Characterizing Heterogeneous Car-Following Behaviors of Human Drivers in Mixed Traffic

Meng Sun
Thrust of Robotics and Autonomous Systems
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
msun897@connect.hkust-gz.edu.cn

Chunxi Huang
Thrust of Intelligent Transportation
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
tracy.huang@connect.ust.hk

Dengbo He*
Thrust of Intelligent Transportation
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
dengbohe@hkust-gz.edu.cn

Abstract*—As more autonomous vehicles (AVs) are tested or deployed on public roads, human-driven vehicles (HVs) and AVs will have to share the road in the near future. However, the interactions between HVs and AVs are under-investigated, especially how HVs behave differently when following AVs versus following HVs and how this potential discrepancy impacts traffic safety and stability. To address this gap, based on the Lyft dataset, our study explored the differences in two types of car-following events, i.e., HV-following-AV and HV-following-HV. Hierarchical clustering was used to classify the driving styles of the following vehicles in all following events. Then, driving volatility measures and surrogate safety measures, i.e., speed volatility and reciprocal time-to-collision (reTTC) value, were used to compare the outcomes of different car-following events. The clustering results showed that human drivers exhibited four different car-following styles in the mixed traffic and when following AVs and HVs, drivers may exhibit different following styles. Compared to the leading HV, the leading AV resulted in a smaller TTC (i.e., a larger reTTC), regardless of the driving styles of following HVs. When driving at low speeds, drivers had lower speed volatility when following AVs, indicating more stable traffic; while at high speed, when drivers were not sensitive to changes of the speed of the vehicle ahead, the leading AVs led to larger speed volatility. Future AV control algorithms should take the styles of the following HV into consideration in order to better optimize traffic stability and improve the safety of mixed traffic.

Keywords—autonomous driving, car-following, driving volatility, traffic safety, cluster analysis

I. INTRODUCTION

In recent years, with the profound development of computer, sensor, and artificial intelligence technology, the autonomous vehicle (AV) has evolved from a laboratory concept to a service that is available in the real world. Many cities have allowed testing and deployment of AVs on public roads, leading to mixed traffic consisting of AVs and human-driven vehicles (HVs). The coexistence of AVs and HVs may fundamentally change traffic characteristics (e.g., impact driving behavior patterns in traffic streams [1][2][3], influence driving safety in mixed AV-HV traffic [4][5][6]) and thus introduce challenges to traffic management.

The car-following scenario is one of the most common but challenging scenarios in mixed traffic. Previous research based on crash reports found that rear-end crashes dominate among AV-involved crashes [7][8][9] and that AVs are more likely to be rear-ended by the following HVs compared to those of HVs [10][11] in the car-following events. It is possible that the AVs are new to most drivers [9], and thus their less predictable behaviors may have contributed to the high chance of AVs being rear-ended [3]. However, the crash reports cannot provide operational-level information regarding how HVs respond to AVs and thus can hardly guide the optimization of the AV control algorithms and traffic management policies. To address this research gap, empirical studies were conducted, for example, in field experiments. It was found that human drivers drove more smoothly and had lower crash risk when following AVs compared to when following HVs [12]. Another study found that when following AVs, HVs exhibited lower driving volatility in terms of speed and acceleration, and exhibited larger TTCs [13].

However, the field experiments are still contrived to some extent and may not reveal the naturalistic behaviors of the HVs. Thus, with the AVs being tested on public roads, more analyses were conducted based on real-world datasets, and the Waymo open dataset [14] has been most widely adopted in previous studies. For example, it was found that, compared to following HVs, when following AVs, drivers exhibited longer response time [15], lower speed and acceleration volatility, shorter time headway, higher TTCs, [16] and larger gap distance [3].

However, the previous analysis was not able to account for the heterogeneity of the HV drivers in the car-following events. When following HVs, previous research found that drivers may exhibit different driving styles (e.g., risky or conservative) in different situations [17][18][19]. When following AVs, although it was found that human drivers may still exhibit different styles based on acceleration volatility, time headway, and vehicle speed in the Waymo dataset [16], the results were inconclusive given the small sample size in the Waymo dataset (i.e., 264 HV-following-AV events). Further, with the small sample size, it was infeasible to analyze the outcomes (i.e., safety and speed volatility) of the HV-following-AV events. Recently, a new car-following dataset based on the Lyft level-5 dataset was released [20], which contains 29,449 HV-following-AV and 42,892 HV-following-HV events. Thus, to address the above-mentioned research gaps, a cluster analysis was conducted to investigate the heterogeneity of HV drivers' car-following behaviors in mixed traffic and further quantify the impact of the driving style on the safety and volatility of mixed traffic.

II. METHOD

A. Data Description and Data Pre-processing

The Lyft level-5 dataset is a large-scale dataset collected by a fleet of 20 AVs [21]. This dataset includes over 1000 hours of perception and motion data collected in cities and suburbs over a period of 4 months (from October 2019 toMarch 2020) on fixed routes in Palo Alto, California, and with a sampling frequency of 10 Hz. However, the raw trajectory data is noisy and can hardly be applied for car-

* Corresponding author.

979-8-3503-1579-0/24/$31.00 ©2024 IEEE

following behavior analysis directly. Thus, in this study, we adopted the cleaned data published by [20], in which the car-following pairs (consisting of one lead vehicle and one following vehicle) were identified. In addition, as we are only interested in car-following events when the traffic is flowing, we added two criteria to further refine the screening results: 1) the leading and following vehicles are both traveling in the same direction and do not change the direction of travel; 2) The speed of the following vehicle should be equal to or over 1 m/s [22].

As a result, we obtained 19,754 HV-following-AV and 22,840 HV-following-HV events. The average length of the event was 21.25 seconds, ranging from 14.2 seconds to 92.2 seconds.

B. Variables Extraction

In order to characterize the car-following behaviors in the events, multiple metrics were extracted from each event, including the mean (mean_lead_vel) and standard deviation (SD) of the lead vehicle speed (Std_lead_vel), the minimum and maximum acceleration of the following vehicle (min_follow_ac and max_follow_ac), the speed volatility of the following vehicle (V_f_follow_vel), and the time headway (TH) and the reciprocal time-to-collision (reTTC) in the event. In this study, all metrics were calculated at the event level. In other words, one set of metrics was obtained from each car-following event. The calculations of the V_f-follow-vel, TH and reTTC are as follow, while the calculations of other metrics are straightforward and thus omitted in this paper.

1) Speed volatility of the following vehicle

Previous studies have shown that speed volatility is not only closely related to collision risk but also to energy consumption and traffic efficiency [13][23]. Thus, time-varying stochastic volatility (V_f) was used to quantify speed fluctuations as the observed proportional changes:

$$V_f = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \bar{r})^2}{n-1}} \tag{1}$$

$$r_i = ln\left(\frac{x_i}{x_{i-1}}\right) * 100 \tag{2}$$

where, ln is the natural logarithm, $\bar{r}$ is the mean of r_i, x_i is the i^{th} observation, $\bar{x}$ is the mean of observations and n is the sample size.

2) Time headway

Time headway (TH) is defined as the elapsed time between two consecutive vehicles passing through the same point [24] and can be calculated as follows:

$$TH(t) = \frac{X_{L(t)} - X_{F(t)}}{V_{F(t)}} \tag{3}$$

where $X_{L(t)}$ and $X_{F(t)}$ are the positions of the leading and following vehicles at time t, and $V_{F(t)}$ is the velocity of following vehicle at time t. Time headway is a safety measure of car-following behavior as it is related to the available time for the driver to respond to an event [10].

3) Reciprocal time-to-collision

Time-to-collision (TTC) is an surrogate safety measure introduced by [25] and can quantify rear-end crash risks. If two consecutive vehicles continue to travel at the current speed while occupying the same lane, TTC represents the time it takes for the vehicle to collide. To avoid infinite or negative TTC, in this study, the reciprocal of the TTC (i.e., reTTC) was used, which is positively proportional to the risk of collision, where a higher value of reTTC indicates a higher risk of collision. When the speed of the leading vehicle was equal to or faster than that of the following vehicle, reTTC was set to 0 [3]. The reTTC was calculated as follows:

$$reTTC = \frac{V_{F(t)} - V_{L(t)}}{X_{L(t)} - X_{F(t)} - L} \tag{7}$$

where $V_{L(t)}$ is the velocity of lead vehicle at time t, and L is the length of the leading vehicle.

C. Hierarchical Clustering

To explore the heterogeneity of the car-following styles, agglomerative hierarchical clustering [26] with a bottom-up approach was implemented. The linkage method was selected from Ward, single, complete and average. The Euclidean distance functions were utilized [27]. To better differentiate driving styles under different conditions, based on previous studies describing car-following behaviors [16][10][28][29], we selected the following features for cluster analysis, including Std_lead_vel, mean_lead_vel, min_follow_ac, max_follow_ac, mean TH (mean_TH), and minimum TH (min_TH) within a car-following event. Among the metrics, the Std_lead_vel and mean_lead_vel can describe the characteristics of the traffic condition (i.e., speed and the stability of the traffic flow) and the mean_lead_vel, min_follow_ac, max_follow_ac, mean_TH, and min_TH can reflect the driving style of the following vehicle. We adopted the Z-score standardization method [30] based on the mean and SD of each feature to standardize the impact of the metrics on the clustering results.

Then, a dendrogram was used to visualize the results of hierarchical clustering, and the Calinski-Harabasz Index and Silhouette Coefficient [31] were used to support our decision on the type of linkage and the number of clusters. The Calinski-Harabasz Index is the ratio of the inter-cluster distance to the intra-cluster distance, with a larger value indicating a better result. The Silhouette Coefficient takes into account the compactness within the clusters and separation between the clusters, ranging within [-1, 1], and a larger value indicates a better clustering result.

D. Statistical Analysis

First, to compare the difference in car-following styles when the lead vehicle was HV versus AV, a multinomial logistic regression model was conducted to explore how car-following types affect drivers' car-following styles.

Then, to evaluate the outcomes of different driving styles under different conditions, multiple linear regression models were constructed with the car-following type (i.e., HV-following-HV and HV-following-AV) and driving style (based on clustering results) as independent variables, and maximum reTTC (max_reTTC) and V_f_follow_vel in the car-following events as dependent variables. Specifically, max_reTTC within each following event can assess the safety of the following behavior, while V_f_follow_vel can indicate the stability of the speed during the event. Post-hoc analyses based on Tukey's test were conducted for all significant (p < .05) independent variables. All models were built in *R 4.3.1.*

III. RESULTS

A. Clustering Results of Driving Styles

The Wald linkage yielded the best performance and the dendrogram of agglomerative hierarchical clustering is shown in Fig. 1. Based on the Calinski-Harabasz Index (Fig. 2) and Silhouette Coefficient (Fig. 3), we decided to categorize the car-following styles into four clusters in different car-following conditions.

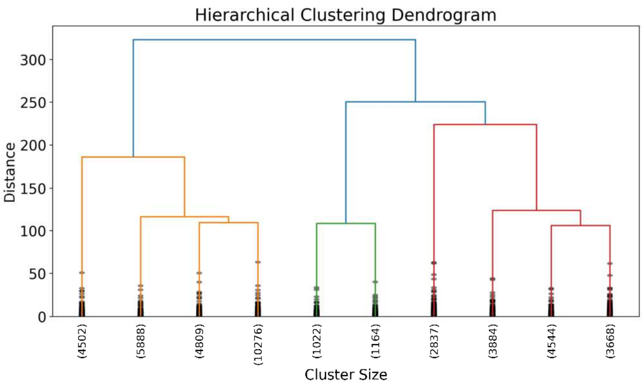

Fig. 1 Dendrogram of agglomerative hierarchical clustering. The numbers along the x-axis are the number of cases in each cluster.

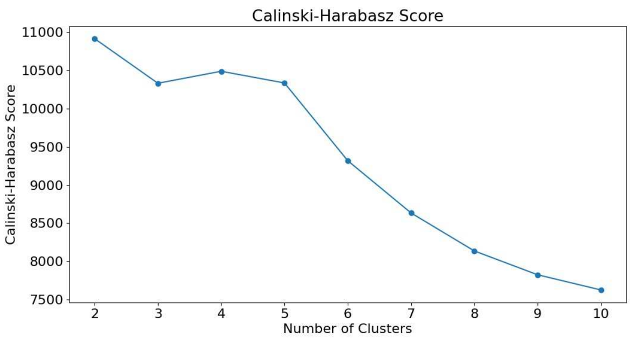

Fig. 2 Calinski-Harabasz Index for different number of clusters

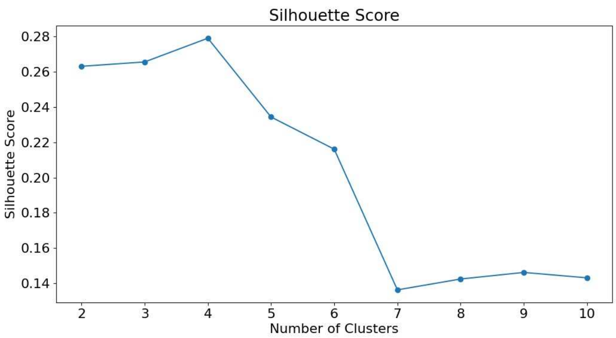

Fig. 3 Silhouette Coefficient for different number of clusters

As shown in Fig. 4, the mean speed of the lead vehicle ($F = 21635$, $p < .0001$), SD of the lead vehicle speed ($F = 28914$, $p < .0001$), minimum acceleration of the following vehicle ($F = 9349$, $p < .0001$), maximum acceleration of the following vehicle ($F = 3911$, $p < .0001$), time headway mean ($F = 1325$, $p < .0001$), and minimum time headway ($F = 155.1$, $p < .0001$) were different across different clusters. Table I further provides the pair-wise comparisons of the features that can differentiate different clusters, including the mean difference (Δ), 95% confidence interval (95%CI), t-value and *p*-value. It should be noted that the degree-of-freedoms for all t tests in Table I are 42592.

TABLE I. POST-HOC COMPARISONS OF FETURRES ACROSS CLUSTERS

Features	Cluster	Δ (95%CI)	t-value	*p*-value
mean_ lead_vel	1-2	4.86 (4.75, 4.96)	t = 91.07	< .0001
	1-3	1.21 (1.11, 1.30)	t = 25.51	< .0001
	1-4	4.24 (4.19, 4.29)	t = 160.75	< .0001
	2-3	-3.64 (-3.78, -3.51)	t = -53.57	< .0001
	2-4	-0.61 (-0.72, -0.50)	t = -10.95	< .0001
	3-4	3.04 (2.94, 3.13)	t = 60.89	< .0001
Std_lead_vel	1-2	-1.08 (-1.12, -1.05)	t = -59.26	< .0001
	1-3	-0.44 (-0.47, -0.41)	t = -27.22	< .0001
	1-4	-1.62 (-1.64, -1.61)	t = -179.71	< .0001
	2-3	0.64 (0.59, 0.69)	t = 27.48	< .0001
	2-4	-0.54 (-0.58, -0.51)	t = -28.56	< .0001
	3-4	-1.18 (-1.22, -1.15)	t = -69.30	< .0001
min_ follow_ac	1-2	0.32 (0.29, 0.35)	t = 23.44	< .0001
	1-3	0.49 (0.46, 0.51)	t = 40.06	< .0001
	1-4	0.64 (0.62, 0.65)	t = 94.06	< .0001
	2-3	0.17 (0.13, 0.20)	t = 9.50	< .0001
	2-4	0.32 (0.29, 0.3445)	t = 22.21	< .0001
	3-4	0.15 (0.13, 0.18)	t = 11.78	< .0001
max_ follow_ac	1-2	-0.23 (-0.26, -0.19)	t = -11.75	< .0001
	1-3	-4.60 (-4.64, -4.57)	t = -270.84	< .0001
	1-4	-0.46 (-0.48, -0.44)	t = -48.51	< .0001
	2-3	-4.38 (-4.43, -4.33)	t = -179.15	< .0001
	2-4	-0.24 (-0.27, -0.20)	t = -11.78	< .0001
	3-4	4.14 (4.11, 4.18)	t = 231.29	< .0001
mean_TH	1-2	-3.03 (-3.06, -3.00)	t = -196.49	< .0001
	1-3	-0.02 (-0.05, 0.01)	t = -1.57	.1
	1-4	-0.41 (-0.42, -0.39)	t = -53.54	< .0001
	2-3	3.01 (2.97, 3.05)	t = 152.78	< .0001
	2-4	2.62 (2.59, 2.65)	t = 162.99	< .0001
	3-4	-0.39 (-0.42, -0.36)	t = -26.85	< .0001
min_TH	1-2	-2.17 (-2.20, -2.15)	t = -166.80	< .0001
	1-3	0.05 (0.03, 0.08)	t = 4.56	< .0001
	1-4	-0.12 (-0.14, -0.11)	t = -18.99	< .0001
	2-3	2.23 (2.19, 2.26)	t = 133.80	< .0001
	2-4	2.05 (2.02, 2.08)	t = 150.93	< .0001
	3-4	-0.18 (-0.20, -0.15)	t = -14.38	< .0001

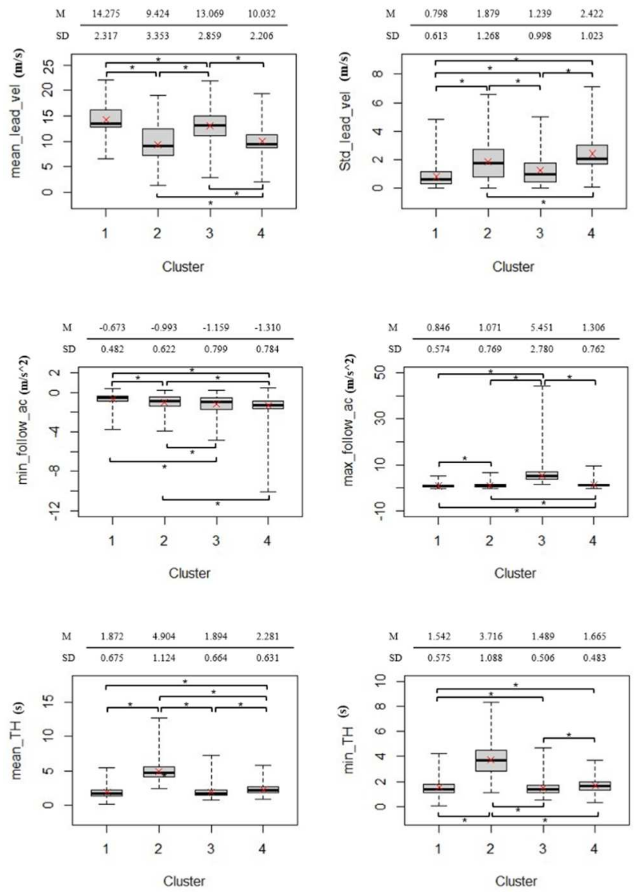

Fig. 4 Boxplots of features across clusters. Boxplots show the five numbers summary as well as the mean values indicated by the red crosses. Mean (M) and standard deviation (SD) values are also provided at the top of each plot. Note that * marks significant results ($p < .05$).

Based on the features, we label the four clusters as follows:

HiVel-LSens-Agg (cluster 1, n = 25,475): high-speed low-sensitivity aggressive ones are those who kept a small headway in a car-following situation at relatively high speed and were insensitive to speed changes of lead vehicles (relatively small max_follow_ac and min_follow_ac).

LoVel-Con (cluster 2, n = 2,186): low-speed conservative ones are those who kept a large headway in a car-following situation at relatively low speed.

HiVel-HSens-Agg (cluster 3, n = 2,837): high-speed high-sensitivity aggressive ones are those who kept a small headway in a car-following situation at relatively high speed and were sensitive to speed changes of lead vehicles.

LoVel-Agg (cluster 4, n = 12,096): high-speed aggressive ones are those who kept a small headway in car-following situation at relatively low speed.

The results of the multinominal model showed that, compared to following HVs, when following AVs, drivers were less likely to be identified as **HiVel-LSens-Agg** (odds ratio [OR] = 0.51, 95% CI: [0.49, 0.54], $\chi^2(1) = 888.60$, $p < .0001$**), LoVel-Con** (OR = 0.33, 95% CI: [0.30, 0.37], $\chi^2(1) = 495.78$, $p < .0001$) and **HiVel-HSens-Agg** (OR = 0.62, 95% CI: [0.57, 0.67], $\chi^2(1) = 133.23$, $p < .0001$) instead of **LoVel-Agg** ones; drivers were more likely to be identified as **HiVel-LSens-Agg** (OR = 1.54, 95% CI: [1.41, 1.70], $\chi^2(1) = 83.06$, $p < .0001$**)** instead of **LoVel-Con** ones; and drivers were less likely to be identified as **HiVel-LSens-Agg** (OR = 0.83, 95% CI: [0.77, 0.90], $\chi^2(1) = 21.38$, $p < .0001$**)** and **LoVel-Con** (OR = 0.54, 95% CI: [0.48, 0.61], $\chi^2(1) = 108.28$, $p < .0001$) instead of **HiVel-HSens-Agg** ones.

B. Statistical Analysis of Outcomes of Driving Styles

Table II summarizes the statistical results for the speed volatility of the following vehicle (V_f_follow_vel) and maximum reciprocal time-to-collision (max_reTTC) in all car-following events. Significant interaction effects of clusters and car-following types have been observed for max_reTTC and V_f_follow_vel and post-hoc comparisons are presented in Table III, Table IV and in Fig. 5. The degree-of-freedoms of t tests in Table III and IV are all 42592.

TABLE II. TYPE III CONTRASTS OF OUTCOME-RELATED VARIABLES

Dependent Variables	Independent Variables	F value	*p*-value
Vf_follow_vel	Cluster	F(3, 42586) = 2493.925	< .0001
	Car-following type	F(1, 42586) = 22.276	< .0001
	Cluster*car-following type	F(3, 42586) = 352.405	< .0001
max_reTTC	Cluster	F(3, 42586) = 3072.56	< .0001
	Car-following type	F(1, 42586) = 997.92	< .0001
	Cluster*car-following type	F(3, 42586) = 64.92	< .0001

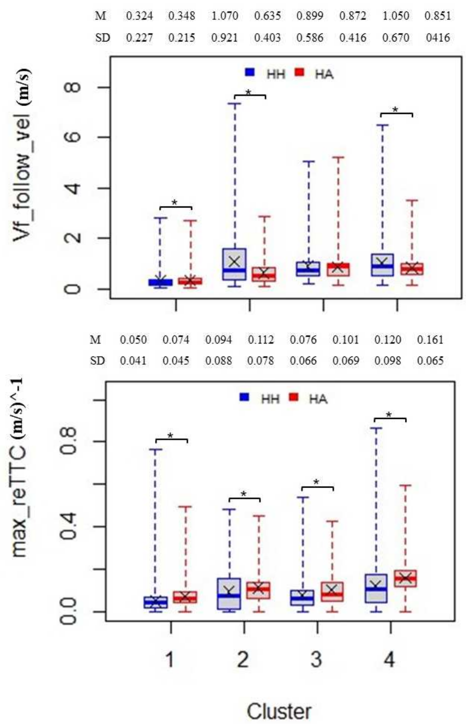

Fig. 5 The post-hoc results for car-following types. Mean (M) and standard deviation (SD) values are also provided at the top of each plot. Note that HA stands for the HV-following-AV events, HH represents the HV-following-HH events, and * marks significant results (p < .05).

TABLE III. POST-HOC RESULTS FOR SPEED VOLATILITY

Cluster	Car-following type	Δ (95%CI)	t-value	*p*-value
HiVel-LSens-Agg	HA-HH	0.02 (0.01, 0.03)	t = 4.72	< .0001
LoVel-Con	HA-HH	-0.44 (-0.47, -0.40)	t = -23.64	< .0001
HiVel-HSens-Agg	HA-HH	-0.03 (-0.06,0)	t = -1.81	.07
LoVel-Agg	HA-HH	-0.20 (-0.22, -0.19)	t = -53.57	< .0001

TABLE IV. POST-HOC RESULTS FOR RETTC

Cluster	Car-following type	Δ (95%CI)	t value	*p*-value
HiVel-LSens-Agg	HA-HH	0.024 (0.023, 0.025)	t = 31.59	< .0001
LoVel-Con	HA-HH	0.018 (0.013, 0.024)	t = 6.69	< .0001
HiVel-HSens-Agg	HA-HH	0.025 (0.020, 0.029)	t = 10.98	< .0001
LoVel-Agg	HA-HH	0.042 (-0.039, 0.044)	t = 37.80	< .0001

IV. DISCUSSION

In this study, we investigated the impact of driver heterogeneity and lead vehicle type (i.e., AV vs. HV) on the traffic stability and safety based on a real-world mixed traffic dataset, the Lyft level-5 dataset. Using a cluster analysis, we identified four types of drivers in car-following events.

In general, drivers kept relatively short headways at high speed, with mean headways smaller than the recommended safety headways of 2 seconds [32] and minimum headways smaller around 1.5 seconds. This finding is not surprising, as in the Waymo dataset, we also found that drivers kept small headways regardless of lead vehicle type [3]. However, in this study, we further observed heterogeneity in drivers' min and max accelerations in the car following events, which may reflect drivers' sensitive to speed variation of the lead vehicle. Based on this feature, we further categorized drivers into high and low sensitive ones, depending on the extreme acceleration values in the events. While at low speed, some drivers tended to keep large headway (over 4 seconds) and some tended to keep moderate headway (slightly over 2 seconds). However, the authors should be aware that, given the limitations of the open dataset (including Lyft and Waymo), we were not able to fully recover the driving scenario, which the car-following styles might be highly associated with. For example, in our dataset, the large accelerations among HiVel-HSens-Agg drivers might be attributed to the large speed standard deviation of the lead vehicle. Further research with more complete scenario information is needed to better reveal driver heterogeneity with the influence of traffic scenario decoupled.

In our study, a further multinominal model showed that, when following AVs versus HVs, the chance of drivers exhibiting different car-following styles differed. Still, there can be two interpretations regarding this finding. It is possible that drivers would exhibit different behaviors when following AVs versus following HVs, similar to what has been found in [3]. It is also possible that the different speed profile of the lead vehicle has led to different following styles of the following vehicles. A future study that matches the speed profile of the lead vehicle [16] should be conducted to further explain the discrepancy in the distribution of the driving styles.

Further, we quantified how lead vehicle type can impact traffic efficiency and traffic safety among drivers with different car-following styles. It was found that, in general, a leading AV led to a smaller TTC (i.e., a larger reTTC) compared to a leading HV, regardless of the style of the following vehicle. It is likely that the speed profile of the lead AV is less predictable compared to HV, and thus, drivers were less responsive to the speed change of the AV. This finding has been observed in [3], but we further found that the style of the following vehicle can influence the car following safety: aggressive following drivers were more dangerous compared to conservative ones at low speed, but the car following style did not make obvious differences at high speed.

However, though there is an increased safety concern with the leading AV, previous research found that the speed volatility was smaller in HV-following-AV events [13][16]. Our study found that this may not always be the case. The benefits of speed stability can only be observed at relatively low speed. In contrast, the leading AV may even lead to greater speed volatility if the following driver is not sensitive to the change of the leading vehicle speed. In [13][16], the drivers' max/min acceleration in the event was not captured. It is possible that when the drivers were less sensitive to the speed change of the lead vehicle at high speed, the unpredictable speed profile of the leading AV led to delayed and more abrupt responses to the speed change of the AV. However, we should point out that the effect size (Δ) of the difference between HV-following-AV and HV-following-HV events at high speed was small and the readers should be cautious when interpreting this result.

In summary, our results found that drivers tended to exhibit different driving styles when following AVs versus following HVs. Further, different driving styles may lead to different outcomes when the leading vehicle is an AV. Future AV control algorithms should take the style of the following HV into consideration in order to better optimize traffic stability and improve the safety of mixed traffic. Given that the differences in the car-following styles exist when following leading AVs versus HVs, future research should also consider driver heterogeneity when modeling mixed traffic. Future deep learning research may also be conducted to re-construct drivers' car-following behaviors in Lyft level-5 dataset and compare it with car-following models based on other datasets (e.g., [33]) in order to better understand how drivers' behaviors in mixed traffic change over time.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 52202425), and partially by the Municipal Science and Technology Project (No. 2023A03J0011) and the Guangzhou Science and Technology Program City-University Joint Funding Project (No. 2023A03J0001).

REFERENCES

[1] A. Sinha, S. Chand, K. P. Wijayaratna, N. Virdi, and V. Dixit, "Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles," *Accid. Anal. Prev.*, vol. 142, no. January, p. 105567, 2020, doi: 10.1016/j.aap.2020.105567.

[2] P. Liu, Y. Du, L. Wang, and J. Da Young, "Ready to bully automated vehicles on public roads?," *Accid. Anal. Prev.*, vol. 137, no. January, 2020, doi: 10.1016/j.aap.2020.105457.

[3] C. Huang, X. Wen, D. He, and S. Jian, “Sharing the Road: How Human Drivers Interact with Autonomous Vehicles on Highways,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 1437–1441, 2022, doi: 10.1177/1071181322661165.

[4] R. Arvin, A. J. Khattak, M. Kamrani, and J. Rio-Torres, “Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections,” *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 25, no. 2, pp. 170–187, 2020, doi: 10.1080/15472450.2020.1834392.

[5] M. S. Rahman and M. Abdel-Aty, “Longitudinal safety evaluation of connected vehicles’ platooning on expressways,” *Accid. Anal. Prev.*, vol. 117, no. December 2017, pp. 381–391, 2018, doi: 10.1016/j.aap.2017.12.012.

[6] L. Ye and T. Yamamoto, “Evaluating the impact of connected and autonomous vehicles on traffic safety,” *Phys. A Stat. Mech. its Appl.*, vol. 526, p. 121009, 2019, doi: 10.1016/j.physa.2019.04.245.

[7] A. M. Boggs, B. Wali, and A. J. Khattak, “Exploratory analysis of automated vehicle crashes in California: A text analytics & hierarchical Bayesian heterogeneity-based approach,” *Accid. Anal. Prev.*, vol. 135, no. October 2019, p. 105354, 2020, doi: 10.1016/j.aap.2019.105354.

[8] Q. Liu, X. Wang, X. Wu, Y. Glaser, and L. He, “Crash comparison of autonomous and conventional vehicles using pre-crash scenario typology,” *Accid. Anal. Prev.*, vol. 159, no. June, p. 106281, 2021, doi: 10.1016/j.aap.2021.106281.

[9] C. Huang *et al.*, “Characteristics of Rear-End Collisions : A Comparison between ADS-Involved Crashes and ADAS-Involved Crashes,” no. 52202425, pp. 1–16.

[10] G. Zhang, Y. Wang, H. Wei, and Y. Chen, “Examining headway distribution models with urban freeway loop event data,” *Transp. Res. Rec.*, no. 1999, pp. 141–149, 2007, doi: 10.3141/1999-15.

[11] Y. Song, M. V. Chitturi, and D. A. Noyce, “Automated vehicle crash sequences: Patterns and potential uses in safety testing,” *Accid. Anal. Prev.*, vol. 153, no. September 2020, p. 106017, 2021, doi: 10.1016/j.aap.2021.106017.

[12] Y. Rahmati, M. K. Hosseini, A. Talebpour, B. Swain, and C. Nelson, “Influence of Autonomous Vehicles on Car-Following Behavior of Human Drivers,” *Transp. Res. Rec.*, vol. 2673, no. 12, pp. 367–379, 2019, doi: 10.1177/0361198119862628.

[13] I. Mahdinia, A. Mohammadnazar, R. Arvin, and A. J. Khattak, “Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles,” *Accid. Anal. Prev.*, vol. 152, no. August 2020, p. 106006, 2021, doi: 10.1016/j.aap.2021.106006.

[14] Waymo LCC, “Waymo Open Dataset.” Accessed: Feb. 17, 2022. [Online]. Available: waymo.com/open/

[15] A. Sinha, A. Radwan, V. Dixit, “Interactions Between Human-Driven and Autonomous Vehicles on Public Roads,” *Res. Sq.*, pp. 0–15, 2021, 10.21203/rs.3.rs-849376/v1.

[16] X. Wen, Z. Cui, and S. Jian, “Characterizing car-following behaviors of human drivers when following automated vehicles using the real-world dataset,” *Accid. Anal. Prev.*, vol. 172, no. January, p. 106689, 2022, doi: 10.1016/j.aap.2022.106689.

[17] R. Wang and S. M. Lukic, “Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles,” *2011 IEEE Veh. Power Propuls. Conf. VPPC 2011*, pp. 1–7, 2011, doi: 10.1109/VPPC.2011.6043061.

[18] P. Sun, X. Wang, and M. Zhu, “Modeling Car-Following Behavior on Freeways Considering Driving Style,” *J. Transp. Eng. Part A Syst.*, vol. 147, no. 12, pp. 1–12, 2021, doi: 10.1061/jtepbs.0000584.

[19] M. Zhu, X. Wang, A. Tarko, and S. Fang, “Modeling car-following behavior on urban expressways in Shanghai: A naturalistic driving study,” *Transp. Res. Part C Emerg. Technol.*, vol. 93, no. March, pp. 425–445, 2018, doi: 10.1016/j.trc.2018.06.009.

[20] G. Li, Y. Jiao, V. L. Knoop, S. C. Calvert, and J. W. C. van Lint, “Large Car-following Data Based on Lyft level-5 Open Dataset: Following Autonomous Vehicles vs. Human-driven Vehicles,” pp. 5–10, 2023, [Online]. Available: http://arxiv.org/abs/2305.18921

[21] J. Houston *et al.*, “One Thousand and One Hours: Self-driving Motion Prediction Dataset,” no. CoRL 2020, pp. 1–10, 2020, [Online]. Available: http://arxiv.org/abs/2006.14480

[22] B. E. Hammit, A. Ghasemzadeh, R. M. James, M. M. Ahmed, and R. K. Young, “Evaluation of weather-related freeway car-following behavior using the SHRP2 naturalistic driving study database,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 59, pp. 244–259, 2018, doi: 10.1016/j.trf.2018.08.023.

[23] M. Kamrani, R. Arvin, and A. J. Khattak, “Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at Intersections,” *Transp. Res. Rec.*, vol. 2672, no. 38, pp. 290–301, 2018, doi: 10.1177/0361198118773869.

[24] A. Ben-Yaacov, M. Maltz, and D. Shinar, “Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance,” *Hum. Factors*, vol. 44, no. 2, pp. 335–342, 2002, doi: 10.1518/0018720024497925.

[25] J. C. Hayward, “Near-Miss Determination Through,” *Highw. Res. Board*, pp. 24–35, 1972, [Online]. Available: http://onlinepubs.trb.org/Onlinepubs/hrr/1972/384/384-004.pdf

[26] S. C. Johnson, “Hierarchical clustering schemes,” *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967, doi: 10.1007/BF02289588.

[27] C. O. Wu and J. Z. Huang, “Hierarchical Grouping to Optimize an Objective Function,” *J. Am. Stat. Assoc.*, vol. 98, no. 463, pp. 588–591, 2003, doi: 10.1198/016214503000000486.

[28] X. Chen, J. Sun, Z. Ma, J. Sun, and Z. Zheng, “Investigating the long- and short-term driving characteristics and incorporating them into car-following models,” *Transp. Res. Part C Emerg. Technol.*, vol. 117, no. November 2019, p. 102698, 2020, doi: 10.1016/j.trc.2020.102698.

[29] Z. Sun, X. Yao, Z. Qin, P. Zhang, and Z. Yang, “Modeling car-following heterogeneities by considering leader–follower compositions and driving style differences,” *Transp. Res. Rec.*, vol. 2675, no. 11, pp. 851–864, 2021, doi: 10.1177/03611981211020006.

[30] E. Kreyszig, H. Kreyszig, and E. J.Norminton, *Advanced Engineering Mathematics, 10th Edition*. 2019. doi: 10.1007/978-3-030-17068-4.

[31] B. Desgraupes, “Clustering Indices,” *CRAN Packag.*, no. April, pp. 1–10, 2013, [Online]. Available: https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf

[32] New York State Department of Motor Vehicles, “Chapter 8: Defensive Driving.” [Online]. Available: https://dmv.ny.gov/about-dmv/chapter-8-defensive-driving

[33] X. Wen, S. Jian and D. He, "Modeling the Effects of Autonomous Vehicles on Human Driver Car-Following Behaviors Using Inverse Reinforcement Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 12, pp. 13903-13915, 2023, doi: 10.1109/TITS.2023.3298150.