Using Physiological Metrics to Improve Reinforcement Learning for Autonomous Vehicles

Michael Fleicher
Computer Science Department
Ariel University
Ariel, Israel
michael.fleicher.tal@gmail.com

Oren Musicant
Industrial Engineering Department
Ariel University
Ariel, Israel
musicant.oren@gmail.com

Amos Azaria
Computer Science Department
Ariel University
Ariel, Israel
amos.azaria@ariel.ac.il

***Abstract*—Thanks to recent technological advances Autonomous Vehicles (AVs) are becoming available at some locations. Safety impacts of these devices have, however, been difficult to assess. In this paper we utilize physiological metrics to improve the performance of a reinforcement learning agent attempting to drive an autonomous vehicle in simulation. We measure the performance of our reinforcement learner in several aspects, including the amount of stress imposed on potential passengers, the number of training episodes required, and a score measuring the vehicle's speed as well as the distance successfully traveled by the vehicle, without traveling off-track or hitting a different vehicle. To that end, we compose a human model, which is based on a dataset of physiological metrics of passengers in an autonomous vehicle. We embed this model in a reinforcement learning agent by providing negative reward to the agent for actions that cause the human model an increase in heart rate. We show that such a "passenger-aware" reinforcement learner agent does not only reduce the stress imposed on hypothetical passengers, but, quite surprisingly, also drives safer and its learning process is more effective than an agent that does not obtain rewards from a human model.**

***Index Terms*—reinforcement learning, autonomous vehicles, passengers, driving style, physiological sensing, comfort**

I. Introduction

In a world where 90% of all motor vehicle accidents are caused by some type of human error [1] the search for an automated solution for driving a vehicle seems inevitable. Still, there are safety issues that are wildly publicized, and recent accidents have initiated concerns regarding the drivers' understanding and capability of safely using AV technology [2].

Therefore, it is plausible that individuals would be concerned when they riding an AV. Passenger trust-related barriers prevent universal acceptance of such technology, and recent studies show that 65-75% of American drivers are afraid to ride in a fully self-driving vehicle [3]. Another path to mitigate people's concerns is to influence the driving style of the vehicle. Furthermore, a recent survey [4] suggests that people are open to more sophisticated vehicle technology and ready to embrace new technology, especially if it makes driving safer. Therefore, to promote acceptance of such technology it is necessary to monitor users' reactions to it to adjust the driving style to their reactions. A review paper [5] proposed autonomous passenger awareness factors and compared them with traditional driving-comfort measures, which eventually linked drivers' preferred driving style to their comfort during driving. The comfort feeling during driving can be represented by the lack of stress, and several studies tried to decrease the passenger's stress during driving: [6] proposed a neural network-driven based solution to learning driving-induced stress patterns and correlating it with statistical, structural, and time-frequency changes observed in recorded bio-signals. They envisaged that such a driver-centric safety system would help save precious lives by providing fast and credible real-time alerts to drivers and their coupled cars. [7] presented a system where a simple and low-complexity classification algorithm is used to identify a person's stress while driving a car. We conducted this study to address the lack of an integrated system that aims to drive both more safely and in a calm, smooth, and less stressful experience.

In this paper we aim to link the autonomous driving style expressed by kinetic-characteristics to the passengers' stress-related response and utilize this linkage in a vehicle control system that takes the passenger stress-state into account. We argue that physiological metrics may be utilized to improve vehicle safety and reduce stress levels of potential autonomous vehicles passengers. Indeed, physiological metrics provide essential information related to vehicle safety. For example, a sudden stop may cause physiological metrics to rise, indicating that the vehicle performed a dangerous sequence of actions—even if it eventually resulted in no harm. In this paper, we attempt to harness these physiological metrics to improve the performance of a reinforcement learning agent for an autonomous vehicle. We study a dataset [8] of twenty volunteers who participated in a field experiment in which the AV driving style was adjusted to allow a variety of velocity and acceleration values. We compose a human model based on the relationship between the intensities of the AVs' kinematic events (braking, accelerating, and turning) characterized by velocity, acceleration and jerk, as well as the distance from other vehicles, and changes in heart rate, heart rate variability, and skin conductance. This human model is embedded into a reinforcement learning agent by providing negative rewards

The research presented in this paper was partially funded by the Ministry of Science and Technology of Israel and the Israeli Smart Transportation Research Center.

to actions that result in an increase of levels of stress for hypothetical passengers. We show that such a reinforcement learning based agent that receives input from the human model does not only reduce the stress imposed on hypothetical passengers, but, quite surprisingly, also drives safer and its learning process is more effective than an agent that does not obtain rewards from a human model.

II. Related Work

This section reviews works related to physiological reactions while riding a vehicle, how to mitigate such reactions, and previous studies that consider the passengers' experience while controlling an AV.

A. Stress and Physiological Responses

State anxiety, defined by [9] as a complex emotional response to a perceived threat, characterized by feelings of tension and heightened autonomic nervous system activity. [10] measured participant's heart rate (HR) and found that mean pulse rate was moderately correlated with the anxiety level and claimed that wearable sensors have the potential to be used for assessing anxiety level objectively and unobtrusively to facilitate mental-stress related studies. [11] explored measures of HR and HR variability (HRV) with an imposed stressful situation and suggested that HR and HRV change with a mental task, and that HR and HRV recordings may have the potential to measure stress levels. [12] claimed that "Stress is triggered by something called stressor. A stressor is a stimulus initiating or sparking changes. In general, stressor is classified further into internal stressor and external stressor." they manged to detect an individual's level of stress by measuring heart rate, blood pressure, and Galvanic Skin Response (GSR). [13] used GSR to objectively evaluate stress and arousal levels, and showed that GSR readings significantly increase when task's cognitive load level increases.

B. Manual driving as an External Stressor

In [14] the researchers present methods for collecting and analyzing physiological data during real-world driving tasks to determine a driver's stress level (during manual driving). HR, HRV, and skin conductance level (SCL) were recorded continuously alongside other stress-related measurements while drivers followed a set route through open roads. Data from 24 drives of at least 50-min duration were collected for analysis. The data were analyzed in two ways. Analysis I used features from 5-min intervals of data during the rest, highway, and city driving conditions to distinguish three levels of driver stress with an accuracy of over 97% across multiple drivers and driving days. Analysis II compared continuous features, calculated at 1-s intervals throughout the entire drive, with a metric of observable stressors created by independent coders from videotapes. The results show that skin conductivity and heart rate metrics are most closely correlated with driver stress levels for most participants. These findings indicate that physiological signals can provide a metric of driver stress in future cars capable of physiological monitoring. Such a metric could be used to help manage noncritical in-vehicle information systems.

C. Autonomous driving as an External Stressor

Another work [8], measured physiological signals (HR, eye-movement patterns, and SCL) and self-reported comfort and anxiety levels from passengers in an autonomous vehicle followed by correlation of passengers' response with driving style parameters, including acceleration, jerk (the third derivative of position), and dynamic object distance (proximity of the AV to other objects, like other cars), as well as four events: following a lead vehicle; stopping at a sign; passing a vehicle; a tight turn. The study took place on a closed track in an autonomous vehicle. The results managed to explain the passengers' responses to various driving style in a physical autonomous vehicle, and how the kinetic characteristics of the ride effects on those responses. the presence and proximity of a lead vehicle not only raised the level of all measured physiological responses but also exaggerated the existing effect of the longitudinal acceleration and jerk parameters. SCL response was also found to be a significant estimator of passenger comfort and anxiety. Using multiple independent events to isolate different driving style parameters demonstrates a method to control and analyze such parameters in future studies.

D. Stress due to kinetic indices during driving

Reference [15] used skin conductance responses (SCRs) to measure learner, novice and experienced drivers' psychophysiological responses to the development of driving hazards and found that experienced drivers were twice as likely to produce an SCR to developing hazards as novice drivers and three times as likely when compared with learner drivers. [16] explored whether slight differences in real-world driving task demands could be discriminated by the electrodermal response (EDR). The likelihood of EDR and, whenever present, its duration was both correlated with workload as represented by the deceleration demand. A higher base travel speed and the unexpected demand of the emergency braking situation impacted EDR, thus attesting higher workload level. EDR explained why stopping the vehicle from 50 km/h and slowing down from 80 to 50 km/h was of similar strain. The results further demonstrate that EDR measures can be successfully employed to discriminate multiple levels of workload. [17] conducted a field experiment and manipulated braking demands such as pre-braking speed and and the target speed for braking (30 km/h, a complete stop, or responding to an impending collision) and found that all SCL, HR and HRV were associated with deceleration intensity, and especially when $|g| > 0.5$, and suggested that SCL, HR and HRV can mirror the mental workload elicited by varying braking intensities.

E. Vehicle control

AVs relay on accurate sensory data, utilizing multi-sensor setups and sensors, such as LIDAR, accurate GPS antennas, and high resolution cameras, to provide environment

perception. Control of early versions of AV's was handled via rule-based controllers, where the developers hand-tuned the parameters after simulation and field testing [18], [19]. Recently, deep learning has gained attention due to the success it had achieved in fields such as image classification and speech recognition [20]–[22]. AV use deep learning for various tasks such planning and decision making [23], [24]; perception [25], [26]; as well as mapping and localisation [27]. In early works towards vehicle control through deep learning, [28] introduced an autonomous driving system based on Q-learning combined with learning from the experience of a professional driver. The reward value of the professional driver's strategy and the Q-value learned through the Q-learning method were combined in the pre-training phase to improve convergence speed during training. A filtered experience replay stores a limited number of episodes and allows the elimination of poor experimental rounds from memory, improving convergence on a control strategy. The proposed Deep Q-learning with filtered experiences (DQFE) approach was compared to a naive neural fitted Q-iteration (NFQ) [29] algorithm without pre-training by an experienced driver. During training, it was shown that the DQFE approach reduced the training time by 71.2% for the 300 training episodes. Moreover, during 50 tests on a competition track, the proposed approach completed the track 49 times, compared to only 33 with NFQ. Additionally, DQFE performed better in terms of the mean distance from the center of the track. Therefore, adding filtered experience replay improved the speed of convergence as well as the performance of the algorithm.

F. Discretized vs continuous decision-making

Comparing two neural networks for lane-keeping systems, [30] investigated the effects of discretized and continuous actions. Two approaches, DQN and a Deep Deterministic Actor-Critic (DDAC) algorithm were evaluated in a TORCS simulator [31]. In the two networks developed by the authors, the DQN could only output discretized values (steer, gear, brake, and acceleration), while the DDAC supports continuous action values. The DDAC consisted of two networks; an Actor-Network, a neural network responsible for taking actions based on perceived states, and the Critic Network, which criticizes the value of the action taken. The experimental results showed that the DQN algorithm suffered in performance since it cannot support continuous actions or state spaces. The DQN algorithm is suitable for continuous (input) states. However, it still requires discrete actions since it finds the action that maximizes the action-value function. In this study, we explored different approach for using DQN, which discretely changes continuous outputs to be able to output high range of discrete values without increasing the dimensionality and complexity of the network.

III. Methodology

This section describes the processes of this study and its two primary outcomes:

A. Estimation of stress-related responses of AV passengers based on kinematic indices.

B. The utilization of outcome 1 in a reinforcement learning agent for an autonomous vehicle that receives a penalty according to the expected stress level of the passengers.

A. Relation Between Kinetic Indices and Stress Reactions

The data used in this study was collected by Dillen et al. [8]. We describe the data in brief: The study was conducted on a closed and circular test track in Waterloo, Ontario, Canada, and involved using an AV, a "normal" vehicle, and human participants. The AV, a Lincoln MKZ hybrid research platform developed at the University of Waterloo [32] to reach Level 3 autonomy. The AV was fitted with an array of sensors, including a Novatel IMU, Novatel GPS, Velodyne LIDAR. The usage of the GPS and IMU allowed the researchers to collect the raw velocity, acceleration, and Jerk (the third derivative of position) with respect to both longitudinal and lateral directions during the entire experiment. The LIDAR allowed the researchers to measure the distance to surrounding objects such as "normal" vehicles deliberately parked alongside the road. The motion planning algorithm [33] onboard the vehicle used the sensor information to select a trajectory in accordance with the intended driving style and the current environmental constraints. The participants' GSR, ,HR, and HRV were measured. A Shimmer3+ device was used to measure GSR (500Hz sample rate) and obtain a PPG signal for HR and HRV. Using this experiment's data allowed us to investigate the relation between the kinetic indices (including the distance to surrounding vehicles) and the stress-related physiological responses.

Preprocessing of kinetics and physiological data: To filter out long-term drifts in acceleration values, we analyzed the raw acceleration signal minus its moving median (30sec). Then the acceleration data (lateral and longitudinal) was subjected to outlier detection and removal according to the 1.5 IQR method. That is, any sample above the third quartile plus 1.5 times the interquartile range or below the first quartile minus 1.5 times the interquartile range, was removed. **Detection of kinematic and physiological events:** The kinetics of the vehicle apply forces on the passenger. One way to estimate the linkage between the forces applied to the passengers and their physiological response is to look at time-series data. Reference [8] report that the presence and proximity of a lead vehicle not only raised the level of all measured physiological responses but also exaggerated the existing effect of the longitudinal acceleration and jerk parameters. Skin response was also found to be a significant predictor of passenger comfort and anxiety. Our approach is different and does not assume that the kinematic forces and physiological responses occur simultaneously. To illustrate, Fig. 1 describes the longitudinal acceleration and SCL versus measurement number (measured in 500[Hz]). The figure presents three acceleration events: Braking, with peak braking at 21.72K measurement, accelerating with peak acceleration at 21.77K measurement, and braking with peak at 21.84k measurement. The letters S and E signify the start and end of each kinematic (and physiological)

event identified in the data. A kinematic event (acceleration state 1 or -1 for throttle and brake events, respectively) begins when the acceleration is higher than 0.1[$\frac{m}{sec^2}$] or lower than -1.1[$\frac{m}{sec^2}$]. There are also events (that we call Skin Conductance Responses - SCR), with peaks at times: 21.69K measurement, 21.75K measurement, and 21.84K measurement.

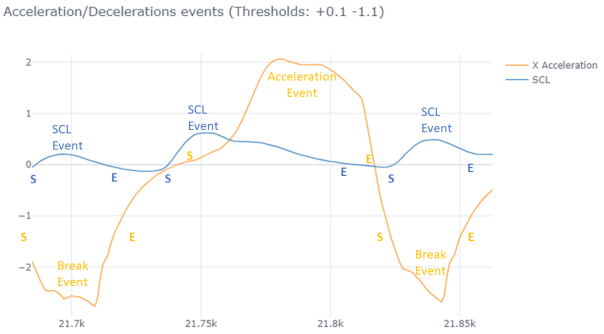

Fig. 1: Acceleration, and SCL response during some trials of the experiment. x axis represents measurement number sampled in 10Hz, orange line represents actual acceleration [$\frac{m}{sec^2}$], blue line represents SCL response, thresholds of -1.1, +0.1 [$\frac{m}{sec^2}$] defines the acceleration/braking event type.

The time of the peaks of the kinematic events is not perfectly synchronized with those of the physiological event, suggesting that the psychological events can appear at the beginning of the kinematic event (proactive response) at the end of it (reactive). For example, in [15], SCRs began before a hazardous event onset reflecting the driver's (in their case) ability to anticipate hazards. Here too, the peak in the physiological event (at time 21.75k measurement) preceded the peak of the acceleration event. The peak of the braking events and their corresponding SCR are almost aligned. While the analysis of the correlation between two time-series (physiological and kinematic) requires an assumption about the time lag between the kinematic time series and the physiological time series, analysing data per event allows more flexibility: the correlation uses the pairs of peaks in the kinematic and the physiological data in the overlapping events even if the peaks did not occur at the same time. Many other acceleration thresholds were explored using grid search. we tested different thresholds for both negative and positive accelerations, in the longitudinal and lateral directions. For each threshold, we calculated linear regression to relate the kinetic characteristics of the driving, to the passenger's HR, HRV and SCL. Then we choose the threshold which facilitated the highest R^2 score between kinetic indices to the passenger's stress response. This procedure yielded 418 braking events, 486 acceleration events, 692 right turn. Due to the circular structure of the driving track, this dataset contains turns to the right side only.

B. Vehicle control

This section outlines how we use the stress estimation model in an AV motion control planning algorithm called a "passenger-aware" agent. The motion control of a vehicle can be broadly divided into two tasks - lateral and longitudinal motion; the steering of the vehicle controls the lateral motion of the vehicle, while longitudinal motion is controlled by manipulating the gas and brake pedals of the vehicle. Lateral control systems aim to control the vehicle's position on the lane and carry out other lateral actions such as lane changes or collision avoidance maneuvers. In the deep learning domain, this is typically achieved by capturing the environment using the images from onboard cameras as the input to the neural network. Longitudinal control manages the vehicle's acceleration such that it maintains the desirable velocity on the road, keeps a safe distance from the preceding vehicle, and avoids rear-end collisions. While lateral control is typically achieved through vision, longitudinal control relies on relative velocity and distance measurements to the preceding/following vehicles. The data set [8] was used to develop several SCL and HR estimators, and their performance in terms of accuracy will be described in the results section. Due to its relatively high $R_{squared}$, HR response was chosen as the estimated human stress indicator in the further work described in this section. Due to its relatively small number of sub-estimators and parameters, the $BaggingRegressor$ model [34] is used as the primary estimator used in this section. To estimate the passenger's HR as a response to the vehicle kinetics state, the following variables were used as the estimator's input:

TABLE I: Variables which used as an input to our HR estimator model

Variable	Description	Units
V	Vehicle velocity	[$\frac{m}{sec}$]
A_x	Forward acceleration	[$\frac{m}{sec^2}$]
J_x	Forward jerk	[$\frac{m}{sec^3}$]
A_y	Lateral acceleration	[$\frac{m}{sec^2}$]
J_y	Lateral jerk	[$\frac{m}{sec^3}$]
D_x	Forward distance from surrounding objects	[m]
D_y	Lateral distance from surrounding objects	[m]
N_{steps}	Episodic step count	[#]
T	1 if a turning event occurs, and 0 otherwise	[Binary]
B	1 if braking event occur, and 0 otherwise	[Binary]
A	1 if acceleration event occur, and 0 otherwise	[Binary]
N_T	Turning events episodic-count	[#]
N_B	Braking events episodic-count	[#]
N_A	Acceleration events episodic-count	[#]

To design a vehicle control mechanism, we used an API of LGSVL-SIM [35] simulator as a training environment, which facilitated the same conditions as conducted on Dillen et al. [8], such as maximum acceleration and velocity and nearby NPC's (non-player characters). The simulated vehicle (hereafter Ego-vehicle) uses several sensors to feed the observation space, control space, reward function, and our HR estimator.

Segmentation camera sensor: An image obtained from

a front camera on the Ego-vehicle's roof as input and transformed using a segmentation algorithm. Objects in the image are colored corresponding to their tag: Ego-vehicle (Black), NPC (Blue), Road (Purple), Horizon (White).

Controller Area Network (CAN bus) sensor: Sends data about the Ego-vehicle chassis. The data includes Velocity[m/s], Angular velocity[rad/sec], Longitudinal Acceleration [m/sec^2], Angular Acceleration[rad/sec^2], Lateral Acceleration [m/sec^2], Longitudinal Jerk[m/sec^3], Lateral Jerk[m/sec^3].

Observation space: Contains data retrieved from the Simulator's sensors. Include the current image and the last three images obtained from the segmentation camera sensor. The observation space also includes data from CAN bus sensor.

Control State: Represents the Ego-vehicle's current throttle [%], Brake [%], Steering [+- %] states.

Action space: Contains the following nine combinations of using the gas/brake pedals and the steering wheel:
no action, throttle, brake, throttle+right, throttle+left, brake+right, brake+left, right, left. Before every step, the agent chooses an action from the action space. every chosen action is increasing/decreasing/resets the current Control state in the following form:
The throttle pedal increases its previous state by 5% if the actions throttle, throttle + right, throttle + left is chosen and drops back to 0% for any other action chosen. The brake pedal is increasing its previous state by 5% if the actions brake, brake + right, brake + left is chosen, and drops back to 0% for any other action chosen.
The steering wheel increases (decreases) by 5% if the actions right, throttle + right, brake + right (left, throttle + left, brake + left) action is chosen and keeps the same for any other action chosen.

Reward Function: After each step, the agent receives a reward according to the results of its action. The reward function uses information about the vehicle location, and while the vehicle's position is inside the road, the reward function calculates the following variables:
Distance traveled D_t: The euclidean distance from the current to the last step's position.
Passenger factor P: The extent at which the passenger's stress exceeds some threshold:

$$P = \begin{cases} (\frac{H_p}{H_t})^2, & \text{if } H_p \geq H_t \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

While H_p represents the HR estimated by the $BaggingRegressor$ model using the current step's kinetic information, and H_t represents the threshold which, if passed, indicates that the passenger feels some level of stress. In our case, H_t is set to 90[BPM].

Delay penalty D_p: a value which increases itself in every step. We used the delay penalty to "punish" the agent for standing still and encourage the agent to move along the road. The delay penalty calculates as follows:

$$D_p = 0.05\frac{S_n - 1}{1000} \quad (2)$$

While S_n is the current session's step count. As a result, the reward is computed using the following formula:

$$R = D_t - D_p - P. \quad (3)$$

If the vehicle location is outside of the road or it collides with another vehicle or any obstacle, the reward function resets the current driving session and the environment is initialized with the vehicle at the starting position. We compare the passenger-aware agent's driving performance to a "standard agent", which is identical to the passenger-aware agent in all aspects except that it does not account for the human stress value P. That is, the reward for the standard agent is computed by:

$$R = D_t - D_p. \quad (4)$$

Architecture: We integrated the heart rate estimation model alongside with kinetic and visual features into a DQN architecture [36] which after trained, called "passenger aware agent". The DQN uses the observation space as an input and, as an output, returns one element from the action space as an action. Fig. 2 describes the entire DQN network structure.

As an agent takes actions and moves through an environment, it learns to map the observed state of the environment to an action. An agent will choose an action in a given state based on a "Q-value" - a weighted reward based on the expected highest long-term reward. A Q-Learning Agent learns to perform its task such that the recommended action maximizes the potential future rewards. This method is considered an "Off-Policy" method, meaning its Q values are updated assuming that the best action was chosen, even if the best action was not chosen. This network learns an approximation of the Q-table - a mapping between the states and actions that an agent will take. For every state, we will have nine actions that can be taken. The environment provides the state, and the action is chosen by selecting the larger of the nine Q-values estimated in the output layer. Our network is built as a combination of two different neural networks merged to decide one action as an output. The upper part (Fig. 2) is a network that can process images data (4 stacked images in our case) and use convolution layers to understand features in the image, such as the road and NPCs. The upper network receives 4 84x84 gray images and then performs a transformation on two convolution layers while using max pulling in between. As an output, the upper network outputs a dense layer of 512 units. This output will merge with the lower network's output.
The lower network receives the kinetic information of the vehicle from the CAN bus sensor and uses it as an input. The input is then processed through two dense layers of 256 units. The last layer of 256 units are merged with the upper network's last layer of 512 units to a 768 dense layer, which then outputs 9 Q-values. Later on, the agent chooses the action in which its representative Q-values have the highest values.

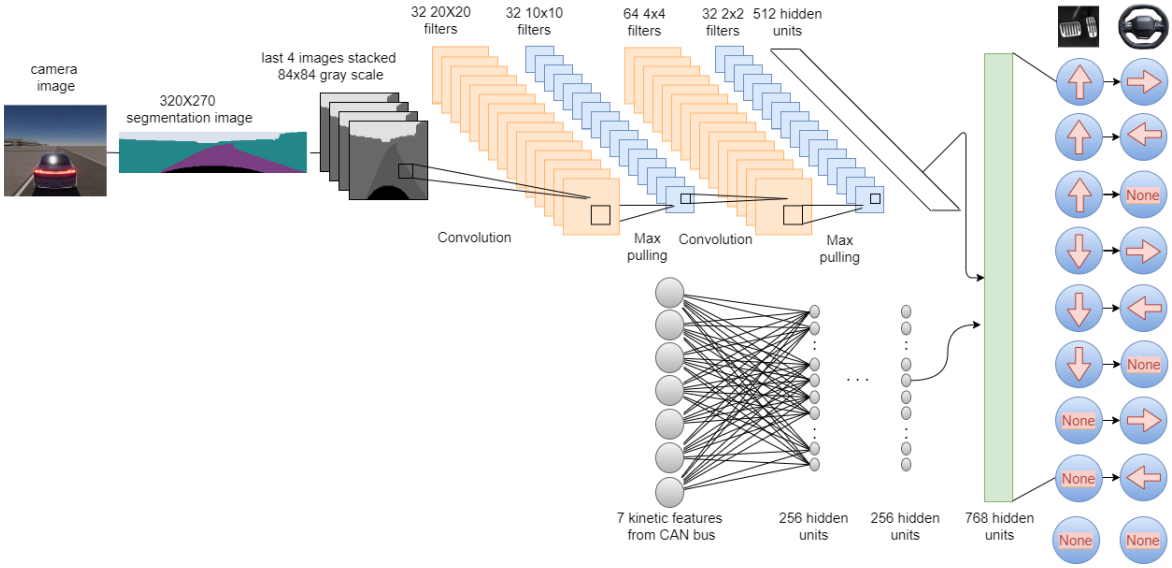

Fig. 2: Visualization of the DQN architecture we used. the network receives 4 images and kinetic features as and input and outputs one action from the action space for controlling the vehicle's wheel and brake/throttle pedals.

IV. Results

This chapter describes in detail the results of two primary outcomes:
A. Estimation of stress-related reactions to kinetic indices on AV passengers.
B. The utilization of outcome 1 in a motion control planning algorithm that considers the passenger's responses in its actions.

A. Estimation of stress-related reactions to kinetic indices on AV passengers

We used non-linear regressors to produce a function that receives several vehicle kinetics as inputs and estimates the passenger's HR as output. the regressors were trained on the entire dataset published by [8], measured with a 500Hz sample rate (after performing the pre-processing described in Section III-A. Thus, we trained several regressors for estimating the passenger's HR, HRV and SCL. Furthermore, the accuracy of each model is described below:

TABLE II: R^2 and RMSE of different models for estimating passengers' HR, HRV and SCL during riding AV.

	HR		SCL		HRV	
Model	R^2	RMSE	R^2	RMSE	R^2	RMSE
RandomForestRegressor	0.8	0.07	0.73	0.09	0.17	0.23
ExtraTreesRegressor	0.8	0.07	0.77	0.08	0.14	0.24
BaggingRegressor	0.77	0.07	0.7	0.1	0.21	0.22

Table II shows that the ability to estimate passengers' stress levels during AV riding exists and the average error of such estimations is less than 1 BPM for HR estimations. However, there is high importance to consider that those estimations will estimate accurately only on the passengers who participated in [8]'s experiments. For the generalization of such estimators, there is a need for data from a larger population. However, in our work, the provided dataset was sufficient for implementing the BaggingResgressor in an AV motion planning and control algorithm, as described in the next section.

B. Motion control planning algorithm which considers the passenger's responses in its actions

After training both passenger-aware and standard agents five times each, we compared the distance travelled by each during an episode. The episode represents a driving session, and while training our algorithms, the agents made thousands of episodes. In total, out of 2,300,000 sampled steps, the passenger-aware agent raised the estimated passenger's stress level 1.9% of time, while the standard agent, (which wasn't designed to address the passenger's stress) made it 39% of time. In Fig. 3, we plotted the distance travelled by the agents during each episode to compare their performance.

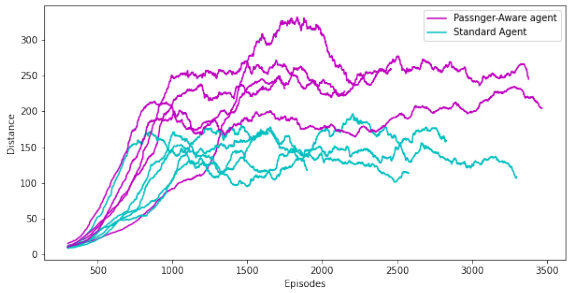

Fig. 3: Distance travelled vs. the number of episodes by agent type (colors). each agent made five different learning sessions

In Fig. 3, if each agent is examined separately, we can see a significant difference between the agent's performance from one learning session to another. This phenomenon can be

explained by the fact that before each step, the agent chooses its actions in some level of randomness. The use of epsilon greedy [37] to deal with the exploration-exploitation dilemma [38] during the training process can explain this randomness and possibly explain the temporary high distance travelled on one of the passenger-aware agent's trials (around 1500-2000 episodes). Fig. 3's significant outcome is the existence of a mechanism that can learn how to avoid increasing the passenger's stress reactions and control and plan the motion of a vehicle - all at the same time. Furthermore, while comparing the performance of our passenger-aware agent to a standard agent (which is the same algorithm with just one difference - the passenger's factor calculated in the algorithm's reward function, as described in Section III-B), we can see that the learning process of passenger-aware agent converges to a higher distance of driving without resets the episodes. Those results can even tell that it is recommended to implement the passenger factor in other algorithm's reward functions to gain higher performance. We then analysed how each agent interacts with other vehicles during the ride, and mainly focused on the proximity to other vehicles during an episode. We collected all the events where the agent's vehicle passed near a surrounding vehicle, and for each event, we took the minimum Euclidean distance between those two vehicles. Fig. 4 presents the results.

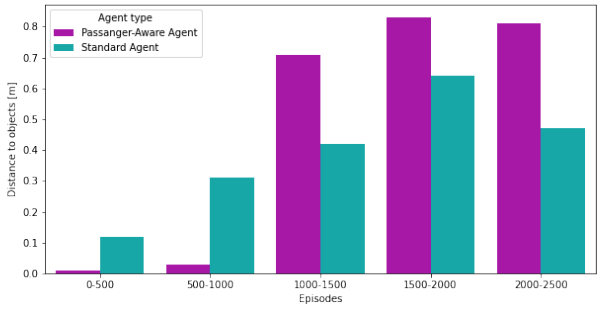

Fig. 4: The distance between the agent's vehicle and the closest vehicle next to it for each type of agent averaged across 500 episodes.

By Fig. 4, we can learn that as time passed, the passenger-aware agent learned to keep its distance from surrounding vehicles, significantly compared to the standard agent. Once reaches 1000 episodes, the passenger aware agent learns to keep greater distance from the other vehicles. Furthermore, we assume the agent acts in such way to avoid triggering stress reactions in passengers, like [8], which found that the presence and proximity of a lead vehicle raised the level of all measured physiological responses. As a result, such behavior can be interpreted as a bit safer. The primary outcome of these results is that eventually, and as expected, driving with passenger awareness can also contribute to a safer driving strategy, especially when comparing the two above control agents' distance-keeping abilities. We then inspected how driving episodes where ended during training. In our environment, an episode ends if one of the following occurs: the vehicle went off the road, the vehicle collides with another vehicle, and the session reaches the maximum steps defined for each episode (in our case is 1000 steps). Fig. 5 shows distribution of averaged episode-ending cause around episodes 2000-2500 for all training sessions. We did so to observe how each agent was able to sense the surroundings and avoid collisions.

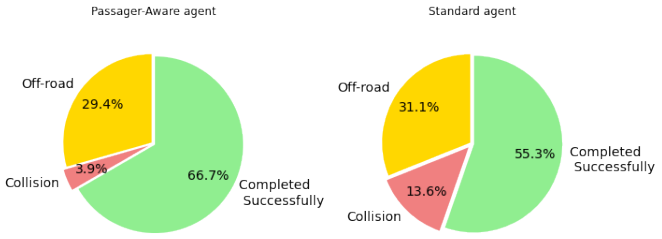

Fig. 5: for each learning session, we averaged the collisions/off-road/completed successfully episode-ending cause percentage of episodes 2000-2500, where the agents' driving ability was mature, and learning curve converged.

Fig. 5 shows that the passenger-aware agent eventually managed to collide with other vehicles less than the standard agent. It can be explained by Fig. 4 outcomes which is, as expected, the idea that driving with higher distance keeping can contribute to collision avoidance. Furthermore, and in general speaking, the passenger-aware agent was able successfully complete its episodes with higher probability than the standard agent.

V. Discussion

The results presented in Section IV-A connect AV passengers' stress to the ride's kinetic characteristics. Some key points to remember while considering those results are that estimations were trained on a significantly small sample size (20 participants). Furthermore, the data set that this study used to develop the HR estimations upon, contains driving sessions of limited speed range (of 0-34 $[\frac{km}{h}]$) and, therefore, may not be accurate when driving at higher speeds. To address those limitations, we intend to run another study that will include a larger sample size and a higher range of speeds. Data collected from our future study can increase generalization of the current study's results. Furthermore, the results presented in Section IV-B support the usage of the passenger stress due to the potential safety benefits – our passenger aware AV had fewer collisions rate during training. However, to proceed with on-road testing, there should be a usage of control agents with a much lower collision rate (around zero), which is measured by much more simulations compared to only five as we did. A key point to remember while considering this study's results is that even though the passenger factor was reliable by presenting high accuracy of HR estimation, it was calculated by a significantly small population size and thus, cannot be

generalized to a higher or external population other than its population. Therefore, those results need to be addressed only as proof of the feasibility of such an AV control strategy. Another way to handle the disadvantages of those results is to examine how such a passenger-aware agent can learn its HR estimation upon individual users, and by then, to learn how to provide personal driving experiences custom-made to individuals instead of to some population.

REFERENCES

[1] N. N. C. for Statistics and Analysis, "2016 fatal motor vehicle crashes: Overview." https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456.

[2] "Tesla update halts automatic steering if driver inattentive." https://phys.org/news/2016-09-tesla-halts-automatic-driver-inattentive.html. (Accessed on 21/06/2021).

[3] E. Edmonds, "Aaa: Three in four americans remain afraid of fully self-driving vehicles." https://newsroom.aaa.com/2019/03/americans-fear-self-driving-cars-survey/, 2019.

[4] E. Edmonds, "Aaa: Today's vehicle technology must walk so self-driving cars can run." https://newsroom.aaa.com/2021/02/aaa-todays-vehicle-technology-must-walk-so-self-driving-cars-can-run/, 2021.

[5] M. Elbanhawi, M. Simic, and R. Jazar, "In the passenger seat: investigating ride comfort measures in autonomous cars," *IEEE Intelligent transportation systems magazine*, vol. 7, no. 3, pp. 4–17, 2015.

[6] R. R. Singh, S. Conjeti, and R. Banerjee, "A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 740–754, 2013.

[7] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Low-complexity classification algorithm to identify drivers' stress using electrodermal activity (eda) measurements," in *The World Thematic Conference-Biomedical Engineering and Computational Intelligence*, pp. 25–33, Springer, 2018.

[8] N. Dillen, M. Ilievski, E. Law, L. E. Nacke, K. Czarnecki, and O. Schneider, "Keep calm and ride along: passenger comfort and anxiety as physiological responses to autonomous driving styles," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[9] C. D. Spielberger, "Theory and research on anxiety; in anxiety and behavior," *Spielbergered*, pp. 3–22, 1966.

[10] Y. Zheng, T. C. Wong, B. H. Leung, and C. C. Poon, "Unobtrusive and multimodal wearable sensing to quantify anxiety," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3689–3696, 2016.

[11] J. Taelman, S. Vandeput, A. Spaepen, and S. V. Huffel, "Influence of mental stress on heart rate and heart rate variability," in *4th European conference of the international federation for medical and biological engineering*, pp. 1366–1369, Springer, 2009.

[12] N. Widanti, B. Sumanto, P. Rosa, and M. F. Miftahudin, "Stress level detection using heart rate, blood pressure, and gsr and stress therapy by utilizing infrared," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, pp. 275–279, Ieee, 2015.

[13] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, "Galvanic skin response (gsr) as an index of cognitive load," in *CHI'07 extended abstracts on Human factors in computing systems*, pp. 2651–2656, 2007.

[14] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[15] N. Kinnear, S. W. Kelly, S. Stradling, and J. Thomson, "Understanding how drivers learn to anticipate risk on the road: A laboratory experiment of affective anticipation of road hazards," *Accident Analysis & Prevention*, vol. 50, pp. 1025–1033, 2013.

[16] C. Collet, E. Salvia, and C. Petit-Boulanger, "Measuring workload with electrodermal activity during common braking actions," *Ergonomics*, vol. 57, no. 6, pp. 886–896, 2014.

[17] O. Musicant, A. Botzer, I. Laufer, and C. Collet, "Relationship between kinematic and physiological indices during braking events of different intensities," *Human factors*, vol. 60, no. 3, pp. 415–427, 2018.

[18] T. Le-Anh and M. De Koster, "A review of design and control of automated guided vehicle systems," *European Journal of Operational Research*, vol. 171, no. 1, pp. 1–23, 2006.

[19] M. Pasquier, C. Quek, and M. Toh, "Fuzzylot: a novel self-organising fuzzy-neural rule-based pilot system for automated vehicles," *Neural networks*, vol. 14, no. 8, pp. 1099–1112, 2001.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[21] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[22] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[23] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 187–210, 2018.

[24] S. M. Veres, L. Molnar, N. K. Lincoln, and C. P. Morice, "Autonomous vehicle control systems—a review of decision making," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 225, no. 2, pp. 155–195, 2011.

[25] R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Ten years of pedestrian detection, what have we learned?," in *European Conference on Computer Vision*, pp. 613–627, Springer, 2014.

[26] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transportation research part C: emerging technologies*, vol. 89, pp. 384–406, 2018.

[27] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.

[28] W. Xia, H. Li, and B. Li, "A control strategy of autonomous vehicles based on deep reinforcement learning," in *2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, pp. 198–201, IEEE, 2016.

[29] M. Riedmiller, "Neural fitted q iteration–first experiences with a data efficient neural reinforcement learning method," in *European conference on machine learning*, pp. 317–328, Springer, 2005.

[30] A. E. Sallab, M. Abdou, E. Perot, and S. Yogamani, "End-to-end deep reinforcement learning for lane keeping assist," *arXiv preprint arXiv:1612.04340*, 2016.

[31] B. Wymann, E. Espié, C. Guionneau, C. Dimitrakakis, R. Coulom, and A. Sumner, "Torcs, the open racing car simulator," *Software available at http://torcs. sourceforge. net*, vol. 4, no. 6, p. 2, 2000.

[32] I. Colwell, "Runtime restriction of the operational design domain: A safety concept for automated vehicles," Master's thesis, University of Waterloo, 2018.

[33] Y. Zhang, H. Chen, S. L. Waslander, J. Gong, G. Xiong, T. Yang, and K. Liu, "Hybrid trajectory planning for autonomous driving in highly constrained environments," *IEEE Access*, vol. 6, pp. 32800–32819, 2018.

[34] G. Louppe and P. Geurts, "Ensembles on random patches," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 346–361, Springer, 2012.

[35] G. Rong, B. H. Shin, H. Tabatabaee, Q. Lu, S. Lemke, M. Možeiko, E. Boise, G. Uhm, M. Gerow, S. Mehta, *et al.*, "Lgsvl simulator: A high fidelity simulator for autonomous driving," in *2020 IEEE 23rd International conference on intelligent transportation systems (ITSC)*, pp. 1–6, IEEE, 2020.

[36] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[37] M. Rawson and R. Balan, "Convergence guarantees for deep epsilon greedy policy learning," *arXiv preprint arXiv:2112.03376*, 2021.

[38] O. Berger-Tal, J. Nathan, E. Meron, and D. Saltz, "The exploration-exploitation dilemma: a multidisciplinary framework," *PloS one*, vol. 9, no. 4, p. e95693, 2014.