Extracting Microscopic Interactions in Road Users from Continuous Road User Trajectory

Yi-Shin Lin, *Member*, *IEEE*, Amir Hossein Kalantari, Aravinda Ramakrishnan Srinivasan, Jac Billington, Matteo Leonetti, and Gustav Markkula

***Abstract*— Understanding how microscopic interactions occur in traffic is essential to successfully transition our society from solely human drivers to a combination of robots and humans. However, a critical understanding of microscopic interactions is patchy. This paper presents a new algorithm that automatically zooms into microscopic interactions from the output of a customized artificial intelligence algorithm capturing traffic trajectories. The new algorithm enables researchers to conduct further investigations into microscopic interactions. We demonstrated one application of the new algorithm by investigating five traffic manifestations. We captured these manifestations and contrasted them between the interactive road users. We showed the different relationships in the interactive pair when a vehicle decelerated sharply, stopped too early away from the stop line, asserted its right of way, and when a pedestrian crossed roads prematurely and hesitantly. Finally, we discussed how one might incorporate the module into their work and considered the output of a different neural network model that estimates the time to arrival. (156 words; must be < 200).**

I. Introduction

To improve road safety and bridge a safe transition to a world of fully autonomous transport, we must be able to examine microscopic interactions automatically and intelligently. However, prior research in modelling focused primarily on meso- and macroscopic interactions [1] [2], relied on controlled and simulation data [3], and so far, primarily focused on one metric without also examining its interactive counterpart [4]. Although a recent surge in research extended to modeling micro-interactive behaviour in drivers [6], it has yet to incorporate an efficient method to systematically study the driver's and the pedestrian's micro-interactions in real traffic. We proposed a new decision-tree algorithm to extract the micro-interaction in traffic automatically.

* This work was supported by Engineering and Physical Sciences Research Council (EPSRC) under Grant number EP/S005056/1. For the

Yi-Shin Lin is with the Institute for Transport Studies, University of Leeds, England, UK (the first corresponding author to; e-mail: y.s.lin@leeds.ac.uk; yishinlin001@gmail.com).

Amir Hossein Kalantari is with the Institute for Transport Studies, University of Leeds, England, UK.

Aravinda Ramakrishnan Srinivasan is with the Institute for Transport Studies, University of Leeds, England, UK.

Jac Billington is with the School of Psychology, University of Leeds, England, UK.

Matteo Leonetti is with the Department of Informatics, King's College London, London, United Kingdom.

Gustav Markkula is with the Institute for Transport Studies, University of Leeds, England, UK. (3278 words)

One challenge in analyzing interactive nuances associated with road safety is to extract massive numbers of microscopic interactions automatically. The technique of extracting trajectory data has improved tremendously in the last decades due to the convolutional neural network (CNN) [7], object recognition [8], and object tracking [9]. The improvement results in software, such as YOLOv3 [8] and Detectron2 [10], capturing moving objects with high precision and increasingly fine differentiation among different entities. However, one drawback in these methods is that the data must be sent to a back office for post-processing to produce high-accuracy traffic trajectories.

Therefore, we partnered with a commercial company that collects and analyses traffic data for road safety applications. We deployed a 3D-vision sensor developed by our retail partner, Viscando, who uses customized artificial intelligence algorithms to analyze stereovision data to classify and track road users [11]. These sensors captured accurate trajectories of different road users over the ground, fused data among several cameras, and thus, covered an extensive range of areas and mitigated the occlusion problem. In addition, its ability to process data onboard resulted in real-time anonymous trajectories, enabling us to work in line with the Data Protection Act 2018 [12]. In this article, we built a new decision-tree module to read data from 3-D vision sensors and, from there, extracted the road user who might interact. The module separated the road users who might share intimate space in close time from those who did not. Therefore, we could use microscopic observations as secondary data for further analyses. We explored the key features and benefits of such an approach. We discussed one method to further integrate our new module with another NN model, dubbed *DeepTTE*, which estimated travel time to a destination [14] for bettering the customer experience, for instance, food deliveries. We converted the application of *DeepTTE* to calculate a road safety metric, the time-to-arrival to the crossing zone, using microscopic observations and exploring whether *DeepTTE* can also apply to study road safety.

Our module makes the automation of highlighting a pair of road users smoothly derived from the output of object tracking and, from there, extracts important nuances and safety metrics. We showcased our method by analyzing five exemplar microscopic manifestations that shared features with those examined before [13], [15], [16]. They are commonly observed in traffic and critical to conflict resolution in vehicles and between cars and pedestrians [17]. The first three were associated with a driver's intention: the sharp deceleration, the stopping short, and the priority assertion at the encounter, and the others were associated with a pedestrian's crossing decisions: the premature and hesitant crossings. The first two manifestations, considered from a car perspective, may inform a driver's intent to accommodate a pedestrian crossing. The third manifestation is associated with a driver's plan to communicate their right of way. The pedestrian perspective manifestations reflect how a pedestrian may assess the probability of crossing safely and react to the car's presentation.

To preview, our result demonstrated a new way of enhancing the value of object recognition and tracking output. Whether the output is generated from the established CNN method or provided by a commercial service, the new module enables one to examine macro-, meso-, and micro-interactions and therefore to study plausible causality in various traffic manifestations critical to road safety. Together with this paper, we released our code in the repository https://github.com/yxlin/Rose so that others can apply the module to their research.

II. Road User Interactions

A. The Gap Between Traffic Surveillance and Road Safety

How interactions at pedestrian crossings evolve is a critical area of research in transportation engineering. Accurately detecting and analyzing these interactions can provide valuable insights into traffic flow, safety, and efficiency. Although NN may accurately capture and recognize cars and pedestrians, it cannot explain how road users interact and how the interaction evolves [18]. Therefore, its usefulness could be better, with a new module filling the gap. There have been efforts on this front by building tools and knowledge to achieve explanations of the evolution of road user interactions. One action was to study the metrics that serve as surrogate measures predicting critical conflicts resulting from road user interaction [15]–[17], and another was a systematic catalogue and study of commonly observed microscopic traffic manifestations [13], [15]. We contributed to this effort by building a module to extract road user interactions and calculating the surrogate metrics from the interactive parties to study the traffic manifestation [12].

B. Driver Perspective Interactions

Sharp deceleration: When a driver is on a crossing path with a pedestrian determined to cross, the driver must react with a sharp drop in speed at a close distance to the encounter point. In the analysis of the vehicle-pedestrian conflict, this controlled deceleration was classified as a "severe evasive action" [17], potentially resulting in casualties. Here we quantified the risky interaction with the metric, sharp deceleration times (SDTs), which is the time from when a car began a drastic deceleration with a speed drop exceeding 3.5 m/s to the moment it entered the crosswalk. The number 3.5 was selected based on the assumption that it is slightly less than the length of a regular car (i.e., 4 meters). We also examined small changes in the threshold value and did not find it altered the included interactive instances.

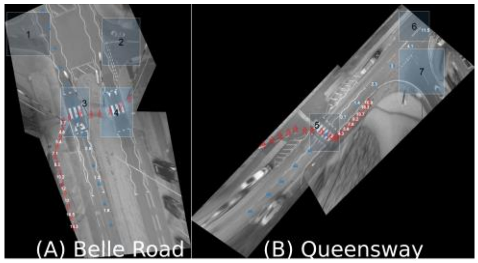

Figure 1. The two investigation sites. We examined seven crossing zones, highlighted in light blue. The sharp teeth and dashed (stop) lines are standard local road markings to signal drivers to slow down and yield to pedestrians at the crosswalk. The street map showed examples of interactive instances at each site. The white texts are the times of arrival to the encounter. Zoom in to see the fine details.

Stopping short: An interactive manifestation observed previously was that a car "stopped short" to yield to a pedestrian. *Stopping short* was defined as that, relative to the stop line, the car stopped at a distance that was longer than one car length. This inference was based on the California traffic regulation mandating that drivers give way to a pedestrian by the stop line [14]. However, a follow-up study also conducted in the USA found, in real traffic, a car rarely stopped completely but slowed to a minimum speed [16]. Therefore, instead of using a zero rate, we defined our threshold speed of *stopping short* as one m/s. Specifically, we examined the vehicle data from 10 to 4 meters from the stop line. When the car's speed dropped below one m/s, we selected these instances as the *stopping short* interactions. Thus, the vehicle manifestation may reflect a driver's active intent to yield, in contrast to the *sharp deceleration*, which may reflect a driver's reaction to a pedestrian's determination to cross.

We captured the *sharp deceleration* and the *stopping short* at Zone 3, 4, and 5 (Fig. 1) but not Zone 1, 2, 6, and 7. The differentiation was due to the observation that drivers showed different behaviors regarding whether they yielded and asserted their right of way at the zebra versus non-zebra crossings [21]–[23].

Priority assertion: In contrast to giving ways, a driver may assert their right of way by keeping a high speed [9] and sometimes even accelerating when driving over a pedestrian crossing. We captured the priority assertion in all seven zones.

C. Pedestrian Perspective Interactions

Premature and hesitant crossings: When standing at a crosswalk, pedestrians may decide prematurely to cross while it is considered risky. On this occasion, a driver must resolve such a situation to keep both parties safe. Two similar manifestations were dubbed "*yield acceptance hesitation"* and "*early yield* acceptance" in [13]. The former is closer to the idea of *premature crossings* here. However, the latter should be viewed as a pedestrian reacting to a driver's initiative to yield and referred to as a pedestrian stepping into the crosswalk before a car comes to a zero or minimum speed [13].

Here we examined the interactive manifestation of *premature crossings* from a pedestrian perspective instead because both *early yield acceptance* and *yield acceptance hesitation* implied a driver's intent to give way. A pedestrian might interpret their observation of the car as such, but *premature crossing* means pedestrians estimate a car's arrival time is sufficient and decide to cross. Undoubtedly, the three traffic manifestations must share features and bear a relationship; the naturalistic trajectory data cannot differentiate who initiated the interaction and who responded to the situation. We studied *premature crossings* because they are one of the likely interactive responses or consequences corresponding to the vehicle manifestations described previously.

In contrast to *premature crossings*, pedestrians may be overly cautious, even though a car has shown signs of yielding. We described this manifestation as a *hesitant crossing*. This manifestation implies that the pedestrian may interpret the driver's intent to yield not as such.

Except for the *sharp deceleration*, we used the surrogate safety metrics, the post-encroachment time (PET) [24], [25], to assess pedestrian safety in the selected traffic manifestation. Two joint criteria define PET and can be applied to most road users. Here, we described PET as a pedestrian safety metric and used it between vehicles and pedestrians. The first criterion was that a pedestrian must pass the prospective conflict zone first, and second, it measured the time when the pedestrian passed the conflict zone to when the car passed it. PET thus reflects the margin in time that the vehicle may endanger the pedestrian's safety. The shorter the PET, the riskier for pedestrians to cross a crossing.

III. METHODS

A. Naturalistic Data

To gauge enough micro-interactions, we dispatched two investigators to five candidate sites to conduct manual counts of the number of road users. This preliminary investigation was to assess whether a site might provide enough micro-interactions. The candidate sites were selected based on the historical traffic data provided by the local city council. Next, we identified the two areas, Belle Isle Road and Queensway, with sufficient traffic volumes. Traffic signals did not control the crossing zones, which included unmarked and marked crossings.

Our commercial partner, Viscando, and engineers contracted with the Leeds City Council deployed two measurement setups at each site. In addition to the supporting infrastructure, each setup comprised two Viscando OTUS3D sensors, which produced complete trajectories of road users using onboard, AI-based object detection, classification and tracking. It produced trajectory data of moving objects, which were projected onto a cartesian coordinate of the two-dimensional street map. Note that sensors provided the trajectories of all captured road users, interactive or not. **Figure 1** shows an example of an extracted interaction pair.

The sensors were set up on two street poles of heights of 6.46 m and 8.3 m, respectively, at Belle Isle Road and the Queensway. The former site recorded the traffic from 08:50:00 pm on the 9th of May to the 15th at 08:05:00 am. The latter recorded the traffic from 11:17:00 pm on the 17th of May to 04:05:00 am on the 23rd of May. The records on the 9th, 15th, and 23rd were to deploy and decommission the sensors. In total, we recorded 319.15 hours.

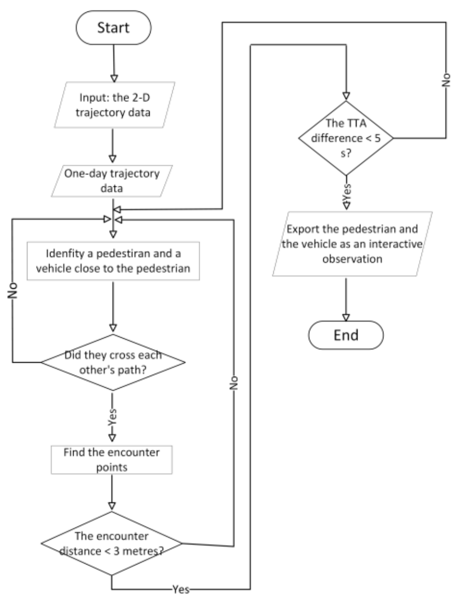

Figure 2. The step to extract interactive instances. TTA = time to arrival.

B. Modelling Methods

Viscando 3D: The sensors used embedded independent *imagers*, taking 10 to 20 images per second. The *imagers* extracted depth information using a proprietary high-performance stereovision algorithm and then separated backgrounds from the identified objects in the depth image. Next, they used a proprietary machine-learning algorithm to classify the identified objects based on their appearance, 3D features, and motion into different categories of road users, including pedestrians, 2-wheelers, and light and heavy vehicles. Then Viscando 3D sensors used an extended Kalman filter, modified with proprietary algorithms to fuse the data from different sensors and identify the road user's motion trajectories. The results are discrete trajectory points sampled at a frequency of 6.25 Hz. Our current analysis examines only the interactions between a pedestrian and a car.

The quantitative model: I took the model developed by Markkula and colleagues [13] and used a cross-validation-like method to test the model against the current naturalistic data. The model has not considered non-straight-line interactions, such as those in Zones 1, 2, 6,

Commented [YT1]: Again, a bit unclear to a reader from outside what kind of model it is.
Is it a model that predicts the future dynamics of the interaction based on initial observations? The one that identifies the type of interaction and its severity post-factum based on a complete trajectory?
Or something different?

Commented [YSL2R1]: This will be removed, so I will leave it.

and 7. Neither does the model prescribe how the model user determines a starting point in real traffic. The following describes my V subjective decisions in applying the model to the new naturalistic data.

Specifically, I defined the initial speed as the first non-zero speed in the camera-captured (not assessed) data. The initial distance was calculated from the x-y coordinate of the corresponding sample relative to the encounter points.

Next, I set up a folder structure for the Python script to work as in Markkula's GitHub source, found at https://doi.org/10.17605/OSF.IO/ZMK9T. Then, I entered the model, "oVAoBEvoAIoDAoSNvoPF" [13], which prescribed 193 optimal parameter sets. The model was set to simulate trajectories for up to 10 seconds. The data, however, are more organic, and few, if any, were presented with the exact 10-s trajectory. Next, I entered each empirical initial distance and speed into the model, which then generated 193 sets of simulations.

Thus, one empirical interactive pair generated 193 possible simulations. There were 2,304, 5,184, 1,920, 2,304, and 6,528 simulations for the five traffic manifestations. See Table I for the number of each traffic behavior. To make a straightforward comparison between the data and the model simulation, I averaged across all simulations over the 193 optimal sets and selected instances and then compared them against the Kalman-filtered average data. I used the Kalman filter here to derive smooth, theoretical speed and acceleration profiles because the trajectory data were sampled at 6.25 Hz, resulting in artificial zigzagging.

Nevertheless, there may be better or more correct ways to use the model than the subjective decisions of how to apply the model described here. For example, the 193 optimal sets might represent different predictions [13]. Moreover, whether the interactive instance captured by my algorithm matched, Markkula et al.'s definition of one interactive observation is disputed. It is argued even between this paper's first and last authors. I recommend the reader to [13] and its online Python code for in-depth details.

C. Interaction Extraction Procedures

The algorithm of micro-interaction extraction was a decision tree to identify a pair of interactive road users. First, it went over each day's data concurrently, ordering all road users based on the time sequence they were identified. Then, it looked for a pedestrian and recruited three other road users, capturing the closest, respectively, before and after the pedestrian.

Next, the algorithm examined each of the six candidates to see whether it was a vehicle and, secondly, the road user had a linear predictive path crossing with the pedestrian. Then, the algorithm looked for the two closest samples in space among the pair, defining them as the encounter points [9]. Due to discrete trajectory sampling, an interactive couple rarely shared an identical sampling point, even when they crossed each other's path in close time; thus, the encounter points were two trajectory samples with fewer than 3 meters.

In addition, the algorithm excluded a few cases by calculating the linear predictive path using the samples near the encounter points. For example, due to the narrow road width in the Queensway, a shopping district, a car, and a pedestrian sometimes traveled in parallel but were at a close distance. Next, the algorithm calculated the moment-to-moment time to arrival (TTA) relative to the encounter point. When the TTA difference between the pedestrian and the vehicle was fewer than five seconds, the couple was considered an interactive instance (Fig. 2).

We identified 535 interactive instances from a pool of 196,866 road users. These instances were then entered in the next module to capture the five traffic manifestations. Excluding the installation and decommission days, the average number of interactions was 53.5 per day, with a standard deviation 12.85.

D. Algorithm to capture the traffic manifestations

We examined the moment-to-moment vehicular data leading up to the stop line or the encounter points to identify specific traffic manifestations**.** The idea of using vehicular movement to assess the driver's intention was, perhaps first, proposed by Risto et al. [15] because the vehicular movement was noticeable, even just at a glance, using peripheral vision [26]. The respective inclusion criteria and the assessed vehicular and pedestrian data for each manifestation were described in **Error! Reference source not found.**. Here, we described the motivations and rationales behind the subjective decisions focusing on pedestrian perspective interactions.

Premature crossings: We used four criteria to capture *premature crossing*s. First, we examined the data where the speed profile in the vehicle had at least 35% speed decreases to gauge apparent vehicular showing. We assumed that the profile reflected the action of a driver stepping on the brake pedal. Second, we conducted a linear regression on the speed-time profile in the assessed data. This test captured an increasing trend in vehicle speed, supporting the assumption that it was still premature for a pedestrian to enter the crossing zone. Third, we included only the instances where the vehicle maintained its speed over 0.1 m/s in the assessed data. Fourth, the pedestrian must cross before the car.

Hesitant crossings: Conversely, the criteria for identifying *hesitant crossings* show a decreasing pedestrian and vehicular speed trend. Further, the vehicle must have 50% of the speed changes negative. These three criteria thus support the assumption that the driver intended to give way, and the pedestrian may hesitate to recognize a driver's intent.

Commented [YT3]: Should it be "we" as there are many authors in the article?

Commented [YSL4R3]: This will be removed, so I will leave it.

Commented [YT5]: Here it would be great to have some illustrations of the trajectories/scenarios to help reader understand the process and the outcomes

TABLE I. INCLUSION CRITERIA FOR TRAFFIC BEHAVIORS

	The observation	The vehicular data to assess	Specific criteria
Vehicle	Sharp deceleration	Ten meters to the stopping line.	1. The vehicle had at least one negative speed change over 3.5 m/s. 2. The pedestrian crossed the road in front of the vehicle.

	The observation	The vehicular data to assess	Specific criteria
	Stopping short	Ten to four meters from the stopping line.	1. The vehicle dropped to a speed of fewer than one m/s. 2. The pedestrian crossed the road in front of the vehicle.
	Priorirty assertation	Ten meters to the encounter points.	1. No data showed negative speed changes. 2. The vehicle passed the crossing zone earlier than the pedestrian.
Pedestrian	Premature crossing	Ten to four meters to the encounter points.	1. The vehicle's speed was on an increasing trend. 2. At least 35% of the data showed negative speed changes. 3. The vehicle must maintain its speed above 0.1 m/s. 4. The pedestrian crossed the road in front of the vehicle.
	Hesitant crossing	Ten to four meters to the encounter points.	1. Both the vehicle's and the pedestrian's speed was on a decreasing trend. (The assessed pedestrian data were four meters from the encounter) 2. At least 50% of the data showed negative speed changes. 3. The pedestrian crossed the road in front of the vehicle.

IV. RESULTS AND DISCUSSION

A. *Safety Metrics*

Table II reported the safety metric. Firstly, the short deceleration time had an average of 2.31 s and was accompanied by an average speed drop of 4.79 m/s, slightly higher than the threshold of 3.5 m/s in speed we used to capture the *sharp deceleration*. This speed change was likely in the range for a driver to safely resolve a potential collision, at least in the present dataset. The PET in the *stopping short* was longer than that of *sharp-deceleration time*, supporting our hypothesis that the former reflected an initiative and the latter a reaction from the driver. However, this hypothesis requires further data to consolidate. The PETs in the pedestrian manifestations aligned with the categorization of hesitant and premature crossings, with the former resulting in a slightly longer PET than the latter.

TABLE II. SAFETY METRICS

	N	Mean (s)	SD (s)	Range (s)
Sharp-deceleration time	12	2.31	1.1	0.96 to 4.96
Stopping short, PET	27	4	0.65	2.96 to 4.88
Priority assertation, PET	10	-2.64	0.85	-3.68 to -1.04
Hesitant crossings, PET	12	3.78	0.85	2.48 to 4.72
Premature crossings, PET	34	3.69	0.71	2 to 4.96

a. PET = post-encroachment time. N = number of instances. SD = standard deviation. Three cases were excluded in *stopping short* because the car passed earlier.

B. *Pedestrian Perspective Interactions*

Fig. 3 shows the speed profile of individual road users and the group average. The quantitative model captured the qualitative pattern of the average data at the *premature crossings* of the car and the pedestrian. Specifically, our inclusion criteria successfully recruited the pedestrians who stepped up speeds when they were walking towards the encounter. The requirements also successfully recruited the drivers, who decelerated to accommodate pedestrians' premature crossings. However, the quantitative model described pedestrians' speed step-up as a smooth and linear increase. Note that the qualitative pattern here shared similarity with the example simulation presented in the green line in the third column of Figure 1C [13], which explained another traffic manifestation, dubbed *yield acceptance hesitation*.

On the contrary, the empirical pedestrian data showed a dynamic pattern, with a drop at about 3 meters to the encounter, following a drastic rise of nearly one m/s and a near-constant speed when the pedestrian got to the encounter. The dynamic difference between the data and the model prediction was also seen in the vehicular speed profile at the *premature crossing*. The misfit in the fine details is likely due to the model being developed primarily concerning simulator data and concerned about a general pattern. It is also expected because of my subjective decision on what constitutes an initial position and speed. These factors are part of the wide range of influences in actual traffic data.

On the other hand, the model also described the qualitative pattern of the car's speed profile in *hesitant crossings*, where a similar fine-point mismatch was also observed. The model predicted a further speed drop slightly farther away from 5 meters from the encounter, but the data showed the driver increased speeds slightly but had not exceeded their early speed.

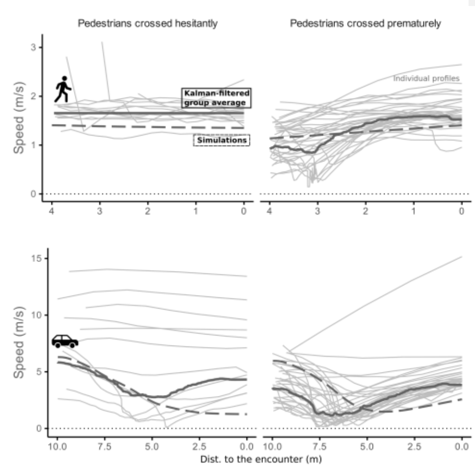

Figure 3. The speed profiles of the pedestrian (the upper row) and the car (the lower row) in the pedestrian traffic manifestations. The figure and the other

following zoomed in to the critical distance range covering 4 and 10 meters to the encounter.

In Fig. 4, the acceleration profile showed the model described the qualitative pattern of vehicular acceleration, although it mismatched the nadir of deceleration. The drivers dropped their speed farther from the encounter than that predicted by the quantitative model. Again, this difference could be attributed to averaging procedure because some individuals did show acceleration nadir close to the one indicated by the model. Moreover, the exact minimal acceleration could be affected by other uncontrolled factors in actual traffic data, such as road marking, traffic regulations, and individuals' changing preferences for the timing of initiating deceleration.

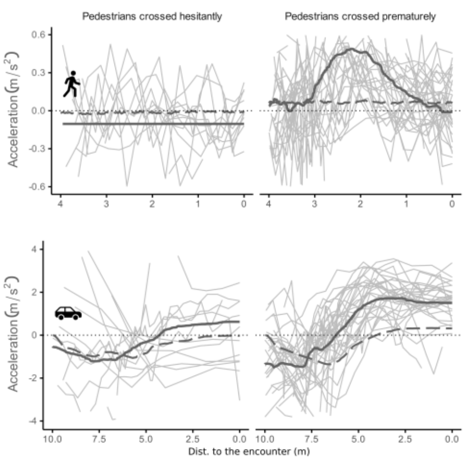

Figure 4. Acceleration profiles in the pedestrian traffic manifestations.

C. *Driver Perspective Interactions*

Sharp deceleration: The middle columns in Fig. 5 and 6 reported the speed and acceleration profiles. The model described the qualitative pattern of the speed profiles well, although it mismatched the timing of the sharp drop in the car profile. The model predicted a sharp decline in speed around 7.5 meters from the stop line, earlier than the observed data. Again, one can find an individual profile that matched the model prediction, as highlighted in red, so the mismatch may reflect the specific naturalistic data we observed, individual differences, and the traffic manifestation selection criteria.

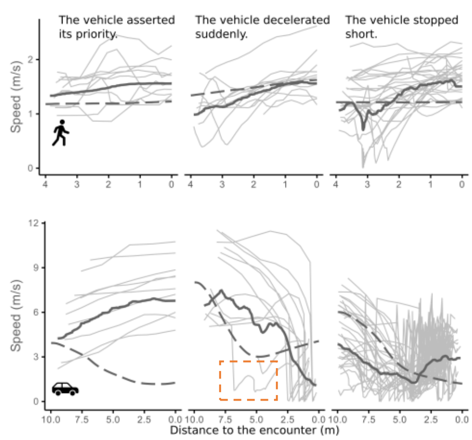

Figure 5. Speed profiles at the driver perspective behavior.

On the other hand, the pedestrian's speed increased linearly, which was also captured by the model prediction, although gentler than the data showed. Of course, one limitation in real-world data is that it cannot inform a causal link regarding whether the pedestrian reacted to the rapid vehicular deceleration or the driver braked hard because they realized the pedestrian showed no clear sign of slowing down. All in all, our data did indicate that such an interactive relationship exists.

Commented [YT6]: Is it really so? You can plot the deceleration profiles on the same time axis and see who started deceleration first.

The vehicular acceleration profile showed a consistent deceleration until the stop line, although the model predicted an earlier return to near-zero acceleration. One way to interpret the data manifestation is that the driver might notice the pedestrian showing no sign of slowing. This pedestrian manifestation showed up in the acceleration profile with an, albeit slight, increase in the magnitude of acceleration between 3 to 2 meters away from the encounter. Unfortunately, the model did not capture such pedestrian behavior.

Stopping short: The right column in Fig. 5 shows the *stopping short*. The speed profile in the car shared the example profile reported in Figure 1C, "Short-stopping" column in [13]. Their green and dark dashed lines shared the qualitative pattern of the solid and dashed lines here, indicating that the car gradually came to a minimal speed. The individual profiles showed further information about the zigzag pattern near the stop line. This pattern is contrasted with the sharp drop in individual speed profiles in the *sharp deceleration*.

The acceleration file in Fig. 6 showed that the car accelerated stronger near the stop line than in the sharp deceleration. One important thing to note is that pedestrians crossed the conflict zone earlier than the vehicle in both traffic manifestations.

Priority assertion: Lastly, the data aligned closely with the definition, showing a consistent acceleration and positively increasing trend at speed. The model could have described the data better, however. Such a model-and-data mismatch is likely due to data insufficiency (10 cases).

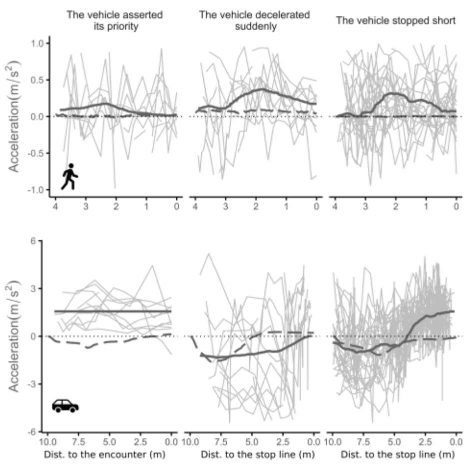

Figure 6. Acceleration profiles at the driver perspective interaction.

V. Limitation and Conclusion

The work to better understand micro-interactions involves examining the time evolution of a pedestrian and a car's manifestations and developing effective strategies to improve how we may harness the output of object tracking to explain these manifestations. The tool we reported here provides one possible solution. For example, the solution may apply to the large datasets of naturalistic trajectories of road users and filter out the secondary interactive data to assess other models that test road safety.

Two plausible future directions of harnessing the extraction module are to (1) incorporate the driver's, instead of the car's, data, such as using the in-car camera to capture a driver's facial expressions, eye viewing direction, and driving styles (e.g., the pattern of using steering wheels, frequencies of stepping on brake and acceleration pedals) and (2) re-purpose the NN model, such as *DeepTTE*, that estimates TTA to promote safe traffic interactions.

In addition to several limitations discussed in the previous section, an expansion of traffic data may help mitigate the rarity of microscopic interactions and reveal the actual ability of the quantitative model. Here the cross-validation test we conducted may partly reflect the problem of insufficient interaction observations.

References

[1] L. Chen, Z.-L. Guo, T. Wang, C.-Y. Li, and T.-Q. Tang, "An evacuation guidance model for heterogeneous populations in large-scale pedestrian facilities with multiple exits," *Phys. Stat. Mech. Its Appl.*, vol. 620, p. 128740, Jun. 2023, doi: 10.1016/j.physa.2023.128740.

[2] B. Anvari, M. G. H. Bell, A. Sivakumar, and W. Y. Ochieng, "Modelling shared space users via rule-based social force model," *Transp. Res. Part C Emerg. Technol.*, vol. 51, pp. 83–103, 2015, doi: https://doi.org/10.1016/j.trc.2014.10.012.

[3] K. Teknomo, Y. Takeyama, and H. Inamura, "Microscopic Pedestrian Simulation Model to Evaluate 'Lane-Like Segregation' of Pedestrian Crossing." arXiv, Sep. 06, 2016. doi: 10.48550/arXiv.1610.00749.

[4] Z. Islam, M. Abdel-Aty, A. Goswamy, A. Abdelraouf, and O. Zheng, "Modelling the Relationship Between Post Encroachment Time and Signal Timings Using UAV Video data." arXiv, Oct. 10, 2022. doi: 10.48550/arXiv.2210.05044.

[5] J. Zhao, V. L. Knoop, and M. Wang, "Microscopic Traffic Modeling Inside Intersections: Interactions Between Drivers," *Transp. Sci.*, vol. 57, no. 1, pp. 135–155, Jan. 2023, doi: 10.1287/trsc.2022.1163.

[6] O. Siebinga, A. Zgonnikov, and D. Abbink, "A Human Factors Approach to Validating Driver Models for Interaction-Aware Automated Vehicles," *J Hum-Robot Interact*, May 2022, doi: 10.1145/3538705.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015. Accessed: Apr. 20, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

[8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *ArXiv Prepr. ArXiv180402767*, 2018.

[9] X. Li, K. Wang, W. Wang, and Y. Li, "A multiple object tracking method using Kalman filter," in *The 2010 IEEE International Conference on Information and Automation*, Jun. 2010, pp. 1862–1866. doi: 10.1109/ICINFA.2010.5512258.

[10] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2." 2019. [Online]. Available: https://github.com/facebookresearch/detectron2

[11] A. Singh, U. Erlandsson, Y. Tarakanov, and K. Bergvall, "Viscando 3D and AI-sensing," *Viscando*, 2023. https://viscando.com/ (accessed May 06, 2023).

[12] *Data Protection Act 2018*. King's Printer of Acts of Parliament. Accessed: May 06, 2023. [Online]. Available: https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

[13] G. Markkula *et al.*, "Explaining human interactions on the road requires large-scale integration of psychological theory," 2022.

[14] D. Wang, J. Zhang, W. Cao, J. Li, and Y. Zheng, "When will you arrive? estimating travel time based on deep neural networks," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, in AAAI'18/IAAI'18/EAAI'18. New Orleans, Louisiana, USA: AAAI Press, Feb. 2018, pp. 2500–2507.

[15] M. Risto, C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan, “Human-Vehicle Interfaces: The Power of Vehicle Movement Gestures in Human Road User Coordination,” *Driv. Assess. Conf.*, vol. 9, no. 2017, Art. no. 2017, Jun. 2017, doi: 10.17077/drivingassessment.1633.
[16] J. Domeyer, A. Dinparastdjadid, J. D. Lee, G. Douglas, A. Alsaid, and M. Price, “Proxemics and Kinesics in Automated Vehicle–Pedestrian Communication: Representing Ethnographic Observations,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2673, no. 10, pp. 70–81, Oct. 2019, doi: 10.1177/0361198119848413.
[17] I. Kaparias *et al.*, “Development and Implementation of a Vehicle–Pedestrian Conflict Analysis Method: Adaptation of a Vehicle–Vehicle Technique,” *Transp. Res. Rec.*, vol. 2198, no. 1, pp. 75–82, Jan. 2010, doi: 10.3141/2198-09.
[18] R. Goebel *et al.*, “Explainable AI: The New 42?,” in *Machine Learning and Knowledge Extraction*, A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 295–303. doi: 10.1007/978-3-319-99740-7_21.
[19] L. Zheng, T. Sayed, and M. Essa, “Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators,” *Accid. Anal. Prev.*, vol. 123, pp. 314–323, Feb. 2019, doi: 10.1016/j.aap.2018.12.007.
[20] D.-G. Kim, Y. Lee, S. Washington, and K. Choi, “Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models,” *Accid. Anal. Prev.*, vol. 39, no. 1, pp. 125–134, Jan. 2007, doi: 10.1016/j.aap.2006.06.011.
[21] A. Várhelyi, “Drivers’ speed behaviour at a zebra crossing: a case study,” *Accid. Anal. Prev.*, vol. 30, no. 6, pp. 731–743, 1998, doi: 10.1016/S0001-4575(98)00026-8.
[22] F. Bella and F. Nobili, “Driver-pedestrian interaction under legal and illegal pedestrian crossings,” *Transp. Res. Procedia*, vol. 45, pp. 451–458, Jan. 2020, doi: 10.1016/j.trpro.2020.03.038.
[23] F. Bella and M. Silvestri, “Effects of safety measures on driver’s speed behavior at pedestrian crossings,” *Accid. Anal. Prev.*, vol. 83, pp. 111–124, Oct. 2015, doi: 10.1016/j.aap.2015.07.016.
[24] P. J. Cooper, “Experience with Traffic Conflicts in Canada with Emphasis on ‘Post Encroachment Time’ Techniques,” in *International Calibration Study of Traffic Conflict Techniques*, E. Asmussen, Ed., in NATO ASI Series. Berlin, Heidelberg: Springer, 1984, pp. 75–96. doi: 10.1007/978-3-642-82109-7_8.
[25] B. L. Allen, B. T. Shin, and P. J. Cooper, “Analysis of Traffic Conflicts and Collisions,” *Transp. Res. Rec.*, no. 667, Art. no. HS-025 846, 1978.
[26] T. Matthews, “Designing and evaluating glanceable peripheral displays,” in *Proceedings of the 6th conference on Designing Interactive systems*, University Park PA USA: ACM, Jun. 2006, pp. 343–345. doi: 10.1145/1142405.1142457.