Driver Distraction Impact on Road Safety: A Data-driven Simulation Approach

Amany A. Kandeel
School of Computing
Queen's University
Kingston, Canada
a.kandeel@queensu.ca

Ahmed A. Elbery
School of Computing
Queen's University
Kingston, Canada
a.elbery@queensu.ca

Hazem M. Abbas
Dept. Computer & Systems Eng
Ain Shams University
Cairo, Egypt
hazem.abbas@eng.asu.edu.eg

Hossam S. Hassanein
School of Computing
Queen's University
Kingston, Canada
hossam@cs.queensu.ca

GLOBECOM 2021 - 2021 IEEE Global Communications Conference | 978-1-7281-8104-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GLOBECOM46510.2021.9685932

***Abstract*—Driver distraction identification is crucial to improve road safety. Through vehicular communications, vehicles can exchange driver behavior information, based on which distracted drivers can be identified, and all drivers can be notified, which can mitigate the impact of driver distraction on road safety. To build such systems, it is essential to understand the different types of distraction, and how they affect driver behavior and their relationship to crashes or near-crashes. This understating should be based on real datasets, which are very limited. Therefore, in this paper, we cover this gap by building a data-driven simulation model to quantify the impact of realistic driver distraction on traffic safety. In particular, we use the 2nd Strategic Highway Research Program Naturalistic Driving Study (SHRP2 NDS) dataset to develop a simulation framework for driver distraction. First, we pre-process and analyze the dataset for different types of distractions. The analysis shows that the data can not be fitted to any of the known distributions. Therefore, we use the Gaussian Mixture Model (GMM) to represent the distraction intervals for the different distraction types. We then use these GMM models and the statistics collected from the data to realistically simulate the driver distraction using the Simulation for Urban MObility (SUMO) software. Finally, we use this framework to simulate the driver distraction in a real network. The data analysis and simulation results revealed important and interesting conclusions, such as decreasing the crash ratio when roads become congested.**

***Index Terms*—Driver Distraction, Traffic Safety, Gaussian Mixture Model, Simulation**

I. Introduction

Road accidents participate significantly in increasing death rates. The World Health Organization (WHO), in its report on road safety [1], declared that the number of annual road traffic deaths reached 1.35 million cases. This large number of accidents and deaths can be reasoned to many factors including road conditions, vehicle conditions, or weather. However, the main reason for this enormous number of crashes is human errors, which cause more than 80% of the road accidents [2]. Although connected and automated vehicles will be mainstream and an important part of transport systems in the future, traditional vehicles are also expected to participate strongly in this future. Therefore, it is essential to study the driver behavior and its effects on the safety and efficiency of these systems [3].

Driving is a multitasking process that involves perception, judgment, decision making, and operation. Such a process needs a high degree of concentration from drivers to achieve safety [4]. The lack of concentration may be due to fatigue, distractions, or emotions [5]. These factors affect the driver's behavior and his ability to precept, judge, and react to on-road events such as deceleration of the leading cars or maneuvers by other surrounding vehicles. This can lead to fatal consequences. Driver distraction is a significant cause of such fatal human errors. For instance, in the United States, the National Highway Traffic Safety Administration (NHTSA) reported that approximately 20,000 people lost their lives from 2014 to 2019 in crashes involving distracted drivers.

Driver distraction may be defined in different ways. One of these definitions is " The diversion of attention away from activities critical for safe driving toward a competing activity, which may result in insufficient or no attention to activities critical for safe driving" [6]. Distracted driving not only affects the driver's performance in the subject vehicle, it also affects the surrounding vehicles. So, vehicle-to-vehicle communication can be utilized to warn other drivers when a distracted driver is detected in a certain vehicle. As shown in Fig. 1, when the follower vehicle receives the distracted driver notification from its leader vehicle, the follower driver will be more cautious and increases his/her headway distance as a reaction to any potential driving misbehavior.

Despite the serious impact of distracted driving, its recording is elusive due to several reasons. Firstly, most police reports do not have videos recording of events and depend on asking the driver about the event. Secondly, even if the event is recorded, it is hard to prove the state of the driver especially if there is no physical evidence such as cases of cognitive distraction or looking at an object on the road. Therefore, reporting the distraction is very difficult, and obtaining the naturalistic dataset that contains distraction details such as its types and duration is not an easy mission. Consequently, understating these distractions and how they affect safety in different

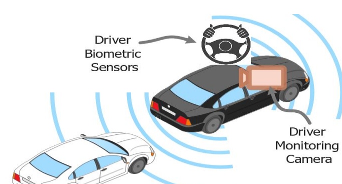


Fig. 1: Vehicular communication and driver distraction

conditions is a challenging task. This is an important barrier in the way of building systems that utilize advanced technologies like vehicular wireless communications and machine learning to solve driver distraction problems or mitigating their impacts. To this end, in this paper, we introduce a novel data-driven simulation framework for driver distraction that can improve the understanding of driver distraction and help overcome its impacts, and consequently, achieve safe driving. This framework combines the advantage of studying naturalistic data and traffic simulation. The significant merit of using naturalistic data is the rational statistics and the accurate results. On the other hand, utilizing the traffic simulation allows studying relations and effects without exposing designated drivers to danger.

The main contributions of this paper are:

- Providing an in-depth study of different types of distractions, their intervals, and distributions based on the SHRP2 dataset.
- Using Gaussian Mixture Model (GMM) to develop models for the different distraction types in the SHRP2 dataset. Each type is represented by a different number of components to best fit its distribution.
- Developing a technique to simulate driver distraction based on the statistics extracted from the dataset and using the developed GMM models. In this technique, distraction tasks are selected stochastically based on their probabilities in the dataset. The distraction intervals are generated based on the GMM model for the selected task.
- Performing sensitivity analysis to quantify the impact of driver distraction on a real road network from where the data was collected and at different traffic conditions using the SUMO simulator [7].

II. Related Work

Many researchers have focused on studying the impact of driving distraction on crash risk using different approaches. One of these approaches is *utilizing large-scale NDS datasets*, in which driver behavior is recorded in the real environment [4] by using cameras and different sensors. For instance, Lv et al. [8] employed 208 events with distracted driving and 373 events without distracted driving extracted from the SHRP2 database [9] to analyze right-turn drivers' distracted driving behavior at intersections based on real driver observation data collected from NDS. Arvin and Khattak [10] utilized SHRP2 to provide analysis of driving impairments and distractions, in addition to studying the duration of these distractions and their impact on critical events, such as crashes or near-crashes. While Zhang et al. [11] studied the influence of using mobile as a distraction task on the driver's control behaviors by utilizing a sample of 134 cases extracted from Shanghai Naturalistic Driving Study data (SH-NDS). Klauer et al. [12] used 100-Car NDS and studied crash and near-crash events and concluded that distracting tasks such as texting, eating and phone calls can significantly increase the probability of crash and near-crash.

Driving simulation experiment is the most commonly used method to study distracted driving. It depends on allowing the driver to conduct operations in a virtual environment. Karthaus et al. [13] investigated the influence of visual and acoustic distraction on the driving process with driving simulation and studied the difference in the performance of the older and younger vehicle drivers.

To study the effects of distracted driving on traffic safety and efficiency, *traffic simulators* is a suitable approach, where information of the speed and the locations of surrounding vehicles is obtained and utilized. Lint and Calvert [14] proposed a multi-level microscopic traffic modelling and simulation framework which studied the distraction impact on increasing traffic jams. Lindorfer et al. [15] employed traffic simulator *TrafficSim* to examine the influence of distracted driving and reaction time variations on traffic safety and efficiency. The *statistical research* is also employed in examining distracted driving, which depends on using datasets, such as accident data released by police reports, to analyze the correlation between traffic accidents and distracted driving [4]. For instance, Pope et al. [16] presented an analysis of adolescent drivers from the Traffic Safety Culture Index (TSCI) from 2011 to 2017 using a sample contains 3565 participants to examine support for distracted driving laws.

Although considerable work has been managed on identifying different types of distractions and studying the impact of distraction on crash probability, the impact of various distraction duration on traffic safety has not been deeply studied. Some weakness points in the literature need to be addressed. First, many studies depend on using driving simulators instead of naturalistic datasets in studying crash risk to achieve safety surrogate measures. Second, some particular research objectives such as studying traffic safety and efficiency can not be achieved by using naturalistic driving experiments. This paper addresses this gap by presenting a study of the different distraction tasks and calculating their distraction intervals with realistic values. Moreover, this work provides a study of the impact of various distraction tasks on traffic safety using traffic simulation.

III. The SHRP2 Dataset Processing and Modelling

SHRP2 NDS is considered the largest study of naturalistic driving behaviors to date [17]. The data is collected continuously from a variety of sensors installed in the vehicles. It contains more than 35 million miles of driving data and more than 5 million trips [18]. The data reduction is processed by the Virginia Tech Transportation Institute (VTTI), where all the crashes and near-crash events are recorded in the final SHRP2 dataset. In addition to these crash and near-crash events, baseline events are selected randomly for normal driving, which include distraction and non-distraction events. Each record in the dataset summaries the content of an individual trip [9]. The SHRP2 dataset has important advantages compared to other datasets. It was collected by volunteers and not from police reports, so, drivers involved in crashes reported the real reasons that they might not report to police investigations. Therefore, it is trusted and dependable. Moreover, it contains diverse variables that can be used in different studies. In this work, we use the sample of the dataset that Queen's University obtained from VTTI. This sample contains a total of 28896 records. Each record has 38 fields. Among these fields, the most important are 1) *Secondary Task Start*, 2) *Secondary Task End*, 3) *Subject Reaction Start*, 4) *Event Severity*, and 5) *Secondary Task*. The first three fields are utilized to compute the distraction duration, the fourth describes the severity of the event (i.e., crash, near-crash, or normal), while the fifth describes the type of distraction if there is a distraction or "No secondary task" if the driver was not distracted when the event happened. Among the 28896 records in this sample, there are almost 20,000 records for baseline and 8758 records for the crash, near-crash, and crash-relevant events.

A. Data Preprocessing and Analysis

The data is processed and analyzed to extract the statistics needed for the simulation of driver distraction. Particularly, we performed data cleaning and merged similar tasks in a single secondary task. Then, the result is used to compute the crash ratio and distraction duration for different types.

Data cleaning: Before processing the data we noticed that some records have invalid values in the *Secondary Task Start*, *Secondary Task End*, and *Subject Reaction Start* fields. In these fields, negative values are used. Therefore, when computing the distraction durations, we exclude these records.

Merging tasks: The dataset contains 55 different distraction types, which are called *Secondary Tasks*, such as *Eating*, *Smoking*, *Drinking*, etc. To simplify the data, similar tasks are combined under the same type. For instance, *Applying make-up*, *Shaving* and *Other personal hygiene* are merged into *Hygiene* distraction type. This way reduces the number of distraction types from 55 to 15 types.

Computing the task ratios: In the simulation, each task must be simulated based on the ratio of its occurrences in the data. So, we computed the ratio of each of the 15 secondary tasks in both the Crash/ Near-Crash/crash-Relevant (CNCR) and the baseline events, which are shown in Fig. 2.

Crash ratio and distraction ratio: The dataset sample we have includes all the crash-related events (8758 events) in the original data. Among them, only 1097 are reported to be caused by driver distraction. Thus, within the 5,411,197 trips [9], there are 1097 accidents, i.e, the crash ratio = 1097/ 5,411,197= 0.02 %. On the other hand, this sample contains

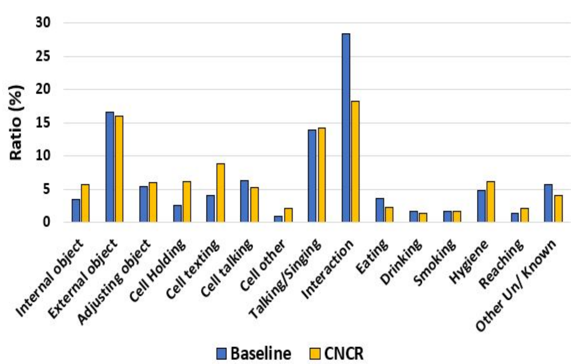


Fig. 2: The secondary tasks ratios

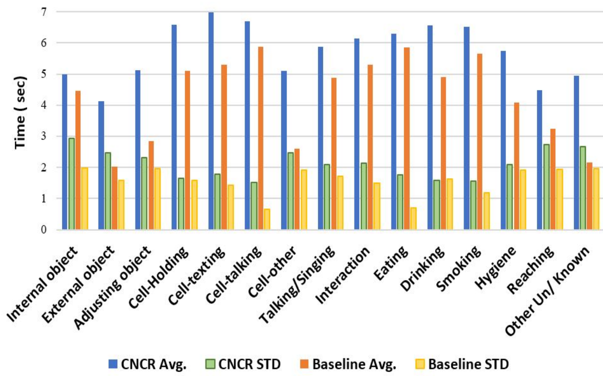


Fig. 3: The distraction intervals

only 20,000 records for other events, which is less than 0.4% of the total records in the original data. Therefore, this is not a representative sample of the data, and we cannot use it to compute the exact distraction ratio among all the trips.

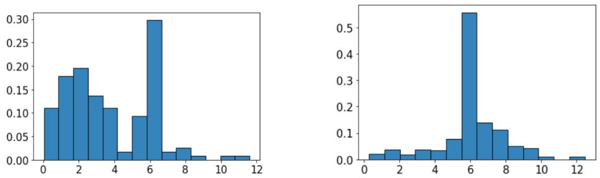
Calculate the distraction durations: Determining the distraction duration is important to know which task has a significant impact on crash risk. It is also a key parameter in the simulation because the simulation behavior will be changed based on the distraction interval. This duration is computed by subtracting the *Secondary Task Start* from *Secondary Task End* for each secondary task in baseline events. In CNCR, subtract the *Secondary Task Start* from *Subject Reaction Start*. The statistics of the duration as well as the standard deviation are shown in Fig. 3.

Based on this analysis, we can conclude that the distraction tasks that contribute the highest to crashes are *Cell phone usage* (including dialing, texting, etc.), *Interaction* (interacting with another person in the vehicle), *External object* (looking at an object outside the vehicle), and *Talking/Singing* with ratios of 22.29%, 18.18%, 16.07%, and 14.24%, respectively (see Fig. 2). Moreover, the tasks that are related to the cell phone have longer distraction durations as shown in Fig. 3.

B. Distraction Duration Modelling Using GMM

The study of data distribution is crucial to represent the data efficiently, particularly if this data is utilized as the first step of more complicated systems, like the simulation framework we present in this paper. The significant challenge we face in representing the distraction duration in SHRP2 is that each secondary task has a different distribution, as shown in Fig. 4. Some types can be modelled using the normal distribution. But others cannot be efficiently modelled by any known distribution. We experimented with several statistical distributions such as *Normal*, *Gamma*, *Beta*, etc. Most tasks did not fit any of the known distributions. Therefore, finding a unimodal distribution to fit the data is not possible, and it is necessary to use a mixture of models to represent the data efficiently. For this purpose, we utilize the GMM [19].

Gaussian distribution is a common distribution in modelling real-world unimodal data. Therefore, GMM is employed in modelling multimodal data with keeping the theoretical and computational benefits of Gaussian models [19]. Using GMM,



(a) Reaching (b) Cell texting

Fig. 4: The histogram of some tasks with different distributions

data is represented by a number ($k > 0$) of Gaussian distributions. A GMM with k components is parameterized by three parameters for each component i; the component weight ϕ_i, the component means μ_i and variances/covariances σ_i [20]. The final data model is the weighted sum of these components, as illustrated by Eq. (1)

$$p(x) = \sum_{i=1}^{k} \phi_i \mathcal{N}(x|\mu_i, \sigma_i) \tag{1}$$

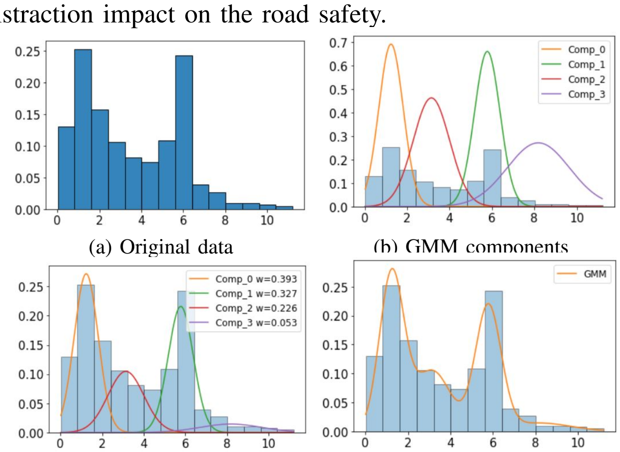
where $\mathcal{N}(x|\mu_i, \sigma_i)$ is a normal distribution, i.e.,

$$\mathcal{N}(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right), \tag{2}$$

and the summation of the wights equals to 1, i.e.,

$$\sum_{i=1}^{k} \phi_i = 1 \tag{3}$$

Fig. 5 shows an example for the generated GMM represent the distraction duration data for the *Adjusting Object* task. The generated data models and the statistics computed from the dataset are utilized to realistically simulate the driver distraction impact on the road safety.



(a) Original data (b) GMM components

(c) Weighted components (d) GMM

Fig. 5: The GMM for "Adjusting Object" task

IV. The Data-Driven Simulation Framework

Modelling the impact of driver distraction on the safety of the vehicles as well as the traffic flow is a challenging task because of the complexity of the system and the parameters that may affect the system's efficiency. Such a study requires information such as vehicle counts, speeds of surrounding vehicles and their locations, road network information and traffic lights and their timing. All these values are difficult to be acquired using naturalistic driving experiments. More importantly, such a study requires microscopic modelling for vehicle mobility and vehicle dynamic, as well as the interactions between vehicles that are in proximity. Therefore, many studies depend on utilizing driving simulation or traffic simulation [4]. In this paper, we use the SUMO software, which is a microscopic traffic simulator that can model all these parameters and capture these impacts.

The proposed simulation framework is illustrated in Fig. 6. In this framework, the analysis and modelling steps of the processed data generate the statistics and the data GMM models, respectively, as described earlier. The core of the framework is the Distraction Simulation Controller (DSC), which uses these data models and the statistics as well as the network and traffic setting to control the simulation parameters of SUMO and the simulated drivers within it.

A. The SUMO Simulator

SUMO is an open-source microscopic road traffic simulation software. It has a suite of tools for different functionalities, such as importing road networks, generating vehicular traffic, and computing routes from different sources to different destinations. One of the main capabilities of SUMO is the modelling of the vehicle dynamics for both longitudinal and lateral movements, which accounts for the vehicle parameters such as maximum speed and maximum acceleration/deceleration rates. It can also model the different types of traffic control strategies (traffic lights, stop signs, and yield signs). The vehicular traffic load in SUMO is represented by traffic flows. Each flows represents a number of vehicles that start at a given time from an origin with the aim to reach a destination. Accidents can happen in SUMO during car-following (a vehicle collides with its leader vehicles), during lane changing (a vehicle collides with a vehicle on the other lane) or when crossing an intersection (when two conflicting vehicles cross an intersection at the same time). However, by using the appropriate configurations the SUMO can run the network scenario without accidents.

There are several parameters to control the simulation environment in SUMO that may affect road safety. Some of these parameters are global that control the overalls simulation system, others are scope-defined parameters that only change the behavior of certain entities. An important global parameter is the $step\_length$ (whose default value is 0.1 second), which determines how often the software

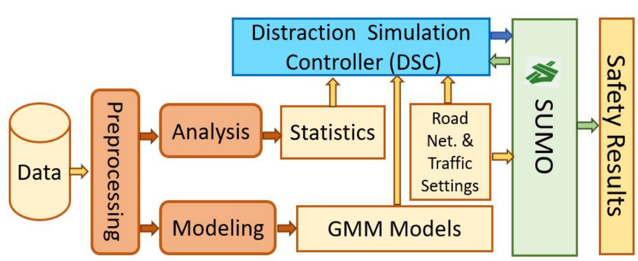


Fig. 6: The data-driven framework

will update the vehicles' parameters (such as location, speed, and acceleration). Another mandatory global parameter is the Traffic Scaling Factor (TSF), which is multiplied by each flow rate to scale it up or down. The TSF determines the total number of vehicles that will be generated in the network, and consequently, the vehicle density during the run time. Examples for vehicle-scope parameters are the vehicle's maximum speed and maximum acceleration, which can be specified for each individual vehicle. An other important vehicle-scope parameter is the $actionStepLength$, and the tau parameters. The $actionStepLength$ is initialized to the $step\_length$ and defines how often this vehicle will update its parameters. The tau parameter is utilized to model a driver's desired time headway (in seconds) which is the minimum time gap between follower and leader vehicle.

B. Driver Distraction Simulation

Since driver distraction means that the driver is not focusing on his main task (driving), and is occupied by a secondary task. This means that during the distraction interval the distracted driver does not account for the surrounding vehicles and does not react to the actions. Based on this concept, the driver distraction can be simulated in SUMO by disabling the updating of the vehicle parameters during the distraction interval. This can be achieved in SUMO by setting the $actionStepLength$ parameter to the distraction duration. Then, at the end of the distraction duration, this parameter should be reset back to its default value.

Time-Step Distraction Probability (TSDP): To simulate the distractions, we need to calculate the ratio of the drivers that are distracted. Since the dataset does not have enough information to calculate this ratio, we use another methodology. First, we define the TDSP in any time step as the probability of distracting one driver in this time step. The TSDP is an input parameter for the simulation framework and it should be computed before the simulation starts. To compute a realistic value for the TSDP, we use a simulation-based method to find the TSDP that generates the average crash ratio in the dataset, as will be explained later in the experimental results section.

C. The Distraction Simulation Controller (DSC)

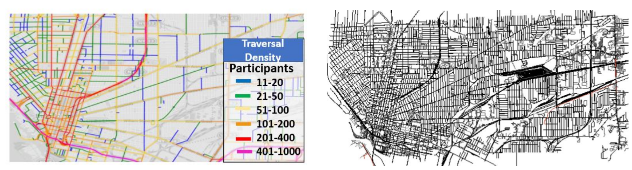
The DSC is responsible for changing the vehicle parameters during the run time to simulate the driver distraction. It follows the following procedures every time step.

1) ***Selecting a secondary task:*** The DSC uses the TSDP to find if a driver should be distracted in this time step. If so, it probabilistically selects a task based on the task probabilities computed previously in the analysis step and saved in the data statistics.
2) ***Generating a distraction interval:*** Then, the DSC uses the GMM model for the selected task to randomly generates a distraction interval.
3) ***Distracting drivers:*** If a driver must be distracted in the current time step, the DSC randomly selects one of the undistracted vehicles that are currently running in the network and changes its $actionStepLength$ to the generated distraction interval.
4) ***Resetting distracted drivers:*** Every time step, the DSC checks if any of the current distractions should be ended, and resets those vehicles to the default $actionStepLength$.

This way, we can model driver distraction and capture its impacts on the vehicle's safety by collecting the accident statistics generated by SUMO.

V. The Experimental Results

We use the developed framework to study the impact of driver distraction on a real network from which the data was collected at different traffic conditions. The SHRP2 NDS dataset was collected from six sites around the United States. The largest collection sites are in Seattle, Washington; Tampa, Florida; and Buffalo, New York, where each of these sites collected over 20% of the data. The advantage of using the SHRP2 networks is showing the number of participants who used each network, and this shows which network is mostly employed and lead to close results. Therefore, we utilized one of these networks; Buffalo, New York. Fig. 7a shows the network with the participant traversal density, and Fig. 7b shows the network snapshot from SUMO.



(a) Traversal density (b) The simulated network

Fig. 7: Buffalo network

A. Calculating the TDSP

Before performing the sensitivity analysis, we need to compute a realistic TDSP value for the network. So, we run the network at different values for the TDSP. And we computed the distraction ratio that generates the same crash ratio of the original real dataset. According to the dataset, the crash ratio is 0.02. Therefore, we use the default simulation parameters, i.e., $step\_length = 0.1sec.$ and $tau = 1sec.$, and change the TSDP until the crash ratio reaches the required value of 0.02. Buffalo network reaches this ratio at TSDP of 0.002. These results are shown in Fig. 8.

B. Traffic Load Sensitivity Analysis

We perform sensitivity analysis to study the impact of the network traffic load on road safety. We use three different values of TSDP including the computed one (0.002). To achieve this goal, we run Buffalo network with different TSF ranging from 1 through 4, and for each TSF we compute the average crash ratio from 5 different runs with different seeds. The results are shown in Fig. 9, which also shows interesting results. It illustrates that the crash ratio increases with the TSF (i.e., the congestion level). However, when increasing the congestion level above a certain level, the crash ratio starts decreasing. This behavior appears with the three different values of TSDP. The reason behind this behavior is that at low traffic loads, the distances between vehicles are long. Hence, most of the distractions will not result in crashes because there are no other vehicles close to the distracted driver. As the traffic load increases (before entering the congested regime), the distances between cars become smaller while speeds are still relatively high, which increases the crash probability in case of distractions. Finally, when the network becomes congested, the distances between vehicles become smaller, vehicle speeds are significantly decreased (because of the congestion), which increases the headway time, and consequently contributes to decreasing the crash probability or being constant.

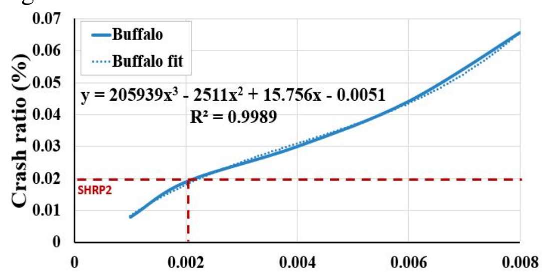


Fig. 8: TSDP for the Buffalo network

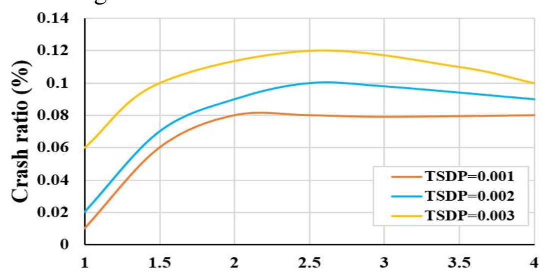


Fig. 9: The distraction impact versus TSF

VI. Conclusion

In this paper, we introduce a data-driven simulation framework for driver distraction, which can be utilized to study the impact of driver distraction on road safety based on the SHRP2 naturalistic dataset. The data analysis reveals that most of the driver distractions are caused by four main types of secondary tasks, namely, *Cell phone usage*, *Interaction*, *External object*, and *Talking/Singing*. We show that, 22% of the accident in the dataset are caused by *Cell phone usage*. Another important conclusion is that the distribution of the distraction intervals of many distraction types are multi-modal distributions and the best way to model them is to use a mixture of models. More importantly, our sensitivity analysis of the distraction impacts under different traffic loads demonstrated that the impact increases with the traffic load, however at a certain point (when the network becomes congested) the impact of distraction decreases. In future work, we plan to study the effects of using vehicular communication to transfer the distraction information on road safety. Another extension for this work is utilizing real dataset to model the driver time headway to represent more realistic driver behavior.

Acknowledgement

This research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number RGPIN-2019-05667.

References

[1] World Health Organization, “Global status report on road safety 2018: Summary,” World Health Organization, Tech. Rep., 2018.

[2] F. Tango and M. Botta, “Real-time detection system of driver distraction using machine learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 894–905, 2013.

[3] A. Sharma, Y. Ali, M. Saifuzzaman, Z. Zheng, and M. M. Haque, “Human factors in modelling mixed traffic of traditional, connected, and automated vehicles,” in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 262–273.

[4] X. Liang, “A brief review of the impact of distracted driving on traffic safety,” in *IOP Conference Series: Earth and Environmental Science*, vol. 580, no. 1. IOP Publishing, 2020, pp. 1–9.

[5] A. A. Kandeel, H. M. Abbas, and H. S. Hassanein, “Explainable model selection of a convolutional neural network for driver’s facial emotion identification,” in *ICPR International Workshops and Challenges*. Springer, 2021, pp. 699–713.

[6] M. A. Regan, C. Hallett, and C. P. Gordon, “Driver distraction and driver inattention: Definition, relationship and taxonomy,” *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1771–1781, 2011.

[7] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, “Microscopic traffic simulation using sumo,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[8] B. Lv, R. Yue, and Y. Zhang, “The influence of different factors on right-turn distracted driving behavior at intersections using naturalistic driving study data,” *IEEE Access*, vol. 7, pp. 137 241–137 250, 2019.

[9] “SHRP2 NDS data,” accessed: 25.04.2021. [Online]. Available: https://insight.shrp2nds.us.

[10] R. Arvin and J. Khattak, “Driving impairments and duration of distractions: Assessing crash risk by harnessing microscopic naturalistic driving data,” *Accident Analysis & Prevention*, vol. 146, pp. 1–17, 2020.

[11] L. Zhang, B. Cui, M. Yang, F. Guo, and J. Wang, “Effect of using mobile phones on driver’s control behavior based on naturalistic driving data,” *International journal of environmental research and public health*, vol. 16, no. 8, pp. 1–13, 2019.

[12] S. G. Klauer, F. Guo, B. G. Simons-Morton, M. C. Ouimet, S. E. Lee, and T. A. Dingus, “Distracted driving and risk of road crashes among novice and experienced drivers,” *New England journal of medicine*, vol. 370, no. 1, pp. 54–59, 2014.

[13] M. Karthaus, E. Wascher, and S. Getzmann, “Effects of visual and acoustic distraction on driving behavior and eeg in young and older car drivers: a driving simulation study,” *Frontiers in aging neuroscience*, vol. 10, pp. 1–15, 2018.

[14] J. Van Lint and S. C. Calvert, “A generic multi-level framework for microscopic traffic simulation—theory and an example case in modelling driver distraction,” *Transportation Research Part B: Methodological*, vol. 117, pp. 63–86, 2018.

[15] M. Lindorfer, C. Backfrieder, C. F. Mecklenbräuker, and G. Ostermayer, “Investigating the large-scale effects of human driving behavior on vehicular traffic flow,” in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2018, pp. 204–215.

[16] C. N. Pope, A. Nwosu, T. M. Rudisill, and M. Zhu, “Support for distracted driving laws: An analysis of adolescent drivers from the traffic safety culture index from 2011 to 2017,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 78, pp. 424–432, 2021.

[17] A. Abdelrahman, H. S. Hassanein, and N. Abu-Ali, “A cloud-based environment-aware driver profiling framework using ensemble supervised learning,” in *IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[18] J. M. Hankey, M. A. Perez, and J. A. McClafferty, “Description of the shrp 2 naturalistic database and the crash, near-crash, and baseline data sets,” Virginia Tech Transportation Institute, Tech. Rep., 2016.

[19] D. A. Reynolds, “Gaussian mixture models.” *Encyclopedia of biometrics*, vol. 741, pp. 659–663, 2009.

[20] N. Shental, A. Bar-Hillel, T. Hertz, and D. Weinshall, “Computing gaussian mixture models with em using equivalence constraints,” *Advances in neural information processing systems*, vol. 16, no. 8, pp. 465–472, 2004.