Comfort-Oriented Motion Planning for Automated Vehicles Using Deep Reinforcement Learning

Rajesh, Nishant; Zheng, Yanggu; Shyrokau, Barys

DOI
10.1109/OJITS.2023.3275275

Publication date
2023

Document Version
Final published version

Published in
IEEE Open Journal of Intelligent Transportation Systems

IEEE Open Journal of
Intelligent Transportation
Systems

Received 12 January 2023; revised 1 April 2023 and 23 April 2023; accepted 8 May 2023. Date of publication 11 May 2023; date of current version 22 May 2023.

Digital Object Identifier 10.1109/OJITS.2023.3275275

Comfort-Oriented Motion Planning for Automated Vehicles Using Deep Reinforcement Learning

NISHANT RAJESH[1], YANGGU ZHENG[2], AND BARYS SHYROKAU[2]

[1]Simulation and Test Solutions, Siemens Digital Industries Software B.V., Helmond, The Netherlands

[2]Department of Cognitive Robotics, Delft University of Technology, 2628 CD Delft, The Netherlands

CORRESPONDING AUTHOR: Y. ZHENG (e-mail: y.zheng-2@tudelft.nl)

This work was supported by the European Union Horizon 2020 Framework Program through Marie Skłodowska-Curie Actions under Grant 872907.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Human Research Ethics Committee of Delft University of Technology under Application No. 2405.

ABSTRACT Automated vehicles promise numerous advantages to their users. The proposed benefits could however be overshadowed by a rise in the susceptibility of passengers to motion sickness due to their engagement in non-driving tasks. Increasing attention is paid to designing vehicle motion to mitigate motion sickness. In this work, the deep reinforcement learning (DRL) method is used to plan vehicle trajectories, with a focus on minimizing low-frequency accelerations. These are known to be the primary cause of motion sickness. The goal is achieved by incorporating a frequency-weighted discomfort term into the reward function during training. The ability of the trained agent to target undesirable frequencies in accelerations is verified by comparing it with another agent trained for improving overall acceleration comfort. A reduction of 9.6% in frequency-weighted discomfort is achieved. The motion plan from the DRL agent is further compared with trajectories generated by human drivers in real-world scenarios. The results demonstrate comparable performance between the DRL agent and human drivers. Meanwhile, a significant reduction in online computation time has been observed when compared to a motion planner based on numerical optimization.

INDEX TERMS Automated driving, deep reinforcement learning, motion planning, motion sickness, proximal policy optimization.

I. INTRODUCTION

AUTOMATED driving is advancing fast in recent years. It is attractive thanks to the potential benefits they offer in terms of improved safety, higher traffic efficiency, and an increase in user productivity. Being freed from the responsibility of driving the vehicle, users of automated vehicles are expected to engage in numerous non-driving tasks ranging from conversing with co-passengers and listening to music to texting and Web surfing [1]. In order for the passengers to be able to perform such tasks, it is imperative that automated vehicles provide a high level of driving comfort. It is anticipated that with the advent of complete automation in cars, there would be an increased susceptibility of passengers to motion sickness [2]. This could be due to a multitude of reasons. The driver would take on a much more passive role, especially with higher levels of automation, which is well known to increase motion sickness [3]. The industry is also re-imagining vehicle cockpit design, unveiling concepts with office-like environments, rearward-facing seats, and passengers facing each other. Combined with the passengers engaging in non-driving tasks, this would lead to a lack of a stable visual horizon and lower predictability of the direction of motion. The combined effect of all these factors may very well lead to a significant increase in the occurrence of motion sickness, posing a substantial threat to the envisioned benefits of automation, and consequently to the acceptance of automated vehicles among customers.

Motion sickness arises from illusory or actual passive self-motion. The symptoms of motion sickness range from drowsiness and fatigue to stomach awareness and nausea [4].

The review of this article was arranged by Associate Editor Jia Hu.

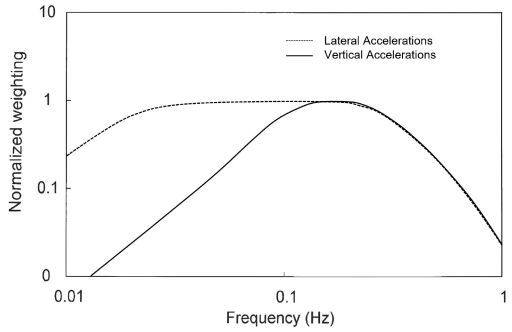

FIGURE 1. Motion Sickness frequency weightings for lateral and vertical accelerations, as adapted from [8].

The most widely accepted theory explaining motion sickness is the sensory mismatch theory, which postulates that motion sickness arises from the conflict between anticipated and sensed motion stimuli [5]. The Central Nervous System (CNS) maintains an internal model of the dynamics of the human body, which estimates the spatial orientation of the body by fusing information from the motor outflow and noisy sensory signals. The conflict between these efference signals with the polysensory afference signals obtained from the sensory organs is used to update and improve the internal observer model but also gives rise to motion sickness. The incidence of motion sickness is predominantly caused by low-frequency accelerations (below 0.5 Hz), with the effect peaking around a frequency of 0.2 Hz for vertical accelerations [6], with a similar peak for longitudinal oscillations [7]. For lateral accelerations, it was found that the incidence of motion sickness was independent of frequency from 0.0315 to 0.25 Hz, followed by decreasing intensity with higher frequency levels [8]. The frequency weighting filters to predict the incidence of nausea for lateral and vertical oscillations have been shown in Fig. 1. As is evident from the frequency weighting, in order to efficiently inhibit the incidence of motion sickness, it is necessary to deal with low-frequency accelerations. This would in turn require motion planning over longer time horizons to accurately predict low-frequency acceleration components.

Although some studies explored the design of automated vehicles to mitigate motion sickness among passengers such as through the layout of the seating arrangement [9], provision of audio and visual cues [10], [11], suspension tuning [12], and the use of advanced suspension actuators [13], they do not directly address the underlying cause of motion sickness, which is the motion regime itself. In a study on comfort in ADS, it has been shown that the motion regime needs to be designed in a manner specific to the maneuver or situation [14]. The magnitudes of accelerations generated by different drivers are found to be independent of the vehicle characteristics and were heavily influenced by the driving style of the individual driver [15]. In [16], the driver was found to be heavily implicated in the generation of motion sickness among passengers, with low-frequency lateral accelerations being primarily responsible for the nauseogenic symptoms. No significant correlation between vertical, roll, or pitch motion and motion sickness was established. Based on these findings, it is safe to conclude that to effectively combat the occurrence of motion sickness in automated vehicles, the vehicle motion itself needs to be planned with particular attention to lateral and fore-aft accelerations.

Passenger comfort has been considered as an objective in motion planning and studied extensively through a variety of methods including planning smooth paths using clothoids, Bezier curves, and polynomial splines, minimizing acceleration and jerk values in an optimization-based framework, or other motion planning methods such as Rapidly-exploring Random Trees [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]. There has also been research into the ride comfort in terms of vertical oscillations in AVs [27], and also high-level route planning [28]. However, very limited research works have directly addressed motion sickness through motion planning. The motion planning problem was formulated as an optimal control problem in [29], where the objective was to minimize the Motion Sickness Dose Value (MSDV). It also revealed the relationship between sickness levels and travel time. A similar goal was attempted in [30], where a high pass filter was employed and the limits in vehicle actuators are expected to filter out higher-frequency motion. A more complex prediction model for motion sickness, namely the 6-degree-of-freedom subjective vertical conflict (SVC) model, was used in [31] for optimizing the vehicle trajectory as well as the gains of the path-following controller for a lane change maneuver. The works mentioned above mitigate motion sickness with motion planning by solving a numerical optimization problem. While they could potentially find an optimal solution, they are highly demanding on computational resources. Given the limited computational power available onboard, it could be challenging to solve the underlying optimization problem in real time. In particular, including the frequency weighting filters in the optimization scheme introduces extra complexity to the planning problem.

To improve computational efficiency, learning-based approaches could be an attractive alternative as they can effectively shift the bulk of the computational burden offline and are hence far less demanding on onboard resources. They have been successfully applied to a plethora of engineering problems including object detection and classification, speech recognition, content recommendation, etc. For our application though, the more popular branch of supervised learning techniques is not applicable. Because they require labeled training data in large quantities that are difficult to collect. To do so, human subjects should be exposed to sickening driving regimes and their responses should be recorded. The inherent variability among humans of susceptibility to motion sickness makes it challenging to reliably quantify the nauseogenicity of imposed motion. Instead, deep reinforcement learning (DRL) is deemed a more suitable

approach to this problem. It is a machine learning paradigm that combines the fields of deep learning and reinforcement learning. Reinforcement learning does not require labeled training data but only needs a representative training environment. DRL has already been applied effectively to various levels of motion planning problems, with the notable advantage of requiring relatively low computational requirements for the trained network [32]. Some automotive applications of DRL include behavioral decision making, path planning to end-to-end vehicle control [33], [34], [35], [36], [37], [38]. DRL has also been shown to be able to successfully capture long-term dependencies in systems when applied in conjunction with methods such as Monte Carlo Tree Search and Long Short Term Memory Neural Networks [39], [40], [41]. This could be interesting to explore with regard to capturing the effects of low-frequency accelerations in motion sickness.

The contribution of this paper is a DRL approach to motion planning that minimizes motion sickness among passengers of automated vehicles by optimizing the vehicle trajectory to reduce nauseogenic accelerations. In particular, the ability of DRL to shape the frequency of longitudinal and lateral vehicle accelerations to reduce overall nauseogenicity has been investigated. This was done by comparing the frequency domain performance of a DRL agent trained to minimize frequency-weighted accelerations to an agent trained on unweighted accelerations. The performance of these agents is also compared to a planner solving the respective environments using an optimization-based technique, to measure how close the agents come to 'solving' the designed training environment. To further establish the applicability of DRL to motion planning with regard to motion sickness, experimental trials with human drivers were carried out on an actual road section to establish a baseline of human performance. The nauseogenicity of the trajectories followed by the human drivers is compared to those generated by DRL agents, over a range of travel times to account for varying driving styles among humans.

The paper is structured as follows. Section II gives a basic overview of DRL and details the setup for training and evaluating the agent. In Section III, the experimental setup for establishing the human baseline has been explained. Section IV details the results of the simulation and the comparisons between the human drivers and the trained agent, with the final conclusions of the study in Section V.

II. DRL TRAINING ENVIRONMENT

A. DEEP REINFORCEMENT LEARNING

Reinforcement learning is a sub-field of machine learning in which an agent learns to perform a task through a trial-and-error process. The agent is trained within an environment where an action taken leads to a reward. In deep reinforcement learning, this agent's observation-action mapping is approximated by a deep neural network. The environment is essentially a system governed by a state transition function:

$$s_{k+1} \sim P(.|s_k, a_k) \tag{1}$$

where $s_k \in S$ is the state, and $a_k \in A$ is the action taken by the agent. The agent acts according to a policy π_θ parameterized by $\theta \in \mathbb{R}^K$, which is given as:

$$a_k = \pi_\theta(a_k|s_k) \tag{2}$$

A series of actions taken by the agent following the policy π till a terminal state is reached is called a rollout or a trajectory $\tau = [s_{0:H}, a_{0:H}]$, where H is the horizon, and the steps from initiation s_0 to the terminal state s_H form an episode. For every action the agent takes, the environment returns a scalar reward r, which is modeled by a reward function:

$$r_k = R(s_k, a_k) \tag{3}$$

The expected value of the accumulated reward over a period of time is called the return $J(\theta)$:

$$J(\theta) = \mathbb{E}\left\{\sum_{k=0}^{\infty} \gamma^k r_k\right\} \tag{4}$$

where $\gamma \in [0, 1)$ is the discount factor. The agent interacts with the environment and samples trajectories, with the objective of learning an optimal policy π_θ^* which maximizes the expected return. Most of the common DRL algorithms are based on some form of policy gradient, and parameter update using gradient ascent

$$\theta_{h+1} = \theta_h + \alpha_h \Delta_\theta J(\theta = \theta_h) \tag{5}$$

where h is the update step of the policy and α_h is the learning rate for updating the weights of the network. The algorithms that use a gradient estimate to perform the parameter update are known as REINFORCE algorithms [42]. These REINFORCE algorithms are easy to implement but suffer from instability during training, low sampling efficiency, and a lack of robustness [43]. The sampling efficiency and reliability could be significantly improved with a clipped objective function [43]. The clipped objective estimates a pessimistic lower bound on the value of the policy performance and consequently limits the size of gradient update steps, preventing drastic performance degradation. This algorithm is known as Proximal Policy Optimization (PPO). The PPO algorithm can handle continuous state and action spaces, offers ease of implementation and reliable performance, and therefore has been chosen for our application.

B. ENVIRONMENT AND OBSERVATION SPACE

We developed a custom training environment using OpenAI Gym. In order to ensure that the agent learns to plan comfortable paths for a wide range of scenarios, random road profiles were generated for training. In each episode, the agent is given a road profile that has a total length of L and consists of intermediate sectors of constant-curvature arcs. The initial and final sectors are straight paths while each of the remaining sectors has a curvature κ_i sampled randomly from the uniform distribution $\mathcal{U}_{[\kappa_{\min}, \kappa_{\max}]}$. The length of each

sector was obtained by partitioning the total length of the road into sectors, again in a manner to ensure that the lengths form a uniform distribution $\mathcal{U}_{[l_{\min}, l_{\max}]}$. The initial velocity is randomly sampled from a uniform distribution $\mathcal{U}_{[v_{\min}, v_{\max}]}$. The road was assumed to be a constant width throughout the entire section. The environment was assumed to be completely observable, and the state vector was defined as:

$$s = \left[\kappa_{0:n-1}, l_{0:n-1}, y_0, v_0\right] \tag{6}$$

where $\kappa_{0:n-1}$ is the array of curvatures of the road sectors, $l_{0:n-1}$ are the lengths of the respective sectors, and y_0 and v_0 are the initial lateral position and longitudinal velocity of the vehicle, respectively. Since the curvatures and lengths of the road profile and the vehicle velocity can take any value within the defined limits, the state space is continuous in nature. The state contains complete information about the vehicle and the road required for the agent to plan the trajectory. The states and observations spaces are normalized to lie between $[-1, 1]$, which ensured that the different quantities are scaled appropriately and the weights in the neural network are not skewed due to different orders of magnitude of the state variables.

C. MOTION DEFINITION AND ACTION SPACE

The vehicle motion was defined in terms of its position and velocity, and it is assumed that a path-following controller would be used to follow the defined trajectory. The position of the vehicle is defined with a lateral deviation with respect to the lane center, measured perpendicular to the normal driving direction and with the left-hand side being positive. To ensure smoothness in the planned trajectory, a cubic spline is utilized to approximate the velocity and position profiles. The reference position was described as a cubic function of the distance traveled along the centerline of the path. The reference position is calculated as the lateral deviation y from the centerline, given by the following equation

$$y_i(u) = a_{y,i}u^3 + b_{y,i}u^2 + c_{y,i}u + d_{y,i} \quad i = 0, \ldots, k-1 \tag{7}$$

where $u \in [0, 1]$ is a normalized distance parameter, 0 and 1 at the beginning and end of each sub-interval P_i of the spline respectively. $a_{y,i}$, $b_{y,i}$, $c_{y,i}$ and $d_{y,i}$ are the cubic spline coefficients for the *ith* polynomial P_i. k is the total number of cubic polynomials that compose the spline. The coefficients are calculated to satisfy the following boundary conditions

- The first derivative at the beginning and end of each polynomial is continuous

$$P_{i-1}^{(1)}(1) = P_i^{(1)}(0) \quad i = 1, \ldots, k-1 \tag{8}$$

- The second derivative at the beginning and end of each polynomial is continuous

$$P_{i-1}^{(2)}(1) = P_i^{(2)}(0) \quad i = 1, \ldots, k-1 \tag{9}$$

- At the start and end of the road, the first derivative is zero. This ensures an initial and final heading along the road direction, and zero initial and final longitudinal acceleration

$$P_0^{(1)}(0) = P_{k-1}^{(1)}(1) = 0 \tag{10}$$

A total of k control points or knots are distributed along the road to shape determine the shape of the trajectory. The velocity profile is calculated in a similar fashion where the knots are placed at the same positions as the spline used to calculate lateral positioning.

$$v_i(u) = a_{v,i}u^3 + b_{v,i}u^2 + c_{v,i}u + d_{v,i} \quad i = 1, \ldots, k-1 \tag{11}$$

Together, the predicted values at the control points for both position and velocity comprise the action space of the DRL agent. The action space of the agent is therefore given as follows

$$\begin{aligned} a = \big[&y_1(0), y_2(0) \ldots, y_{k-1}(0), y_{k-1}(1), \\ &v_1(0), v_2(0) \ldots, v_{k-1}(0), v_{k-1}(1)\big] \end{aligned} \tag{12}$$

Similar to the state and observation space, the action space was also normalized to lie between $[-1, 1]$. The bounds for normalization are decided based on the vehicle and lane width and speed limits. Considering a typical lane width of 3.3 m according to road design guidelines in the Netherlands, and a vehicle width of 2.1 m, the control knots for trajectory are limited to a maximum deviation of 0.5 m from the lane center to keep the vehicle within the lane boundaries. Inside built-up areas in the Netherlands, the most common speed limit is $50\,\mathrm{km\,h^{-1}}$, which gives the upper limit for velocity. The lower limit for velocity is set to $18\,\mathrm{km\,h^{-1}}$. The knot vectors along the length of the road are placed so as to obtain equal partitions of the total length of the road. A minimum length is enforced on each sector of constant curvature so that it contains at least one spline knot vector. This constraint ensured that for every corner the spline could accommodate a change in the direction corresponding to a change in curvature of the path.

D. REWARD FUNCTION

The agent is expected to learn to plan paths that minimize motion sickness while also maintaining reasonable travel time. Hence a reward function is designed to reflect these factors. First, a discomfort term D is defined as the integral of the squared planar accelerations over the entire duration of the motion:

$$D = \int_0^T \left(a_x^2 + a_y^2\right) dt \tag{13}$$

where a_x and a_y are the longitudinal and lateral accelerations of the vehicle, respectively, and T is the total travel time to traverse the planned trajectory. Our goal is to selectively minimize accelerations with the most significant contribution to the generation of motion sickness in passengers, and therefore the accelerations need to be weighted based on the frequency dependence of motion sickness. Models such as the SVC model [44], [45] can be used to approximate

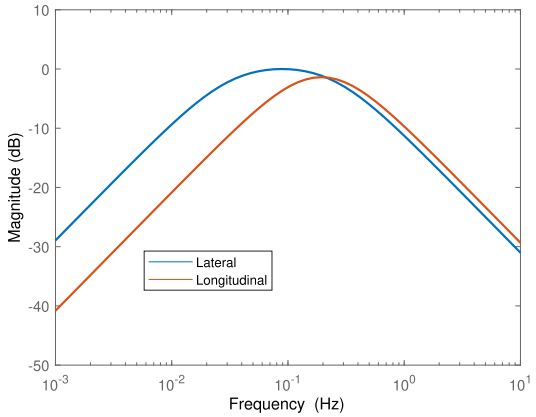

FIGURE 2. The bandpass filters used for weighting lateral and longitudinal accelerations to calculate the discomfort term.

the relation between acceleration frequencies and motion sickness. The SVC model can also account for rotational accelerations. However, the point-mass model used to calculate the accelerations cannot provide any information about rotations. In addition, the application of SVC model to weigh the accelerations would increase the computational complexity of the reward calculation. The frequency dependence of accelerations can also be modelled by frequency weighting filters, as proposed in [7], [8]. To further simplify the calculation of the reward for our purpose, 2nd order transfer functions were used to approximate these frequency weighting filters, and then applied to the accelerations prior to calculating the discomfort term. As can be seen from Fig. 1, for lateral oscillations, accelerations in the frequency range of 0.02 to 0.25 Hz have the most significant contribution toward inducing motion sickness in passengers, with the weighting independent of the frequency of excitation [8]. Incidence of nausea has been shown to peak around 0.2 Hz for longitudinal oscillations, with the frequency dependence dropping off with higher and lower frequencies [7]. To incorporate these findings into the discomfort term, two bandpass filters were adopted for weighting the lateral and longitudinal accelerations separately. The cut-off frequencies are 0.02 Hz and 0.25 Hz for lateral acceleration and 0.15 Hz and 0.25 Hz for longitudinal acceleration. Each bandpass filter is expressed by the following transfer function in continuous time:

$$BP(s) = K\frac{1}{\tau_1 s + 1}\frac{s}{\tau_2 s + 1} \tag{14}$$

where τ_1 and τ_2 are the time constants of the low and high pass filters respectively. To ensure that neither of the lateral or longitudinal acceleration is weighted preferentially, the gain K of the filter for longitudinal acceleration has been adjusted to obtain an equal area under the curve for the frequency range between 0 and 1 Hz. The resulting bandpass filters are shown in Fig. 2. The filters are then converted to state-space models and discretized in order to fit into the calculation for the reward function. A 30 s cooldown period is added to the end of the motion plan where zero acceleration input is given. This is to take into account the long-tail effect of the bandpass filter with slow dynamics where the effect of the previous acceleration input continues to be observed in the filter output. The penalty on the filter output during this period should be included in the reward.

The overall reward is formulated as a negative weighted sum of the discomfort D and the travel time T, with the latter preventing the agent from acquiring the behavior of driving at excessively low speed:

$$R = -(WT + D) \tag{15}$$

where W is a weighing factor and a larger W means travel time is considered with more importance, encouraging the agent to plan for traveling faster while sacrificing some motion comfort. To calculate the accelerations and travel time, the path is first discretized into stations spaced 1 m apart along the length of the road. At each station k, the velocity v_k, and the waypoint's lateral deviation y_k, are determined using the respective spline functions. Given the coordinates of a station and its normal direction, the corresponding waypoint can be located in the global frame. Consequently, it is possible to find the distance between two consecutive waypoints and the curvature, enabling the calculation of the reward:

$$\Delta T_k = 2d_k/(v_{k+1} + v_k) \tag{16}$$

$$a_{x,k} = \left(v_{k+1}^2 - v_k^2\right)/2d_k \tag{17}$$

$$a_{y,k} = \kappa_k\left(v_k + a_{x,k}\Delta T_k\right)^2 \tag{18}$$

$$\Delta D_k = \left(a_{x,k}^2 + a_{y,k}^2\right)\Delta T_k \tag{19}$$

$$R = \sum_{k=1}^{N-1}(W\Delta T_k + \Delta D_k) \tag{20}$$

The calculation is purely based on kinematics so the training requires minimal computation. It also leads to a more general trajectory planner which can be implemented on a range of vehicles, independent of individual vehicle parameters. In order to quickly guide the agent away from planning excessively aggressive motions at the beginning of the training, we impose a conditional penalty that will be given to the agent when the combined planar acceleration exceeds $1g$ at any timestep.

E. DRL AGENT TRAINING

The training consisted of single-step episodes. In each episode, the agent first receives the initial information s_0 from the environment as defined in (6). Based on this, the agent then outputs knot vectors for the entire path and receives the corresponding reward. As described in Section I-A, the PPO algorithm is used to train the agent. We adopt the standard PPO implementation from the Stable Baselines3 library [46]. The hyperparameters listed in Table 1 are chosen after being optimized with Optuna [47]. All training and testing were performed on a laptop PC with an Intel i5-10210U CPU plus an NVIDIA GeForce MX250 GPU.

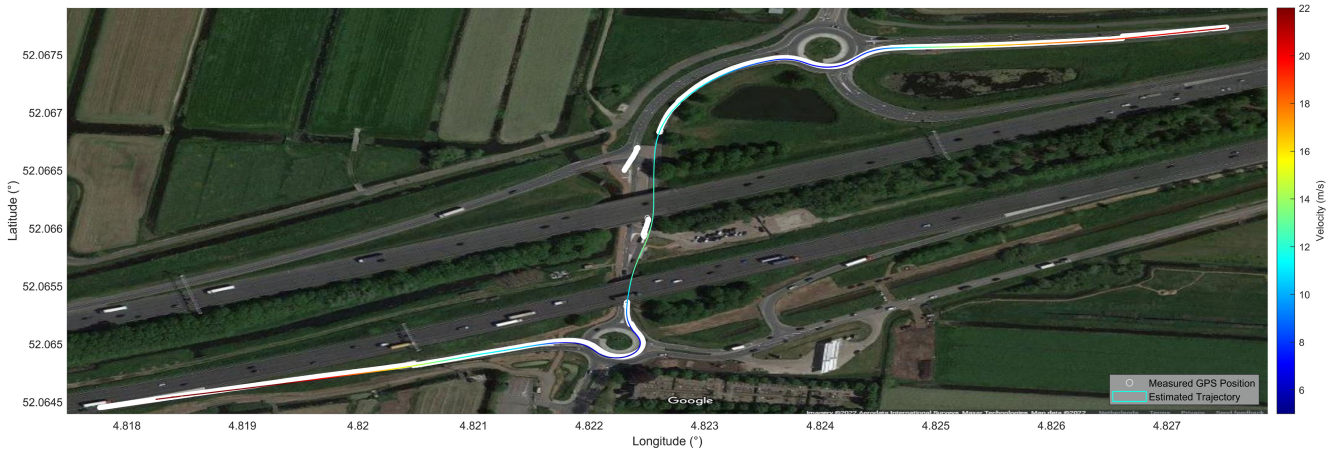

FIGURE 3. A satellite view of the road section including the two roundabouts has been depicted. The vehicle positions as measured by the GPS sensor for one of the test runs have been shown in white, along with the estimated trajectory as obtained from the Kalman filter. As can be seen, the trajectory is reconstructed and the discontinuities in the data have been eliminated.

TABLE 1. Hyperparameters for the PPO algorithm and their corresponding values.

Hyperparameter	Value
Learning rate	0.001
Discount factor	0.99
Steps before update	2048
Clip range	0.2
Batch size	64

Given the limited computational resources, the optimization of hyperparameters is only preliminary. Hence this method may hold more potential than what is presented in this paper.

F. OPTIMIZATION-BASED PLANNER

To determine the upper limit of agent performance in the designed custom environment, we developed an optimization-based spline planner that has the same amount of freedom as the DRL agent:

$$\begin{aligned} \max_{a} \quad & R(s = s_0, a) \\ \text{where:} \quad & a = \left[y_{1:k-1}, v_{1:k-1}\right] \\ \text{s.t.} \quad & y_{\min} \leq y_i \leq y_{\max} \\ & v_{\min} \leq v_i \leq v_{\max} \end{aligned} \tag{21}$$

where a consists of the control points as defined in (12), R is the reward function given in (15), and $s = s_0$ is the initial state of the environment. The constraints on the vehicle velocities and lateral positions are identical to those imposed on the DRL agent. The Sequential Least SQuares Programming (SLSQP) algorithm from the SciPy library is used to solve the optimization problem above. It should be clarified that the trajectory generated by this optimization-based planner is not the best possible motion plan for a given road profile. Instead, it is a measure of the best performance the DRL agent can be expected to achieve in a spline-based planning scheme.

III. HUMAN BASELINE PERFORMANCE

In addition to comparing the performance of DRL agents with regard to targeting motion sickness, it is also interesting to compare the frequency content of the accelerations generated by the DRL planner to the accelerations typically generated by human drivers. In this study, a human performance baseline has been established by measuring an instrumented vehicle's position, velocity, and acceleration values as it is driven by volunteers on a chosen road section. The chosen road section is located in the Netherlands. It begins at the exit ramp of motorway A12 (52.064°N, 4.818°E) and ends at the distributor road N420 (52.068°N, 4.828°E, see Fig. 3). When driving through the road section, the vehicle has to navigate through two roundabouts connected by a path consisting of consecutive turns. The section is considered sufficiently demanding on the driver's capability in planning and controlling vehicle motion.

To evaluate the performance of the DRL agent, we focused only on the trajectory followed on each of the roundabouts, as the remaining portions of the path consisted mostly of long straight sections which would substantially increase the number of control points required, and subsequently the training time. The chosen sections of the roundabouts spanned 134 m in length, including the straight parts at entry and exit. The roundabouts will be referred to as RB1 and RB2 in sequential order respectively. The motion profile as output by the agent was then evaluated in a high-fidelity IPG CarMaker environment with a multi-body vehicle model, as the custom training environment only used a point-mass model which would not be directly comparable to actual vehicle accelerations. Section III-A describes the experimental setup, Section III-B introduces the data processing method, and Section III-C details the controllers and vehicle model used for evaluating the motion planned by the DRL agent.

FIGURE 4. The test vehicle, a Hyundai Tucson (top), equipped with a double antenna GPS (bottom left) and IMU system (bottom right) for measuring the driving performance of human drivers.

A. DRIVING EXPERIMENT

An instrumented vehicle has been driven by volunteers to provide a preliminary human driving performance baseline. The test vehicle was a Hyundai Tucson with an automatic transmission to reduce the workload of the drivers and minimize the impact of gear shifting on comfort. To record the vehicle trajectory, we installed a high-precision 100 Hz dual-antenna GPS sensor from Racelogic in combination with an inertial measurement unit (IMU) from XSENS (see Fig. 4). A total of 6 volunteers were recruited with an average age of 27.5 and an average driving experience of 7.2 years counting from the date of receiving the first driving license. The volunteers were informed about the purpose of the experiment and tried to drive the test route smoothly and swiftly. Each volunteer was given two attempts and the one with lesser interaction with road users is chosen. The trials were conducted in the evening between 19:30-21:00 when the traffic is less busy in the area. This improves the chance of capturing an unobstructed test run.

B. PROCESSING OF EXPERIMENT DATA

The test runs resulted in a collection of vehicle position, velocity, and acceleration profiles. The motion profiles were portioned for the two roundabouts of interest using the GPS position information. To establish the human baseline performance, however, the measurement data could not be used directly. to ascertain the start and end of the trajectories due to errors resulting from measurement noise. To produce reasonable driving data representative of actual vehicle trajectories, a Kalman filter was implemented. A kinematic model is used here with the state given by vehicle position and velocity and the accelerations as inputs:

$$\begin{aligned} p_k &= p_{k-1} + v_{k-1}\Delta t + \frac{1}{2}a_{k-1}\Delta t^2 \\ v_k &= v_{k-1} + a_{k-1}\Delta t \end{aligned} \tag{22}$$

where p_k, v_k and a_k are the position, velocity, and acceleration vectors in the global coordinate system at time step k and Δt is the sampling time. The state is expressed as $x = [p^T v^T]^T$. The acceleration vector is taken as the IMU measurements but converted to global coordinates using yaw orientation. The process noise is assumed to be a result of the noise in IMU measurements. The acceleration data has been low-pass filtered beforehand to remove the high-frequency noise using FFT-based techniques. We assume the system to be fully observable with all state measurements z_k available from the GPS sensor:

$$z_k = \begin{bmatrix} p \\ v \end{bmatrix} + \eta_k \quad \eta_k \sim \mathcal{N}(0, R) \tag{23}$$

Using the system dynamics given by (22) and the measurement model given by (23), and assuming normally distributed Gaussian noise for all sensors, the Kalman filter could obtain a better estimation of vehicle trajectory and velocity profile over the test run. The desired portions of the trajectory of RB1 and RB2 were extracted using the estimated vehicle position. An example of the measured and estimated trajectory and velocity profiles has been shown in Fig. 3.

C. SIMULATION SETUP

The motion profiles generated by the DRL agent were based on a point-pass model, so in order to have a fair comparison with the human drivers, the trajectories were evaluated in a virtual IPG CarMaker environment. The vehicle model used was comparable in dimensions and kerbweight to the actual test vehicle in order to have as close a comparison as possible. To track the reference trajectories generated by the motion planner, a simple Stanley controller was implemented as follows

$$\delta = (\psi_r - \psi) + \operatorname{atan}\frac{k_{\text{steer}}(y_r - y)}{v} \tag{24}$$

where δ and ψ are the steering input and heading of the vehicle respectively, y_r is the reference position, while k_{steer} is a parameter that decides the aggressiveness of the controller. The throttle percentage P_T or brake percentage P_B is decided as a weighted sum of reference forward acceleration $a_{x,r}$ with the error in velocity, scaled by a factor k_{drive} or k_{brake} depending on whether the vehicle is desired to be accelerated or decelerated.

$$P_T = k_{\text{drive}}\big(a_{x,r} + k_{\text{speed}}(v_r - v)\big) \times 100\% \tag{25}$$

$$P_B = k_{\text{brake}}\big(a_{x,r} + k_{\text{speed}}(v_r - v)\big) \times 100\% \tag{26}$$

The parameters of the Stanley controller were tuned to achieve a RMS tracking error of less than 0.1m for all the planners, for all motion plans with different weights W. The focus of the research was on developing the DRL motion planner, and investigating the path tracking performance was not a primary goal of the study. In principle, other path-following techniques can also be used [48].

IV. RESULTS

The results have been divided into three parts: Section IV-A goes into the frequency analysis of two DRL agents trained in a simple environment, one minimizing motion sickness and the other optimizing general motion comfort described by total acceleration energy. The agents have also been compared to optimal planners. The Section IV-B establishes the human baseline performance for the roundabout scenarios and compares the performance of the trained DRL agent with the optimal planner as well as human drivers. Furthermore, we report on the computation efficiency of the proposed method in Section IV-C. To measure the effectiveness of the proposed system, we have used Key Performance Indicators, such as the discomfort term, the RMS accelerations, travel time and computation time for each of the motion plans.

A. FREQUENCY DOMAIN PERFORMANCE

For the purpose of investigating whether the DRL agent is able to target the low-frequency acceleration component, a simple environment was used. The road length was assumed to be 100 m, with a single turn. The trajectory was defined with splines controlled by $k = 5$ control points. Two agents were trained in the same environment, with the only difference in their respective reward functions. Agent A was trained on a discomfort term calculated without frequency-weighted accelerations, while agent B was trained using a reward function incorporating the bandpass filters as described in Section II-D. The accelerations outside the cut-off frequencies are attenuated by the bandpass filters, and so the frequency-weighted discomfort term is generally lower in value for comparable travel times. To compensate for this and ensure similar travel times for both agents, $W = 0.6$ and $W = 1$ were used for agents A and B respectively. Both agents were trained for 1M steps.

The planned trajectories of agents A and B for a randomly generated scenario from the training environment have been shown in Figs. 5 and 6 respectively. In the test case, the vehicle is initialized with a randomly generated speed of 7.25m/s and has to traverse through a sharp left turn. As can be seen from these figures, both agents learn to accelerate in the straight sections of the road and decelerate on approaching the corner. The spatial plans also are close to a path that would be intuitively expected to be the most comfortable around the corner, with the vehicle entering from the outside edge, moving close to the apex, and then exiting toward the outside edge of the corner. Both agents learn to utilize the complete limits of the available lateral deviation. In the particular test case shown here, the vehicle velocity range is not completely used, however, that is in the interest of producing lower vehicle accelerations. The peak lateral accelerations in both cases are around 3.5m/s^2.

For the particular case being analyzed, the frequency-weighted discomfort term is 6.5% lower for the trajectory planned by agent B, with the same travel time as agent A. The drop in the discomfort term mainly arises from the lower lateral accelerations in the motion plan generated by agent B. Meanwhile, travel time is unaffected due to the higher longitudinal accelerations from agent B. The vehicle was commanded to make more aggressive speed changes before and after the corner. This is reflected in the speed range as plan B has a minimum and maximum velocity of 6.4m/s and 11.5m/s respectively, as opposed to 7.0m/s and 9.7m/s in motion plan A. The preferential lowering of lateral accelerations by agent B can be attributed to the frequency filter. The longitudinal accelerations have a narrower bandpass filter as opposed to lateral accelerations and hence are attenuated more. In addition, the longitudinal frequency filter has a lower peak gain, to have the same area under the curve over the frequency range of 0 to 1 Hz. The agent learns to increase accelerations beyond the cut-off frequencies, and target the frequencies of interest.

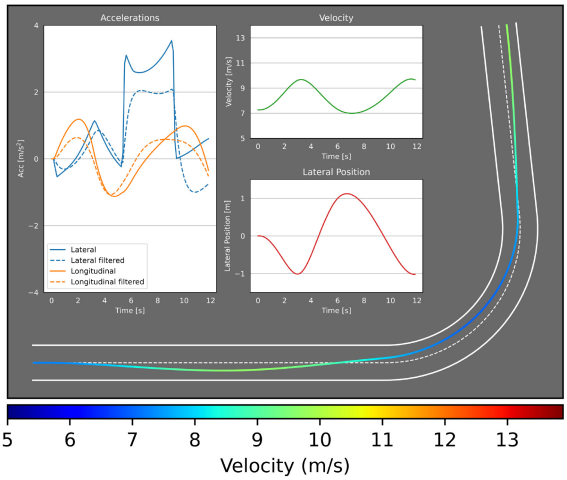

FIGURE 5. Trajectory as planned by agent A, trained without filtered accelerations. The subfigures depict the vehicle accelerations, velocity and positions from the motion plan. The filtered accelerations have also been shown to depict the frequency content of the accelerations causing nauseogenicity. The entire path takes 11.84 s to navigate.

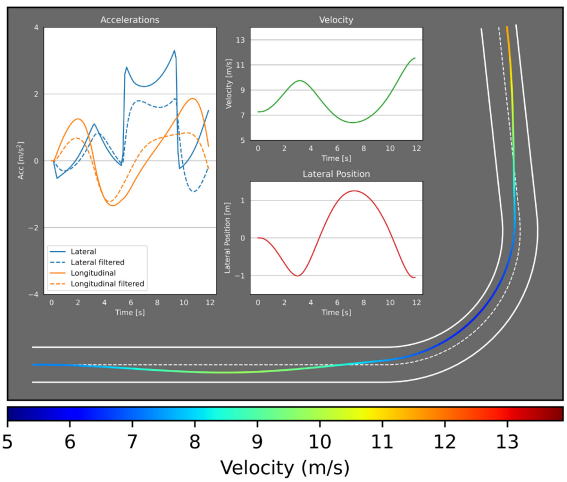

FIGURE 6. Trajectory as planned by agent B, trained with filtered accelerations. The subfigures depict the vehicle accelerations, velocity and positions from the motion plan. The filtered accelerations have also been shown to depict the frequency content of the accelerations causing nauseogenicity. The entire path takes 11.87 s to navigate.

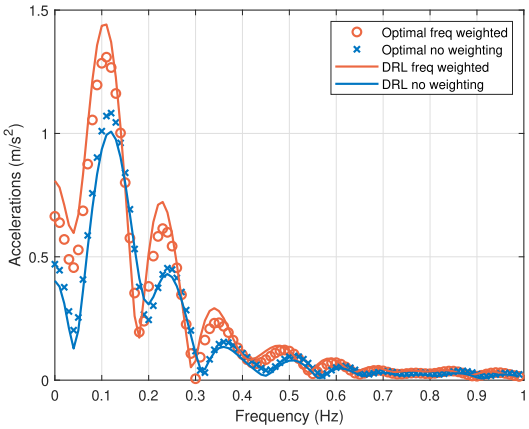

FIGURE 7. Frequency content comparison of the longitudinal accelerations in the motion plans A and B. The line plots represent the DRL agents, while the scatter plots represent the optimal planners.

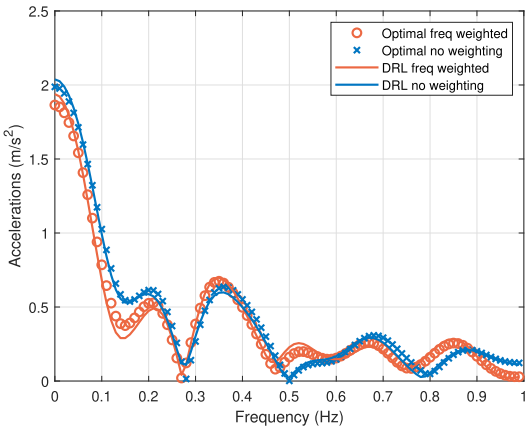

FIGURE 8. Frequency content comparison of the lateral accelerations in the motion plans A and B. The purple and yellow line plots represent the DRL agents A and B respectively, while the scatter plots represent the optimal planners.

To analyze the frequency content of the accelerations, the non-uniform fast Fourier transform has been shown in Figs. 7 and 8. To further provide an insight into the performance of the proposed method, the comparison with the optimization-based planner detailed in Section II-F has also been included. As expected from the acceleration values, it can be seen that the peak amplitudes of lateral accelerations are lower in motion plan B than in A. Throughout the frequency band of 0.0315 to 0.25 Hz, the amplitudes are significantly lower in motion plan B. This does lead to higher peaks between 0.3 and 0.9 Hz, but that is expected and desirable behavior in our case. With DRL agent B, it can be seen that the energy is transferred from lateral to longitudinal accelerations, with significantly higher peaks compared to agent A. However, two important points need to be noted. Near the peak nauseogenic frequency of 0.2 Hz, motion plan B has a lower minimum as compared to agent A. Also, lateral accelerations have significantly higher amplitudes throughout the relevant frequency spectrum as compared to longitudinal accelerations, and so agent B learns to minimize low-frequency

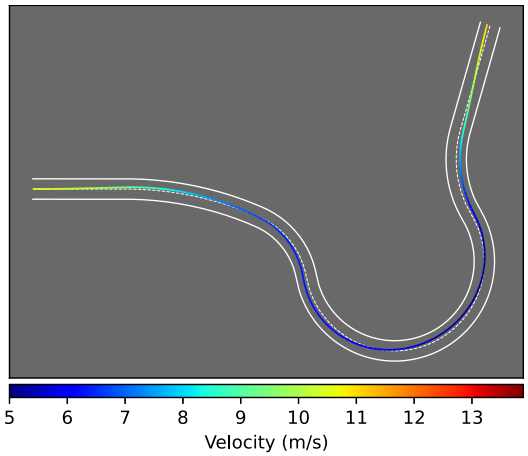

FIGURE 9. Trajectory as planned by DRL agent for RB1. The weight on time for the agent is $W = 8$. The planned path takes 19.04s to navigate.

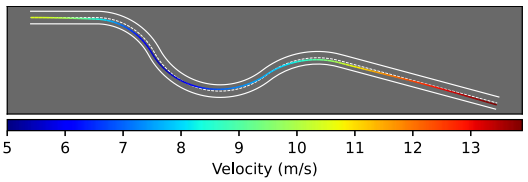

FIGURE 10. Trajectory as planned by DRL agent for RB2. The weight on time for the agent is $W = 8$. The entire path takes 15.52s to navigate.

lateral accelerations at the cost of higher longitudinal accelerations. A previous study [16] found a higher correlation of lateral accelerations with motion sickness in passengers as compared to fore-aft accelerations, so the behavior learned by agent B works in this direction.

The above analysis was performed for a single randomly generated test case. However, to have a holistic idea of the performance of the agents, the average frequency-weighted discomfort value and the travel times over 10,000 episodes were calculated. The frequency-weighted discomfort value D for agent A was 19.61, while the for agent B was 17.73, which is a drop of 9.6%, quite significant for the relatively short and simple road profiles under consideration. The average travel times for the same were 9.78 and 9.87 s respectively, which is a difference of less than 1%, and therefore can be considered comparable.

B. PERFORMANCE IN TIME AND COMFORT

As described previously in Section III-A, the two roundabouts from the human driving data, RB1 and RB2, are of primary interest. Therefore, the training has been done on a 6-section road with a total length of 134 m. A variety of weights on travel time, ranging from $W = 4$ to $W = 16$, has been used in the training to produce an inclusive set of possible trajectories representative of different driving styles. All agents were trained for 1.5M steps. The example trajectories for RB1 and RB2 from the agent trained with $W = 8$ are shown in Figs. 9 and 10, respectively. RB1 is a slower roundabout with smaller radii of curvature, while RB2 could allow a higher speed as it bends less sharply.

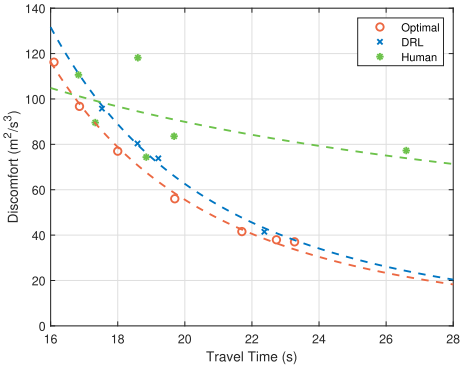

FIGURE 11. Comparison of frequency weighted discomfort values and the travel times for the DRL agent and the optimal planner, for roundabout 1.

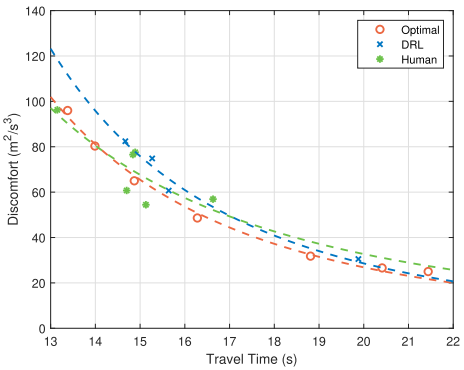

FIGURE 12. Comparison of frequency weighted discomfort values and the travel times for the DRL agent and the optimal planner, for roundabout 2.

The performance of the DRL agent has again been compared with the optimization-based planner described in Section II-F. In addition, the nauseogenicity of planned motion has been compared with the driving comfort of human drivers measured through the experimental setup as described in Section III-A. It is commonly recognized that time efficiency and comfort are conflicting factors. Hence to fairly compare the performance, the discomfort values have been plotted against travel time in Figs. 11 and 12. It is visible that the discomfort value increases as travel time becomes shorter. This is effectively a Pareto front for a multi-objective optimization problem where time and comfort are optimized simultaneously. We fitted the individual points on the time-discomfort plane into a curve of the following form in order to compare the discomfort values at a given travel time or vice versa:

$$y = ax^b + c \tag{27}$$

For RB1, the performance of the learning-based planner is between 10.9% to 12.9% below the optimization-based planner. The performance difference is between 6.2% and 14.2% for RB2. The worse performance over RB2 is due to the increased difficulty of navigating the turn at higher speeds leading to higher accelerations, particularly at lower travel times. In both cases, the DRL-based planner has higher discomfort values than the optimization-based counterpart.

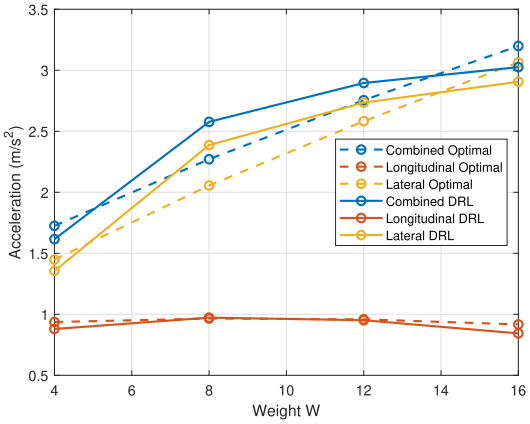

FIGURE 13. Change in RMS acceleration values with varying weight *W*.

As anticipated, the discomfort values are significantly lower in RB2 due to the mellower turns it involves.

These figures also provide a rough description of human driving performance. In both roundabouts, there is significant variation in the performance of human drivers. The travel time ranges from 16.8 to 26.6 s for RB1, and from 13.1 to 16.6 s for RB2. The frequency-weighted discomfort rating varies from 74.3 to 118.1 for RB1 and 54.4 to 96.3 for RB2. The general trend complies with the observations of the motion planners that a lower travel time leads to more discomfort. Meanwhile, some drivers outperformed the others with the location of their test run closer to the lower-left corner on the time-discomfort plane. The DRL-based planner is comparable to the better-performing drivers in RB1 and slightly underperforms the best test run by 11.3%. It can be seen that the discomfort rating of the DRL agents shows a sharper increasing trend than the human drivers with lower travel times, and this can be attributed to the larger modeling errors associated with the point mass model on approaching higher vehicle speeds and acceleration values. The RB2 scenario, on the other hand, does highlight some limitations in the design of the training environment itself. A total of 3 test runs exhibit a lower discomfort value than the optimization-based planner. Apart from the potential mismatch in the road geometry, a major contributing factor to this is the use of splines. It limits the freedom of action of the agent and the optimization-based planner formulated in the same framework. The second reason could be the higher modeling errors with using a point-mass model for trajectories, but this effect will only be pronounced with faster travel times and higher acceleration values.

We further present the trend of root-mean-square (RMS) value of lateral and longitudinal accelerations obtained with varying weights in Fig. 13. The lateral acceleration behaves as expected, showing an upward trend when increasing *W*. Interestingly, the RMS longitudinal acceleration only shows minimal changes and even drops when *W* increases from 12 to 16. This is because when the vehicle approaches the roundabout with a moderate initial velocity, the need for a

shorter travel time is interpreted into a smaller change in velocity as the planner commands to take the corners at a higher speed. The rise in lateral acceleration is tolerated as it does not overshadow the reward of saving time due to a larger W being used. The DRL agent also shows an increasing trend of RMS lateral accelerations, and nearly constant RMS longitudinal accelerations, with a slight drop with increasing weight W. However, the RMS values of the DRL agent cannot be directly compared here with the optimal planner, since their travel times learnt for each weight W are different.

C. COMPUTATION TIME

The metric in which the DRL agent comprehensively outperforms the optimization-based planner is the computational time. Once completed with training, the DRL-based planner can generate a sub-optimal motion plan in less than 2 ms. The computation time is rather consistent compared to an optimization-based approach where the number of iterations before convergence is hardly predictable. The latter takes an average of 5 s to compute the optimal motion plan. The improvement in computation time is by three orders of magnitude while the loss of performance is less than 15%. An essential upside of the DRL-based planner is, even with a more complex training environment, the added complexity will not be reflected in online computation.

V. CONCLUSION

In this research, a novel method of minimizing motion sickness in motion planning with Deep Reinforcement Learning has been presented. The nauseogenicity of the planned trajectory is evaluated by frequency-weighted accelerations according to motion sickness models in the literature. It has been shown that by including such a discomfort term in the reward function, the DRL agent is able to learn to target the frequency range that is primarily responsible for motion sickness. The frequency-weighted discomfort is reduced by 9.6% on average, compared to an agent trained without the frequency sensitivity in accelerations. The improvement is a result of shifting the acceleration energy away from the frequency range of interest.

The DRL agent has been trained using a variety of weights to represent different preferences between comfort and time efficiency. The learning-based planning performance has been evaluated on two roundabout scenarios derived from real road sections in the Netherlands. An optimization-based spline planner has been developed for comparison purposes, next to a human performance baseline established experiments where volunteers drove an instrumented vehicle through the actual road sections. The DRL-based planner outperforms most human drivers in one scenario while providing comparable performance in the other. Compared to the optimization-based planner, the proposed method shows a performance deficit in the range of 10-15% but nevertheless cuts computation time significantly by three orders of magnitude.

It has been shown that for scenarios with a reasonable complexity level, DRL can be used in motion planning for reducing accelerations in the most nauseogenic frequency bands. The performance of the method is however limited by the design of the environment, for example, by limiting the planning freedom to placing control points of a spline. To be applied in more complex environments, training needs to be done with a larger state and action space. For more complex driving situations such as multi-lane highways, highway ramps, or intersections, the performance of the method still needs verification. In addition, the work needs to be extended to dynamic environments with multiple actors, where the comfort benefits of the proposed method remain to be seen. The state representation consists of a constant length of the road, with a fixed number of curvature changes. While this representation can be used to encode road information for a large number of possible road profiles, it is still not general enough to accommodate road profiles with a higher number of curvature changes within the defined length. The use of Recurrent Neural Networks (RNNs) could be investigated for incorporating a variable state space so as to have a more general representation of the road profile. RNNs can deal with variable sizes of the state space so they can be used to represent a varying number of road sections depending on the number of curvature changes present in the road profile. The motion planning approach presented in this work is a proof-of-concept of the applicability of DRL to acceleration frequency shaping and motion sickness minimization, and not a mature algorithm which can be used in its present form. The approach presented is only applicable to the environment in which the agent was trained, and the performance of the DRL agent in a variety of environments remains to be investigated.

In this study, we have encountered challenges that are commonly mentioned in DRL-related studies. The training time increases exponentially along with the dimensionality of the state and action space. Given the basic computational power, the training process is too time-consuming to optimize the hyperparameters, or consequently, to obtain more satisfactory performance. In addition, it is unclear whether the reduction in frequency-weighted accelerations would translate to real-world comfort improvements for passengers. It is possible that focusing on motion sickness mitigation might lead to degraded acceleration comfort in general. The benefits of the method may only be appreciable in longer journeys through curvy roads. The subjective evaluation of the proposed method could be evaluated through simulator-based or on-road experiments.

REFERENCES

[1] B. Pfleging, M. Rang, and N. Broy, "Investigating user needs for non-driving-related activities during automated driving," in *Proc. 15th Int. Conf. Mobile Ubiquitous Multimedia*, 2016, pp. 91–99.
[2] C. Diels and J. E. Bos, "Self-driving carsickness," *Appl. Ergonom.*, vol. 53, pp. 374–382, Mar. 2016.
[3] A. Rolnick and R. E. Lubow, "Why is the driver rarely motion sick? The role of controllability in motion sickness," *Ergonomics*, vol. 34, pp. 867–879, Jul. 1991.

[4] G. Bertolini and D. Straumann, "Moving in a moving world: A review on vestibular motion sickness," *Front. Neurol.*, vol. 7, p. 14, Feb. 2016.

[5] J. T. Reason, "Motion sickness adaptation: A neural mismatch model," *J. Roy. Soc. Med.*, vol. 71, no. 11, pp. 819–829, 1978.

[6] J. F. O'Hanlon and M. E. McCauley, "Motion sickness incidence as a function of the frequency and acceleration of vertical sinusoidal motion," *Aerosp. Med.*, vol. 45, pp. 366–369, Apr. 1974.

[7] J. F. Golding, A. G. Mueller, and M. A. Gresty, "A motion sickness maximum around the 0.2 Hz frequency range of horizontal translational oscillation," *Aviat. Space Environ. Med.*, vol. 72, pp. 188–192, Mar. 2001.

[8] B. E. Donohew and M. J. Griffin, "Motion sickness: Effect of the frequency of lateral oscillation," *Aviat. Space Environ. Med.*, vol. 75, pp. 649–656, Aug. 2004.

[9] S. Salter, S. Kanarachos, C. Diels, P. Herriotts, and C. D. Thake, "Motion sickness in automated vehicles with forward and rearward facing seating orientations," *Appl. Ergonom.*, vol. 78, pp. 54–61, Jul. 2019.

[10] O. X. Kuiper, J. E. Bos, C. Diels, and E. A. Schmidt, "Knowing what's coming: Anticipatory audio cues can mitigate motion sickness," *Appl. Ergonom.*, vol. 85, May 2020, Art. no. 103068.

[11] M. Miksch, M. Steiner, M. Miksch, and A. Meschtscherjakov, "Motion sickness prevention system (MSPS): Reading between the lines," in *Proc. 8th Int. Conf. Autom. User Interfaces Interact. Veh. Appl.*, 2016, pp. 147–152.

[12] G. Papaioannou, J. Jerrelind, L. Drugge, and B. Shyrokau, "Assessment of optimal passive suspensions regarding motion sickness mitigation in different road profiles and sitting conditions," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, 2021, pp. 3896–3902.

[13] Y. Zheng, B. Shyrokau, T. Keviczky, M. Al Sakka, and M. Dhaens, "Curve tilting with nonlinear model predictive control for enhancing motion comfort," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 4, pp. 1538–1549, Jul. 2022.

[14] A. Dettmann et al., "Comfort or not? Automated driving style and user characteristics causing human discomfort in automated driving," *Int. J. Human–Comput. Interact.*, vol. 37, no. 4, pp. 331–339, 2021.

[15] R. J. Koppa and G. G. Hayes, "Driver inputs during emergency or extreme vehicle maneuvers," *Human Factors*, vol. 18, pp. 361–370, Aug. 1976.

[16] M. Turner and M. J. Griffin, "Motion sickness in public road transport: The relative importance of motion, vision and individual differences," *Brit. J. Psychol.*, vol. 90, no. 4, pp. 519–530, 1999.

[17] L. Labakhua, U. Nunes, R. Rodrigues, and F. S. Leite, "Smooth trajectory planning for fully automated passengers vehicles: Spline and clothoid based methods and its simulation," in *Informatics in Control Automation and Robotics*. Berlin, Germany: Springer, 2008, pp. 169–182.

[18] M. McNaughton, C. Urmson, J. M. Dolan, and J.-W. Lee, "Motion planning for autonomous driving with a conformal spatiotemporal lattice," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2011, pp. 4889–4895.

[19] S. Gim, L. Adouane, S. Lee, and J.-P. Dérutin, "Clothoids composition method for smooth path generation of car-like vehicle navigation," *J. Intell. Robot. Syst.*, vol. 88, pp. 129–146, Mar. 2017.

[20] R. Lattarulo, E. Martí, M. Marcano, J. Matute, and J. Pérez, "A speed planner approach based on Bézier curves using vehicle dynamic constrains and passengers comfort," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2018, pp. 1–5.

[21] Y. Zheng, B. Shyrokau, and T. Keviczky, "Comfort and time efficiency: A roundabout case study," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, 2021, pp. 3877–3883.

[22] Y. Zheng, B. Shyrokau, and T. Keviczky, "3DOP: Comfort-oriented motion planning for automated vehicles with active suspensions," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2022, pp. 390–395.

[23] M. Diachuk and S. M. Easa, "Motion planning for autonomous vehicles based on sequential optimization," *Vehicles*, vol. 4, no. 2, pp. 344–374, 2022.

[24] M. Mischinger, M. Rudigier, P. Wimmer, and A. Kerschbaumer, "Towards comfort-optimal trajectory planning and control," *Veh. Syst. Dyn.*, vol. 57, no. 8, pp. 1108–1125, 2019.

[25] A. Artuñedo, J. Villagra, and J. Godoy, "Jerk-limited time-optimal speed planning for arbitrary paths," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8194–8208, Jul. 2022.

[26] I. Bae, J. Moon, and J. Seo, "Toward a comfortable driving experience for a self-driving shuttle bus," *Electronics*, vol. 8, no. 9, p. 943, 2019.

[27] A. Genser, R. Spielhofer, P. Nitsche, and A. Kouvelas, "Ride comfort assessment for automated vehicles utilizing a road surface model and Monte Carlo simulations," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 10, pp. 1316–1334, 2022.

[28] Z. Li, I. V. Kolmanovsky, E. M. Atkins, J. Lu, D. P. Filev, and Y. Bai, "Road disturbance estimation and cloud-aided comfort-based route planning," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3879–3891, Nov. 2017.

[29] Z. Htike, G. Papaioannou, E. Siampis, E. Velenis, and S. Longo, "Fundamentals of motion planning for mitigating motion sickness in automated vehicles," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 2375–2384, Mar. 2022.

[30] D. Li and J. Hu, "Mitigating motion sickness in automated vehicles with frequency-shaping approach to motion planning," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7714–7720, Oct. 2021.

[31] R. Ukita, Y. Okafuji, and T. Wada, "A simulation study on lane-change control of automated vehicles to reduce motion sickness based on a computational mode," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2020, pp. 1745–1750.

[32] S. Aradi, "Survey of deep reinforcement learning for motion planning of autonomous vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 740–759, Feb. 2022.

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

[34] P. Wolf, K. Kurzer, T. Wingert, F. Kuhnt, and J. M. Zollner, "Adaptive behavior generation for autonomous driving using deep reinforcement learning with compact semantic states," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2018, pp. 993–1000.

[35] C.-J. Hoel, K. Wolff, and L. Laine, "Automated speed and lane change decision making using deep reinforcement learning," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 2148–2155.

[36] P. Wang and C.-Y. Chan, "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2017, pp. 1–6.

[37] Á. Fehér, S. Aradi, F. Hegedüs, T. Bécsi, and P. Gáspár, "Hybrid DDPG approach for vehicle motion planning," in *Proc. 16th Int. Conf. Inform. Control Autom. Robot. (ICINCO)*, vol. 1, 2019, pp. 422–429.

[38] N. A. Spielberg, M. Templer, J. Subosits, and J. C. Gerdes, "Learning policies for automated racing using vehicle model gradients," *IEEE Open J. Intell. Transp. Syst.*, vol. 4, pp. 130–142, 2023.

[39] C.-J. Hoel, K. Driggs-Campbell, K. Wolff, L. Laine, and M. J. Kochenderfer, "Combining planning and deep reinforcement learning in tactical decision making for autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 5, no. 2, pp. 294–305, Jun. 2020.

[40] P. Wang and C.-Y. Chan, "Autonomous ramp merge maneuver based on reinforcement learning with continuous action space," 2018, *arXiv:1803.09203*.

[41] C. Paxton, V. Raman, G. D. Hager, and M. Kobilarov, "Combining neural networks and tree search for task and motion planning in challenging environments," 2017, *arXiv:1703.07887*.

[42] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, pp. 229–256, May 1992.

[43] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.

[44] J. E. Bos and W. Bles, "Modelling motion sickness and subjective vertical mismatch detailed for vertical motions," *Brain Res. Bull.*, vol. 47, no. 5, pp. 537–542, 1998.

[45] N. Kamiji, Y. Kurata, T. Wada, and S. Doi, "Modeling and validation of carsickness mechanism," in *Proc. SICE Annu. Conf.*, 2007, pp. 1138–1143.

[46] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable reinforcement learning implementations," *J. Mach. Learn. Res.*, vol. 22, pp. 1–8, Nov. 2021.

[47] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2019, pp. 2623–2631.

[48] Z. Lu, B. Shyrokau, B. Boulkroune, S. van Aalst, and R. Happee, "Performance benchmark of state-of-the-art lateral path-following controllers," in *Proc. IEEE 15th Int. Workshop Adv. Motion Control (AMC)*, 2018, pp. 541–546.