trust in low speed driving. Safety Science, **43**(9): 639–654.

Alberta Transportation. 1999. Highway geometric design guide.

Ben-Bassat, T., and Shinar, D. 2011. Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior. Accident Analysis and Prevention, **43**(6): 2142–2152.

Berndt, H., Wender, S., and Dietmayer, K. 2007. Driver braking behavior during intersection approaches and implications for warning strategies for driver assistant systems. Proc. IEEE Intelligent Vehicles Symposium,: 245–251.

Casucci, M., Marchitto, M., and Cacciabue, P.C. 2010. A numerical tool for reproducing driver behaviour: Experiments and predictive simulations. Applied Ergonomics, **41**(2): 198–210.

Chen, R., Sherony, R., and Gabler, H.C. 2016. Comparison of Time to Collision and Enhanced Time to Collision at Brake Application during Normal Driving. SAE Technical Paper, (2016-01–1448).

City of Ann Arbor. 2013. City of Ann Arbor Speed Limits Map. Available from http://www.a2gov.org/departments/engineering/Pages/default.aspx [accessed 12 April 2017].

Cottrell, W.D., and Mu, S. 2005. Utah Intersection Safety: Recurrent Crash Sites: Identification, Issues and Factors. University of Utah, Salt Lake City.

Cox, M., and Ellsworth, D. 1997. Managing Big Data for Scientific Visualization. ACM Siggraph, **97**: 1–17. Available from http://www.dcs.ed.ac.uk/teaching/cs4/www/visualisation/lectures/98-99/lect16ref.ps.gz.

van Driel, C.J.G., Hoedemaeker, M., and van Arem, B. 2007. Impacts of a Congestion Assistant on driving behaviour and acceptance using a driving simulator. Transportation Research Part F: Traffic Psychology and Behaviour, **10**(2): 139–152.

Elragal, A. 2014. ERP and Big Data: The Inept Couple. Procedia Technology, **16**: 242–249.

Farah, H., Koutsopoulos, H.N., Saifuzzaman, M., Kölbl, R., Fuchs, S., and Bankosegger, D. 2012. Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior. Transportation Research Part C: Emerging Technologies, **21**(1): 42–56.

Feng, F., Bao, S., Sayer, J.R., Flannagan, C., Manser, M., and Wunderlich, R. 2017. Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. Accident Analysis and Prevention, **104**: 125–136.

FHWA, U.S.D. of T. 2009. Manual on Uniform Traffic Control Devices.

Fitzpatrick, K., Carlson, P.J., Wooldrigge, M.D., and Brewer, M.A. 2000. Design factors that affect driver speed on suburban arterials. : 164.

Harding, J., Powell, G., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J., and Wang, J. 2014. Vehicle-to-Vehicle Communications : Readiness of V2V Technology for Application. Washington, DC.

Hashem, I.A.T., Yaqoob, I., Anuar, N.B., Mokhtar, S., Gani, A., and Ullah Khan, S. 2015. The rise of “big data” on cloud computing: {Review} and open research issues. Inf Syst, **47**(Journal Article): 98–115.

Henclewood, D., and Rajiwade, S. 2015. Safety Pilot Model Deployment – Sample Data Environment Data Handbook. **1.3**.

Jensen, M., Wagner, J., and Alexander, K. 2011. Analysis of in-vehicle driver behaviour data for improved safety. International Journal of Vehicle Safety, **5**(3): 197.

Kusano, K.D., Chen, R., Montgomery, J., and Gabler, H.C. 2015. Population distributions of time to collision at brake application during car following from naturalistic driving data. Journal of Safety Research, **54**: 95–104.

Kusano, K.D., Montgomery, J., and Gabler, H.C. 2014. Methodology for identifying car following events from naturalistic data. IEEE Intelligent Vehicles Symposium, Proceedings,: 281–285.

Luckow, A., Kennedy, K., Manhardt, F., Djerekarov, E., Vorster, B., and Apon, A. 2015. Automotive big data: Applications, workloads and infrastructures. Proceedings - 2015 IEEE International Conference

on Big Data, IEEE Big Data 2015,: 1201–1210.
Manyika, J., Chui, M., Brown, B., Bugin, J., Dobbs, R., Roxburgh, C., and Byers, A.H. 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, (June): 156.
Martinez, C.M., Heucke, M., Wang, F.Y., Gao, B., and Cao, D. 2017. Driving Style Recognition for Intelligent Vehicle Control and Advanced Driver Assistance: A Survey. IEEE Transactions on Intelligent Transportation Systems, **PP**(99): 1–11.
Messer, C.J. 1980. Methodology for Evaluating Geometric Design Consistency. Transportation Research Record: Journal of the Transportation Research Board, **757**: 7–14.
Mohammed (Mostafa H. Tawfeek), M., Ke, Y., Gao, J., Zhang, H., El-Basyouny, K., and Qiu, T.Z. 2016. Connected Vehicle V2I Communication Application To Enhance Driver Awareness At Signalized Intersections. *In* 11th International Transportation Specialty Conference, CSCE 2016.
Montgomery, J., Kusano, K.D., and Gabler, H.C. 2014. Age and gender differences in time to collision at braking from the 100-car naturalistic driving study. Traffic injury prevention, **15 Suppl 1**(November): S15-20.
Olia, A., Abdelgawad, H., Abdulhai, B., and Razavi, S.N. 2015. Assessing the Potential Impacts of Connected Vehicles: Mobility, Environmental, and Safety Perspectives. Journal of Intelligent Transportation Systems, **2450**(September): 1–15.
Osafune, T., Takahashi, T., Kiyama, N., Sobue, T., Yamaguchi, H., and Higashino, T. 2017. Analysis of Accident Risks from Driving Behaviors. International Journal of Intelligent Transportation Systems Research, **15**(3): 192–202.
R.Stollof, E. 2008. Intersection and Junction Fatalities in the Context of Access Management. 8th National Conference on Access Management, (July).
Rodriguez Gonzalez, A.B., Wilby, M.R., Vinagre Diaz, J.J., and Sanchez Avila, C. 2014. Modeling and detecting aggressiveness from driving signals. IEEE Transactions on Intelligent Transportation Systems, **15**(4): 1419–1428.
Simons-Morton, B., Lerner, N., and Singer, J. 2005. The observed effects of teenage passengers on the risky driving behavior of teenage drivers. Accident Analysis and Prevention, **37**(6): 973–982.
Staubach, M., Schebitz, N., Köster, F., and Kuck, D. 2014. Evaluation of an eco-driving support system. Transportation Research Part F: Traffic Psychology and Behaviour, **27**(PA): 11–21.
Sundbom, M., Falcone, P., and Sjoberg, J. 2013. Online driver behavior classification using probabilistic ARX models. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC,: 1107–1112.
Tawfeek, M.H., and El-Basyouny, K. Forthcoming 2018. A Perceptual Forward Collision Warning Model using Naturalistic Driving Data. Canadian Journal of Civil Engineering,.
USDOT. 2018. ITS Public Data Hub. Available from https://www.its.dot.gov/data/ [accessed 3 January 2018].
Wang, J., Lu, M., and Li, K. 2010. Characterization of Longitudinal Driving Behavior by Measurable Parameters. Transportation Research Record: Journal of the Transportation Research Board, **2185**: 15–23.
Wang, W., Liu, C., and Zhao, D. 2017. How Much Data is Enough? A Statistical Approach with Case Study on Longitudinal Driving Behavior.
Wang, X., Abdel-Aty, M., Nevarez, A., and Santos, J.B. 2009. Investigation of Safety Influence Area for Four-Legged Signalized Intersections: Nationwide Survey and Empirical Inquiry. Transportation Research Record: Journal of the Transportation Research Board, **2083**(1): 86–95.
Zheng, Y., Chase, R.T., Elefteriadou, L., Sisiopiku, V., and Schroeder, B. 2017. Driver types and their behaviors within a high level of pedestrian activity environment. Transportation Letters, **9**(1): 1–11.