Autonomous vehicles for on-demand transport service: Field experimentation

S. M. Hassan Mahdavi and Neila Bhouri

Abstract—**Autonomous vehicles (AVs) hold great promise for transforming urban mobility, offering potential benefits. In this paper, we provide an evaluation of AV experimentation for six vehicles used in an on-demand service. The assessment considers both user experience and system effectiveness, analyzing service accessibility, responsiveness, and ride quality. From the user perspective, key factors include travel time, waiting time, and in-vehicle experience, while operational efficiency is assessed through traveled distance, vehicle availability, and driving dynamics such as speed and acceleration.**

Keywords—Autonomous vehicles, AV, shuttle, On-demand service, assessment, evaluation, field operational tests.

I. Introduction

Autonomous vehicles (AVs) hold great promise for transforming urban mobility, offering potential benefits such as enhanced road capacities, reduced congestion, and improved speed consistencies [1], [2]. While their operation remains experimental, it is widely expected that these vehicles will progressively integrate into urban networks; yet with some degree of uncertainty. Therefore, to minimize uncertainties, realizing their successful deployment hinges on understanding the importance of related performance aspects and the impacts on operational reliability.

In recent years, manufacturers, and authorities have invested in launching several fields operational tests (FOTs) piloting SAE L3/L4 AV models [3] based on all conceivable overlapping conditions, use cases, restrictions, and scenarios that vehicles might encounter. While notable FOTs efforts have emerged from Europe (e.g., "SHOW", "L3pilot") the USA (e.g., "Relay"), and China (e.g., "Apollo at Wuhan"), there remains a scarcity of appropriate evaluation approaches to assess the operational reliability of tested AVs.

Autonomous vehicle evaluations are often conducted through simulations [6][7] and [8], whereas our study assesses a real-world experiment. Launched in France [4], it aims to formulate a safety validation methodology to amplify comprehension of AV applications. To understand the implications on operational reliability, and in-depth exploration of FOTs (i.e., autonomous shuttle bus, on-demand robotaxis, and private autonomous vehicles).

This paper investigates the experimentation of an AVs use as an on-demand service in an urban area in the Paris region. We leverage data from the project "SAM" (Security and

The paper investigates the on-demand service of AVs on mixed urban road traffic. It investigates, the service from the users' perspective such as the travel time, the waiting time, and in-vehicle time, with a particular emphasis on their operational travel time.

The paper investigates the performance of an on-demand AVs service in a mixed urban traffic. The evaluation considers both user experience and system effectiveness, analyzing service accessibility, responsiveness, and ride quality. From the user perspective, key factors include in-vehicle travel time and waiting time. While speed and acceleration reliability consistency provide indications of the comfort experienced by users inside the vehicles. Operational efficiency is assessed through traveled distance, vehicle availability, and driving dynamics such as speed and acceleration.

The paper is divided into five sections. After the introduction, Section 2 presents the experimentation description, including the experimental site, its environment, and the collected data. Section 3 explains the methodology adopted for the evaluation. Section 4, dedicated to the evaluation results, is divided into three paragraphs. The first discusses the passenger flow analysis, focusing on supply and demand. The second analyzes traveled distances. Finally, the third presents results on speed and acceleration, which directly impact passenger comfort. The conclusion is provided in Section 5.

II. Experimentations

One shuttle with a capacity of 12 passengers and five OALVs with a capacity of 3 passengers each were used. These AVs belonged to two different vehicle manufacturers and were at Level 4 automation. The AVs responded to client demands, either booked through the website or hailed at one of the stops. Clients were driven from a pick-up stop to a drop-off stop at any time wanted.

The site is made up of straight lines, intersections, traffic circles, and crosswalks. The maximum speed of vehicles depends on the nature of the road and varies between 30 and 110 km/h. The automated on-demand cars ran from April 4, 2022, to June 24, 2022, and the shuttle service from April 27 to June 24, 2022.

S. M. Hassan Mahdavi, Human Factors & Economics of Sustainable Mobilit, Institut VEDECOM, 23bis, allée des Marronniers, 78000 Versailles, France; email : hassan.mahdavi@vedecom.fr ORCID ID: 0000-0001-5087-1531.

Neila Bhouri, Laboratory Cosys/Grettia,, University Gustave Eiffel, 14-20 Boulevard Newton, Champs sur Marne, 77447 Marne-la-Vallée, France.. email :neila.bhouri@univ-eiffel.fr ORCID-ID :0000-0003-0775-7922

The shuttle experimentation site is slightly different from the OALVs site. However, for confidentiality reasons, we are unable to disclose the technology of the AVs used or the precise map of the experimental site.

Two kinds of data are gathered for this experiment:

a) Vehicle data: the coordinates of the vehicle position with the timeline, and videos taken from the vehicle at the according to time.

b) Passenger data: client ID, vehicleId, order dates, pick-up/drop-off stops name.

Notice that, in this study, for OALVs. We have only the ID of the main passenger. We assume therefore that there is only one passenger per car.

III. Methodology

We focus in this paper on the performance of AVs as an on-demand transport service. We evaluate key metrics related to both passenger flow and vehicle operations, with calculations made across the entire dataset. The vehicle operations are analyzed using speed and acceleration metrics, mainly:

- Speed Variability: the dispersion of speed values (in km/h) between different experiments and vehicles. This performance indicator allows us to compare speed data over and above average values, on different days of the week. For passengers, this dispersion can mean a less comfortable journey. For other road users (i.e. other vehicles), it can be difficult to predict the vehicle's actions, which can compromise road safety.
- Speed Consistency Index: the index calculates the proportion of time a vehicle maintains a constant speed (km/h) over a given period, which indicates the stability of travel over that period. It therefore represents the percentage of time during which the vehicle's speed remains within a certain threshold of the average speed. A value of 90% of the average was chosen when it was coherent for the acceleration and the speed.
- Acceleration Consistency Index: this indicator represents the regularity and reliability of the acceleration and deceleration behavior (m/s^2) of automated vehicles over time, reflecting their adaptability to variations in the real world.

Passengers' operations are analyzed, by time intervals, through metrics such as:

- Passenger flow distribution is analyzed to identify patterns for the different days of the week, the periods of the day, and pick-up and drop-off locations. Passenger flow distribution is a measure of how passengers are distributed across a service, and therefore within the experiment. This indicator helps operators to better manage and plan vehicle occupancy, thereby improving the overall user experience of the transport systems provided.
- Passenger in-vehicle travel: this indicator assesses the efficiency and quality of the service for the user. Expressed in hours per passenger, it represents the average time a passenger spends inside the vehicle during the journey from origin to destination. A low value indicates a faster service with fewer delays, while a high value suggests longer journey times, due to factors such as vehicle inefficiency, congestion, or frequent stops.
- The passenger in-vehicle travel time distribution: refers to the distribution of the time that passengers spend inside the AV during their journeys. It quantifies the extent and variability of journey times experienced by passengers. This distribution provides insight into the reliability and efficiency of a system.
- The number of passengers served per day: this indicator represents the total number of passengers using the service during the day. Aggregating data across the fleet allows us to measure the overall service capacity.
- The total number of individual journeys or services successfully completed by a particular autonomous vehicle or service in a given time. This measure provides an overview of operational efficiency and utilization rate.
- Waiting time: defined as the delay between ordering and vehicle arrival. Waiting time directly impacts user satisfaction and service quality. It indicates the efficiency of the assignment algorithm. Lower waiting times make the system more competitive against other modes of transport, influencing user choice and overall system acceptance

IV. Results

First of all, we compared the behavior and percentage of use of all the vehicles to ensure they were used comparably, otherwise, we need to differentiate the study according to the vehicle tested. As the sites are slightly different, we differentiate the analysis, on the one hand, the 5 OALVs and, on the other hand, the results for the shuttle.

A. General Overview of Vehicle Usage

As for the shuttle (Table 1), it was used for a total of 26 hours, carrying 342 passengers which represents 4.6 minutes per passenger. 93% of these trips were unreserved boarding and 7% were pre-orders. The distribution of these passengers shows a disparity among the days of the week. Tuesday was the most crowded day.

Table 1: The shuttle's use: Passengers served, Vehicle travel time with corresponding Speed and Acceleration Styles.

<table>
<tr><th>Vehicle 1 service</th><th>Total passengers served</th><th>Total passengers in-vehicle travel time (hours)</th><th>Experienced Average speed</th><th>Average acceleration and deceleration experienced</th></tr>
<tr><td>Monday</td><td>36</td><td>3.3</td><td rowspan="6">36 (km/h)</td><td rowspan="6">0.55
-0.18</td></tr>
<tr><td>Tuesday</td><td>101</td><td>9.0</td></tr>
<tr><td>Wednesday</td><td>55</td><td>4.0</td></tr>
<tr><td>Thursday</td><td>84</td><td>6.1</td></tr>
<tr><td>Friday</td><td>66</td><td>4.6</td></tr>
<tr><td>Total</td><td>342</td><td>26.9</td></tr>
</table>

Regarding the OALVs (Table 2), a total of 1,133 passengers were served for a complete journey with a rate of 91% for immediate boarding and 9% for pre-ordering. It covers 144 hours of total travel time for the 5 vehicles. The number of users ranges from 194 passengers (the minimum, vehicle 3) to 262 passengers (the maximum, vehicle 5). The number of users per day does not indicate a preference for one vehicle over another: the highest number of passengers transported was achieved by V2 on Tuesday, while the lowest was by V4. However, on Monday and Wednesday, V4 carried more passengers than V2.

Table 2: The five OALVs use: Passengers served, Vehicle travel time with corresponding Speed and Acceleration Styles.

	Total passengers served					Total time in vehicle (hours)					Average speed (km/h)					Average acceleration/deceleration				
	V1	V2	V3	V4	V5	V1	V2	V3	V4	V5	V1	V2	V3	V4	V5	V1	V2	V3	V4	V5
Monday	51	39	28	42	47	4.7	5.0	3.1	4.9	4.6										
Tuesday	52	72	54	32	59	6.8	8.6	6.1	4.9	7.8						0.65	0.73	0.61	0.58	0.67
Wednesday	27	48	42	50	68	4.0	6.1	6.1	8.5	9.3	14.75	15.3	13.5	12.7	15.1					
Thursday	49	56	35	38	44	6.1	8.3	4.5	5.3	6.1						-0,48	-0,75	-0,57	-0,47	-0,66
Friday	45	41	35	35	44	5.2	5.1	3.9	4.1	5.2										
Total	224	256	194	197	262	26.9	33.0	23.6	27.7	33.1										
Average	226,6					28,86					14,27					0,648 / -0,59				

B. *Passenger flow analysis*

As shown in Figure 1, OALV vehicles were available in 62.4% of the time. This rate includes the completed journeys (56.8%) and the "no-show" cases when the vehicle was available but the customer did not show up to the pick-up stop (5.6%).

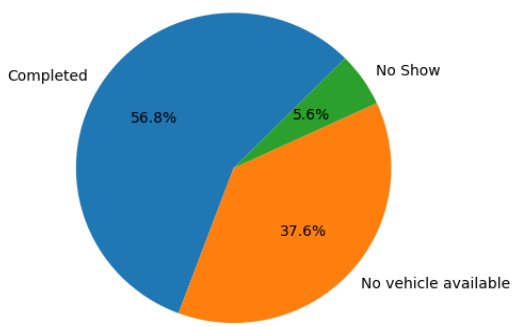

Figure 1: Total availability rate for all vehicles service

In Figure 2 we draw the passenger demand and the AVs availabilities for each day of the experimentation. We can notice, greater disparities in the first days: the demand was far more important than the service availability. This could be explained by the curiosity of users for these new cars. Indeed, as shown by Everett Rogers' Diffusion of Innovations Theory, people are attracted to new technologies; see also [9], however, they were dissuaded in this experiment by the lack of availability. Rapidly, the demand was adapted to the offer (nearly in the two weeks of the experimentation).

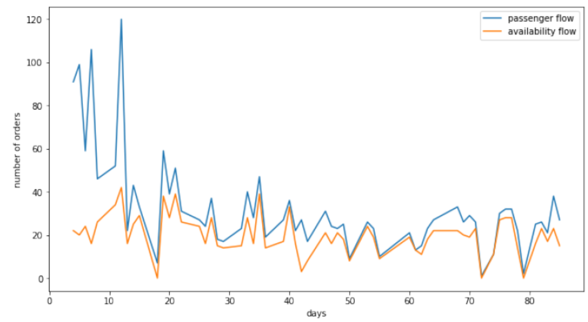

Figure 2: Passenger demand and availability flow offer

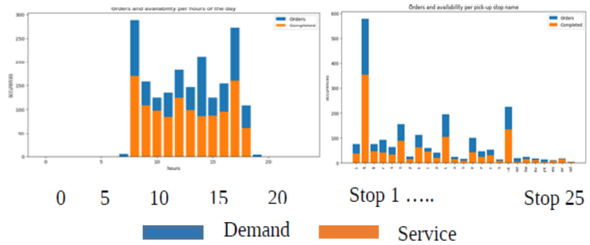

Fig.3.a per hour of the day

Fig.3.b per pick-up stops

Figure 3: distribution of the demand vs the completed service per hour of the day (3.a) and pick-up stops (3.b)

Though the service is available for the 24 hours of the day, orders are concentrated between 7 am and 7 pm, peaking at 8 am and 6 pm, suggesting demand from workers and students, with lunch breaks at 12 pm and 2 pm (Figure 3.a). Pick-up locations vary, with Stop 2 being the most used, likely due to its rail station proximity (Figure 3.b). Vehicles operate from 8 am to 6 pm, meeting only 60% of peak demand, with unexpected availability drops at certain times. Specific dates show lower availability, linked to holidays or weekends

C. *Traveled distance and In-vehicle travel time*

The In-vehicle travel time ranges from 5.5 minutes (Vehicle 1, Monday) to 10.2 minutes (Vehicle 4, on Wednesday) showing short trips. There are disparities between the different vehicles, with vehicle 4 having the highest in-vehicle time (Figure 4). This is due to more conscious driving with a lower speed (12.7 km/h) when the average speed of all vehicles is 14.4km/h (see Table 2).

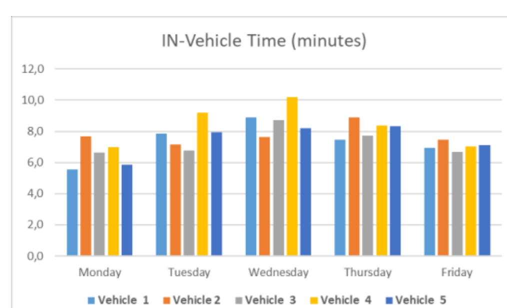

Figure 4: Passengers In-vehicle time per vehicle and day of the week

We also notice differences between the days of the week, Monday is the lowest with a 6.5-minute average. This may be due to traffic congestion and travelers' behaviors. Analysis of the users' behavior showed that 28% of all Monday trips originated from Stop 1, and 14% of those trips took 3.5 minutes. Similarly, 30.2% of trips started from the rail station, and 15% took only 6 minutes to reach their destination (Figure 5).

In contrast, Wednesday's operations revealed a different picture. With 27 passengers served over 4 hours, the average travel time was estimated at 8.9 minutes per passenger. This suggests that on Wednesday passengers traveled to more distant destinations. For example, 29% of all trips that day started from STOP 2 with an average travel time of 11.5 minutes. The trends for the remaining days are as follows: on Tuesday, 52 passengers were served in 6.8 hours, averaging 7.8 minutes per passenger, indicating potentially longer trips or traffic issues. On Thursday, 49 passengers were transported in 6.1 hours, with an average of 7.4 minutes each (see Table 2).

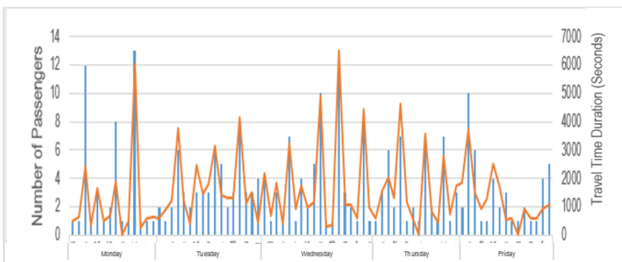

Figure 5: Number of passengers per pick-up station per weekday with corresponding travel time for vehicle 1 as an example,

With an average travel distance of approximately 2 km (Figure 6), there is a clear trend of higher vehicle occupancy on certain days, accompanied by shorter journey times, indicating that more passengers are willing to travel shorter or medium distances. This observation is further confirmed when we plot the number of passengers per vehicle against travel time, as shown in Figure 5 for Vehicle 1

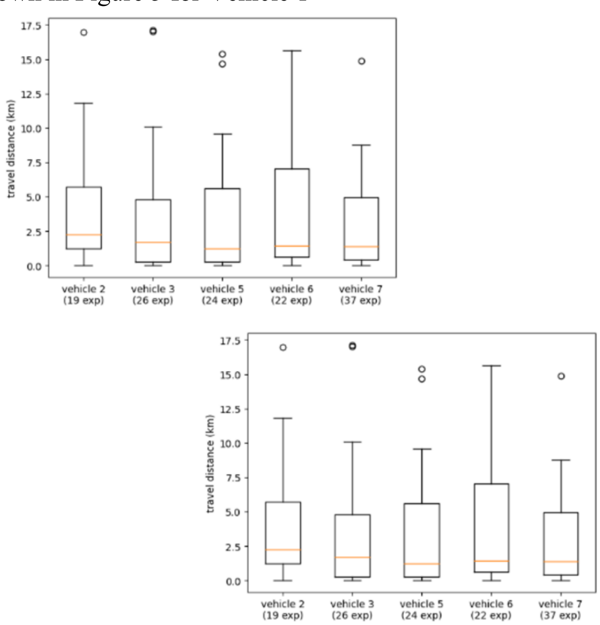

Figure 6: Distributions of travel times (right) and travelled distances (left) per vehicle.

The average travel time on board for all Mondays was approximately 5.5 minutes per passenger. Tuesday's service was the most efficient in terms of serving more passengers with an average travel time of 5.35 minutes per passenger. This suggests that, on this day, the vehicle maintained a consistent service pattern in terms of time taken per passenger. The median speed on Wednesday was 37.2km/h, which is slightly higher than Tuesday's median. On Wednesday, the vehicle reached its maximum speed of 87.9 km/h at the 95th percentile. If we assume that the number of passengers is directly related to the number of pick-up points, then Tuesday had the most pick-up points, while Monday had the fewest. The average travel time in a vehicle for all Fridays was approximately 4.18 minutes per passenger.

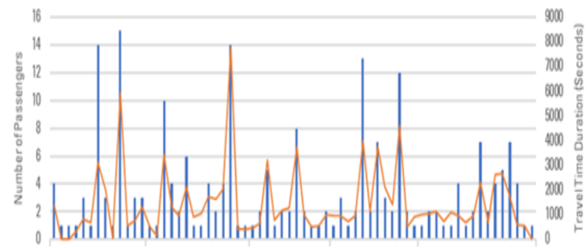

Figure 7: Number of passengers per pick-up station per weekday with corresponding travel time of the shuttle.

D. *Speed, acceleration, and passenger comfort*

Each vehicle has unique acceleration/deceleration patterns based on daily traffic, demand, and pick-up/drop-off points. Vehicles with aggressive acceleration/deceleration maintain a moderate speed with speed spikes, while those with cautious driving have a smoother speed and less acceleration increase.

The operation of five vehicles showed different levels of speed and acceleration regularity. Speed consistency (Figure 8) showed an average of 94.7% for Vehicle 1, 90% for Vehicle 2, 93% for Vehicle 3, 90.2% for Vehicle 4, and 92.5% for Vehicle 5. (Figure 9) shows that the overall average speed consistency level is 92.1% and the overall average acceleration regularity index is 81.7% (Figure 9).

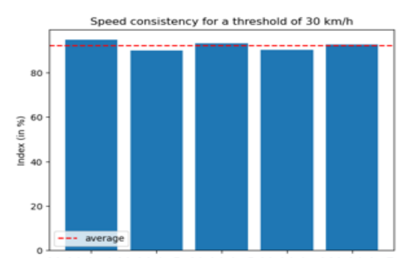

Figure 8: Speed Consistency

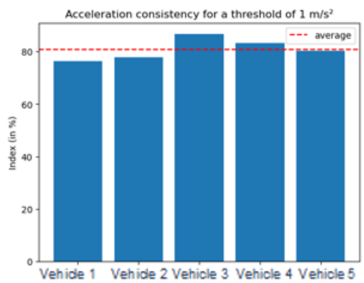

Figure 9: Acceleration consistency

The analysis of vehicle acceleration and deceleration per day shows that each vehicle represented distinct patterns according to daily traffic conditions. As an example, we give in Figure 10, the acceleration and deceleration of Vehicles 1 and 2. We can notice that Vehicle 1, showed strong acceleration, particularly on Tuesdays and Fridays, reaching 1.49 and 1.41 m/s^2 respectively, suggesting its ability to adapt quickly to traffic flow. Conversely, the strong deceleration of vehicle 2 on Mondays highlights its quick reflexes, likely to react to sudden obstacles or unexpected traffic jams, and probably a cautious approach to roundabouts to ensure passenger comfort and safety.

a. Vehicle 1 b. Vehicle 2

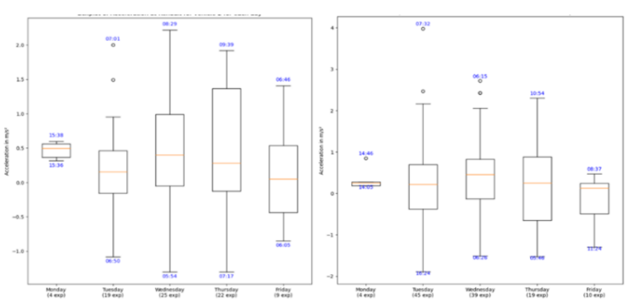

Figure 10: Acceleration and deceleration by day of the week for Vehicle 1 and Vehicle 1

Drawing the "Time required to reach maximum acceleration/ minimum deceleration" (Figure 11) shows that Vehicle 1 reached a maximum acceleration of 1.08 m/s^2, with an average acceleration of 0.65 m/s^2 and an average speed of 14.76 km/h, the vehicle tended to maintain a moderate speed with occasional increases. On the deceleration side, the recorded minimum deceleration was -0.53 m/s^2, while the average deceleration was -0.48 m/s^2.

Vehicle 2 exhibited a maximum acceleration of 1.17 m/s^2, making it one of the most aggressive vehicles in terms of acceleration. Its average acceleration was 0.73 m/s^2, which, combined with an average speed of 15.3 km/h, suggests that this vehicle tended to be faster with periodic rapid accelerations. The deceleration patterns further emphasize the vehicle's aggressive nature. It recorded a minimum deceleration of -0.83 m/s^2 and an average deceleration of -0.75 m/s^2, indicating stronger braking behavior.

In summary, the driving style of Vehicle 2 leans more toward aggressiveness, both in acceleration and deceleration, while maintaining a slightly higher average speed than Vehicle 2.

Figure 11: Min/max acceleration and deceleration: vehicle 1 and vehicle 2

With regard to the shuttle, the level of regularity was different from that of the OALVs. The average regularity of acceleration was estimated at 81.7% compared with 80.8% for the OALVs (Figure 12). On Mondays, the negative median acceleration (-0.01 m/s^2) suggests that the vehicle tends to decelerate slightly more than it accelerates. This may be explained by the stable traffic conditions without many obstacles affecting speed confirmed by the videos taken from the vehicles. On Wednesday, the median acceleration of 0.02 m/s^2, slightly positive, suggests a slightly more frequent acceleration behavior compared to deceleration, but it is minimal. Thursday, however, stands out with the highest median acceleration of 0.04 m/s^2. This indicates that on Thursdays, the vehicle's tendency to accelerate is greater than its tendency to decelerate (see Figure 13).

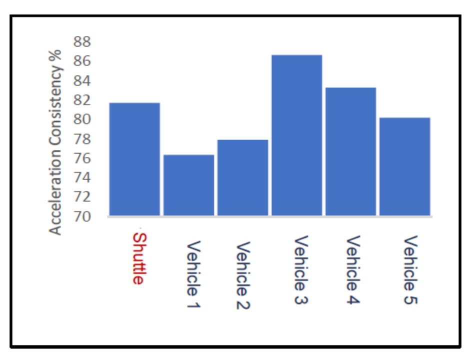

Figure 12: Acceleration consistency for the shuttle and the 5 OALVs.

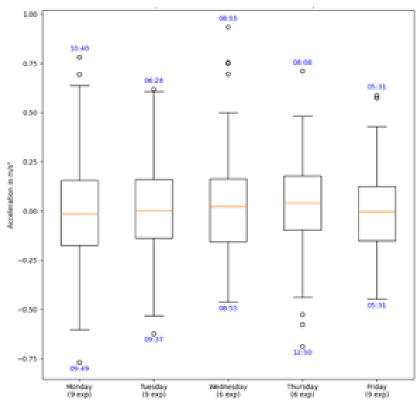

Figure 13: The shuttle acceleration and deceleration by day of the week

Days with negative median acceleration (such as Monday) may involve more frequent stops or slower traffic conditions. This may indicate a need for route optimization or fleet management adjustments on these days to ensure fast service. Days with zero or near-zero median acceleration (such as Tuesday and Friday) suggest a balance in vehicle movement. These days are ideal because they indicate a steady flow of traffic without frequent stops or speed changes. Finally, days with a positive median acceleration (such as Thursday) could indicate smoother traffic conditions, but if this value is too high, it could also indicate more aggressive driving behavior that may not always be comfortable for passengers.

V. Conclusion

This paper evaluates the performance of a field operational test involving an autonomous vehicle (AV) shuttle and five on-demand autonomous low-speed vehicles (OALVs) to determine their effectiveness for users and operators.

Vehicles have demonstrated their ability to adapt to a variety of scenarios in mixed traffic conditions with conventional traffic. It showed its adaptability for the on-demand service respecting the users' demand from specific pick-up to specific drop-off points. The distinct acceleration and deceleration patterns of the five vehicles indicate that their algorithms could be optimized based on the obstacles, priorities, and situations they encounter: for example, prioritizing adaptation to changes in traffic flow, or cautious driving detecting obstacles and slowing down at each. There is a clear trend towards higher vehicle occupancy on certain days, with shorter journey times in the vehicle, suggesting that more passengers are willing to travel short or medium distances.

With a larger capacity of 12 seats, the shuttle offered a different dynamic in terms of passenger service. While the OALVs demonstrated a range of behaviors from rapid adaptation to cautious behavior, the shuttle revealed a more harmonious interplay of acceleration and deceleration, perhaps indicating its design for the shuttle and consistent journeys with more passengers. This balance, combined with its speed data, suggests a driving profile that prioritizes passenger comfort and safety (taking into account Level 4 autonomy). Given its greater passenger capacity, the acceleration and deceleration of the shuttle could be inherently smoother, resulting in a different type of adaptability and responsiveness compared with smaller vehicles. The shuttle is busiest during the evening peak hours (4 pm - 7 pm).

In conclusion, measuring key performance indicators using real data, supplemented by passenger feedback, will facilitate periodic updates to the vehicles' algorithms. This will optimize the balance between speed, efficiency, and passenger comfort. Additionally, the diverse driving and acceleration behaviors observed suggest that the learning and calibration of control algorithms could be tailored to specific use cases and user preferences.

Acknowledgment

We thank the SAM project for allowing this study. We greatly appreciate the support and collaboration of all consortium members.

References

[1] Ma Ke Hao Wang and Tiancheng Ruan. 2021. “Analysis of Road Capacity and Pollutant Emissions: Impacts of Connected and Automated Vehicle Platoons on Traffic Flow.” Physica a: Statistical Mechanics and Its Applications.

[2] Sun Wenbo Shian Wang Yunli Shao Zongxuan Sun and Michael W Levin. 2022. “Energy and Mobility Impacts of Connected Autonomous Vehicles with Co-Optimization of Speed and Powertrain on Mixed Vehicle Platoons.” Transportation Research Part C.

[3] SAE-International, 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

[4] SAM, 2019. Security acceptability for automated driving. URL: https://www.vedecom.fr/sam-project-security-acceptability-for-automated-driving/?lang=en.

[5] Bhouri N and J Kauppila. 2011. “Managing Highways for Better Reliability: Assessing Reliability Benefits of Ramp Metering.” Transportation Research Record No. 2229 (2011): 1-7: 1–7.

[6] Simon Oh, R. Seshadri, C. L. Azevedo, N. Kumar, K. Basak, M. Ben-Akiva.2020. “*Assessing the impacts of automated mobility-on-demand through agent-based simulation: A study of Singapore*”. Transportation Research Part A: Policy and Practice. Volume 138. Pages 367-388. https://doi.org/10.1016/j.tra.2020.06.004.

[7] Johnson, C., Graupe, E., and Kassel, M., "*A Literature Review of Simulation Fidelity for Autonomous-Vehicle Research and Development*," SAE International Journal of Aerospace. 16(3):253-261, 2023, https://doi.org/10.4271/01-16-03-0021.

[8] Alghodhaifi H., Lakshmanan, S.. 2023. "Autonomous Vehicle Evaluation: A Comprehensive Survey on Modeling and Simulation Approaches.” in IEEE Access, vol. 9, pp. 151531-151566, 2021, doi: 10.1109/ACCESS.2021.3125620.

[9] F.Yuen, Y.D.Wong, F.Ma, X.Wang (2020) “The determinants of public acceptance of autonomous vehicles: An innovation diffusion perspective. Journal of Cleaner Production.Vol. 270, 121904, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2020.121904