IEEE Open Journal of
Intelligent Transportation
Systems

Received 19 December 2023; accepted 23 April 2024. Date of publication 29 April 2024; date of current version 17 May 2024.

Digital Object Identifier 10.1109/OJITS.2024.3395149

Adaptive Cruise Control Utilizing Noisy Multi-Leader Measurements: A Learning-Based Approach

YING-CHUAN NI [1,2], VICTOR L. KNOOP [2], JULIAN F. P. KOOIJ [3] (Member, IEEE), AND BART VAN AREM [2] (Senior Member, IEEE)

[1]Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zürich, 8093 Zürich, Switzerland

[2]Department of Transport and Planning, Delft University of Technology, 2628 CN Delft, The Netherlands

[3]Intelligent Vehicles Group, Department of Cognitive Robotics, Delft University of Technology, 2628 CD Delft, The Netherlands

CORRESPONDING AUTHOR: Y.-C. NI (e-mail: ying-chuan.ni@ivt.baug.ethz.ch)

ABSTRACT A substantial number of vehicles nowadays are equipped with adaptive cruise control (ACC), which adjusts the vehicle speed automatically. However, experiments have found that commercial ACC systems which only detect the direct leader amplify the propagating disturbances in the platoon. This can cause severe traffic congestion when the number of ACC-equipped vehicles increases. Therefore, an ACC system which also considers the second leader further downstream is required. Such a system enables the vehicle to achieve multi-anticipation and hence ensure better platoon stability. Nevertheless, measurements collected from the second leader may be comparatively inaccurate given the limitations of current state-of-the-art sensor technology. This study adopts deep reinforcement learning to develop ACC controllers that besides the input from the first leader exploits the additional information obtained from the second leader, albeit noisy. The simulation experiment demonstrates that even under the influence of noisy measurements, the multi-leader ACC platoon shows smaller disturbance and jerk amplitudes than the one-leader ACC platoon, indicating improved string stability and ride comfort. Practical takeaways are twofold: first, the proposed method can be used to further develop multi-leader ACC systems. Second, even noisy data from the second leader can help stabilize traffic, which makes such systems viable in practice.

INDEX TERMS Adaptive cruise control, car-following, deep reinforcement learning, measurement noise, multi-anticipation, string stability.

I. INTRODUCTION

ADAPTIVE cruise control (ACC) is one of the most commonly discussed applications in advanced driver assistance system (ADAS). The development of this vehicle automation function achieves the Society of Automotive Engineers level 1 automation. Various ACC systems have already been implemented on commercial vehicles nowadays. Calvert et al. [1] summarized that the share of ACC-equipped vehicles on roads will reach approximately 20% by 2035. The purpose of ACC is to enhance the ride comfort and convenience by adapting the vehicle speed automatically according to the desired spacing policy. Using the measurements collected by on-board sensors, ACC systems reduce the deviation between the current and desired gaps with the directly preceding vehicle.

Early ACC systems have already shown their positive effect on improving ride comfort and reducing fuel consumption for individual vehicles [2], [3]. However, from a collective point of view, the automated control over the vehicle longitudinal dynamics provided by ACC systems can greatly impact traffic flow efficiency and stability. The effect of ACC systems on traffic flow performance has received a lot of attention [4], [5], [6], [7]. In particular, the results in Milanés and Shladover [8] and Calvert et al. [1] both indicated that the presence of autonomous vehicles (AVs) does not necessarily positively affect traffic flow in scenarios with high traffic demand. It was also demonstrated in many field

The review of this article was arranged by Associate Editor Aaron Yi Ding.

experiments and empirical studies that platoons consisting of vehicles equipped with commercial ACC systems could not ensure string stability, indicating that the disturbance caused by a preceding vehicle would be amplified as the disturbance propagates upstream [9], [10], [11]. String instability not only induces shockwaves which degrade the traffic flow efficiency but also leads to increased energy consumption and unsafe maneuvers in extreme cases [6]. With the increase in the number of ACC-equipped vehicles, this problem needs to be addressed.

One frequently discussed approach to improve string stability of AV platoons nowadays is Cooperative ACC (CACC) systems which employ wireless inter-vehicle communication technologies. When individual vehicles can obtain and utilize the in-car information of their preceding vehicles, e.g., determined vehicle acceleration in the next time step, the string stability of the platoon can be guaranteed. In addition, with the communication between multiple vehicles, CACC systems allow the measurements of leaders further ahead to be collected and utilized by the ego-vehicle. Therefore, it is possible for the ego-vehicle to respond to the downstream car-following dynamics earlier to prevent from overreacting to the disturbance. This also resembles the behavior of human drivers who tend to look at more than one vehicle ahead to adapt their car-following behaviors. This kind of driving behavior, which enhances drivers' situation awareness, is called multi-anticipation [12]. Many studies have already designed CACC systems which possess the property of multi-anticipation [13], [14], [15].

The improved traffic flow performance brought by CACC systems has been demonstrated by many studies using theoretical analysis and simulation approaches [16], [17], [18]. However, the employment of CACC systems heavily relies on communication technologies to achieve information-sharing between vehicles [19]. There are still various difficulties for practitioners to tackle before such inter-vehicle communication can be broadly accepted and implemented. Hence, a great penetration rate of connected and autonomous vehicles (CAVs) is not expected for the near future [20].

According to the aforementioned concerns, we seek to improve the string stability performance of ACC systems (without communication technologies) by relying only on a vehicle's own sensing capabilities. In this study, we argue that multi-anticipation should be technically achievable even without the help of communication technologies. ACC-equipped vehicles are able to collect range and speed measurements from not only the direct leader but also those occluded leaders further downstream by only relying on on-board sensors. A similar system design has been proposed and modeled by Donà et al. [21]. In view of the challenge of market adoption of inter-vehicle communication systems, Donà et al. [22] used a simulation approach to compare the performance of the M-ACC system proposed in Donà et al. [21] and three CACC systems in a platoon with mixed traffic. The results showed that the M-ACC system outperforms the investigated CACC systems at low penetration rate.

FIGURE 1. Multi-leader detection using only the ego-vehicle's radar.

Given the state-of-the-art development of sensor technologies, it is believed that ACCs are able to detect at least two leaders ahead, which are the direct preceding vehicle and the pre-preceding vehicle (hereinafter referred to as the first and second leaders, respectively). For instance, when using radar to detect multiple leaders, the radio waves can propagate to the second leader through diffraction or reflection under the bottom of the first leader, as illustrated in Fig. 1. Descriptions regarding the non-line-of-sight detection of the occluded vehicle through the multi-path propagation of radio signals can be found in Kopp et al. [23] and Zhou et al. [24]. Scheiner et al. [25], Hayashi et al. [26], and Palffy et al. [27] also applied such a characteristic of automotive radar to the detection of occluded vulnerable road users.

However, unlike CACC systems which use communication technologies to obtain information, leveraging sensors to collect measurements from leaders further downstream may entail relatively high sensor measurement noise, as graphically indicated by the radio wave reaching the second leader in Fig. 1. The erroneous measurements may affect string stability and ride comfort. The gap addressed in this study is the design of an ACC system which leads to stable traffic by incorporating noisy relative distance and speed measurements of the second leader. To develop and validate this design, we use microscopic traffic simulation and deep reinforcement learning (DRL).

It is worth noting that there are indeed other sources of sensor measurement uncertainties which can be considered in this problem context. For instance, the problem of losing detection (false negative) may often happen to the second leader detection task when the power of the received radar signal becomes too small, which significantly affects the reliability of the ACC system. However, this type of sensor uncertainty creates another dimension to explore and is hence left out in this study.

The remainder of this paper is structured as follows. The next section reviews articles regarding ACC controller design and how measurement uncertainty was considered. The methodology section then introduces the proposed control system architecture and explains the controller design method in detail. The experimental setup and performance evaluation framework are also elaborated here. The fourth section presents the results and compares the system performance in different scenarios. Finally, the research findings based on the evaluation results are discussed.

II. LITERATURE REVIEW

The control methods which have been applied to design ACC controller are first reviewed in this section. The second

part discusses the studies which considered measurement uncertainty in the design process of ACC systems.

A. ACC CONTROLLER DESIGN

Proportional-derivative feedback control is one of the most commonly-seen methods for the design and modeling of ACC systems in the literature. The controller computes the acceleration based on range and range rate error with the preceding vehicle. The feedback gains in the control law can be tuned to represent the sensitivity of these two factors to the next control input. Classic examples of the control law can be found in VanderWerf et al. [4] and Xiao et al. [28].

Model predictive control (MPC), or receding horizon control, is another popular control method nowadays. With the optimization modeling framework in MPC, various objectives and requirements regarding traffic stability, ecology, comfort, and driveline limitation can easily be integrated into the ACC controller design. Corona and De Schutter [29] proposed an MPC-based control method for the ACC. The cost function consists of the number of gear switching, control input variation, and deviation from the leader trajectory. Factors including the operational speed range, largest range error, and comfortable acceleration are considered in the constraints. Wang et al. [30] designed an ecological ACC system considering the carbon emission rate via a function of speed. The optimization problem was solved by dynamic programming. It was found that the eco-driving strategy also produces a smoother acceleration profile.

Employing learning-based methods to design ACC controller is another approach which has been gaining a lot of attention. Ng et al. [31] applied Monte Carlo reinforcement learning (RL) to develop an ACC controller, which updates the feedback gains in its control law at every time instance. Desjardins and Chaib-Draa [32] used the policy gradient algorithm in RL to develop a control policy which determines the vehicle acceleration action for a CACC controller.

DRL is a method which leverages the strength of both RL and deep neural networks (DNNs). DRL trains a DNN agent to solve a sequential decision-making problem through trial-and-errors. In pursuit to optimize the reward obtained from the environment, the agent explores how to make appropriate action decisions based on the state information. To dampen traffic oscillations and reduce energy consumption, Qu et al. [33] developed a car-following model based on Deep Deterministic Policy Gradient (DDPG), a DRL algorithm, for CAVs. A platoon of 10 vehicles which drive on a ring road were simulated in the environment to train the policy. Lin et al. [34] compared the DRL-based and MPC-based ACC controllers. It was pointed out that DRL may still suffer from the generalization problem of learning-based methods. However, it performed better than MPC when there are large errors, such as control delay or unexpected disturbances in the simulation experiment. Shi et al. [35] also proposed a cooperative vehicle platoon control strategy for CAVs using DRL. However, it is out of the scope of the discussion in this study. Zhou et al. [36] provided an in-depth review for the longitudinal motion planning of AVs using learning-based methods. They pointed out that automakers tend to focus on the safety performance of the longitudinal motion planning systems. Hence, particular attention was placed on the great potential of learning-based methods in integrating traffic-related domain knowledge into the motion planning design to mitigate congestion.

B. ACC MEASUREMENT UNCERTAINTY

For CACC systems, the loss and delay of inter-vehicle communication play a big role in their performance. Studies have proposed various strategies to mitigate their impact [37], [38], [39], [40]. For the proposed multi-leader ACC system in this study, on the other hand, measurement accuracy is the main concern. Obtaining accurate information from the second leader only relying on sensors is difficult. Therefore, an ACC design which considers the uncertainty is crucial. Several works have aimed to handle the uncertainty, for instance by including the stochasticity in the MPC formulation [41] or by applying the concept of robust control [42] instead of using a deterministic controller.

Learning-based methods can also be applied for autonomous driving problems under uncertainty. By formulating the problem as a Partially Observable Markov Decision Process (POMDP) and utilizing a Recurrent Neural Network (RNN) to consider temporal information, the agent in the RL algorithm is allowed to learn from the patterns of variance in the environment, e.g., from sensor noise, other perception limitations, or surrounding vehicle behavior, and handle the uncertainty accordingly [43], [44]. To develop adaptive driving behavior, Mani et al. [45] used recurrent deterministic policy gradients to train the agent how to react to the surrounding traffic condition instead of the normal DDPG. For controlling longitudinal vehicle motion specifically, Zhou et al. [46] proposed a car-following model which can react to the preceding dynamics by predicting traffic oscillation using RNNs. Albeaik et al. [47] designed a cruise controller for trucks with unknown mechanical configuration and internal state by also solving a POMDP using DRL.

Given the capability of DRL in capturing stochastic dynamics and handling uncertainty, this study proposes a DRL-based approach to design the multi-leader ACC system considering sensor noise and hence ensure the string stability of a platoon.

III. METHODS

This section first provides an overview of the proposed control system design. The second part formulates the ACC car-following problem for training the controllers (agents) and elaborates the parameter setting in the DRL algorithm. The adopted simulation approach and details of the training setup will also be covered. The last two subsections then describe the experiment scenarios and the performance indicators of interest.

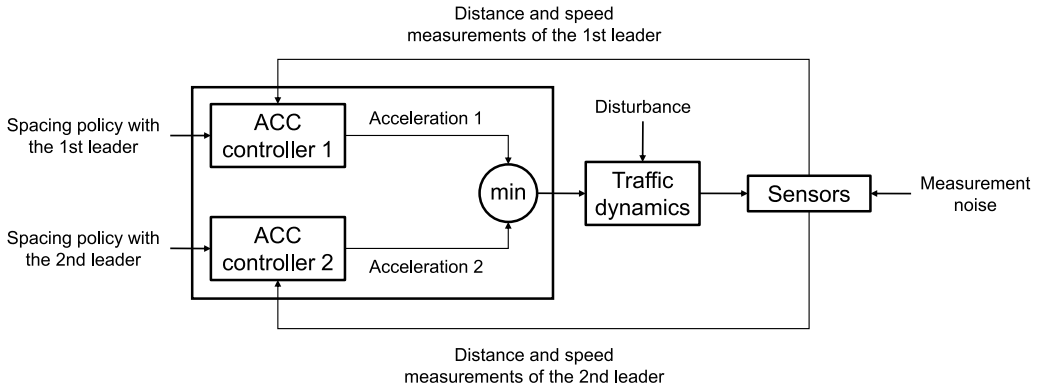

FIGURE 2. Framework of the proposed multi-leader ACC system.

A. CONTROL SYSTEM DESIGN

Fig. 2 presents the framework of the proposed multi-leader ACC system. Typical ACC systems can only utilize the distance gap and relative speed measurements with the first leader. The proposed system has two upper-level controllers, which allow it to also consider the movement of the second leader.

Based on the specified spacing policy, controller 1 uses the measurements collected from the first leader and generates the acceleration command for the ego-vehicle to follow the first leader, while controller 2 enables the ego-vehicle to follow the second leader through the same logic. The minimal acceleration command generated by these two controllers would be the control signal in the next time step, which was the same control logic adopted by Hallouzi et al. [37].

The *Traffic dynamics* block in the figure contains the vehicle motion model which is applied to simulate the car-following behavior. As mentioned in the previous section, other factors, such as the lower-level ACC controller, internal driveline, and road gradient, etc., which can affect the vehicle longitudinal motion are not considered in the defined problem context. A disturbance may occur and deviate the system dynamics from its equilibrium state. In this study, a disturbance caused by a speed fluctuation of the leading vehicles, which affects the spacing in the car-following problem, is of interest. When encountering such disturbance in the driving environment, the controllers would seek to guide the ego-vehicle back to the desired spacing state.

The on-board sensor is another important component in the system. It collects exteroceptive measurements as state input information for the controllers. As mentioned in the introduction, radar is selected for the proposed ACC system thanks to its suitability for the non-line-of-sight multi-leader detection function.

For the first leader, we assume the measurements to be accurate. The second leader measurements are believed to be comparatively less accurate. Instead of controllers which passively generate decisions based on measurements filtered by an explicit state estimator, e.g., Kalman filter, we design a controller which can directly utilize the inaccurate measurements.

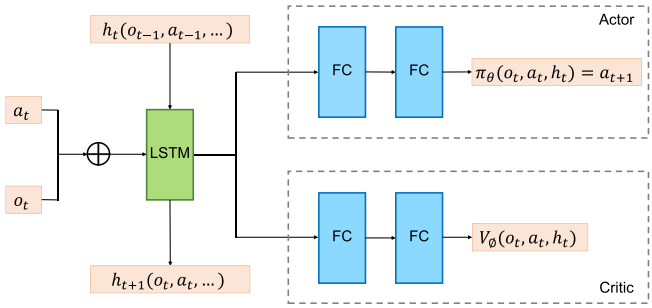

FIGURE 3. Network structure of the agent using PPO.

B. DRL FORMULATION

In the DRL problem, ACC controllers are the learning agents. Each agent learns a policy network to map the state input to the output action, which is the vehicle acceleration command. In this study, Policy Proximal Optimization (PPO) [48], an on-policy method which can be used for both discrete and continuous action spaces, is selected as the training algorithm given the relatively simple search space of the action defined in the problem.

1) DNN STRUCTURE AND PARAMETERS

There are two networks in the PPO algorithm, the actor and critic. The actor represents the policy π_θ which determines the next action, while the critic V_ϕ estimates the value function of the state-action pair. θ and ϕ denote the hyperparameters of the actor and critic, respectively.

Different from the typical DRL problems, the DNN takes the *observation* as input instead of the *state* with perfect information due to the noisy sensor measurements. A shared Long-Short-Term-Memory (LSTM) network layer is implemented before the two networks to serve as a hidden internal state estimator, as illustrated in Fig. 3. By doing so, recurrency is considered in the DRL policies, which enables the agents to learn through a sequence of data in the time series. The agents can then be trained directly with noisy measurements in a POMDP framework. The LSTM layer is expected to infer the *belief state* at every time step t using

TABLE 1. DNN setup in the agent.

	Shared LSTM	Actor	Critic
# hidden layers	1	2	2
# units (neurons) per layer	128	64	64
Activation function	—	Tanh	Tanh

the history h_t, current observation o_t, and current action a_t as input.

Both actor and critic networks contain two hidden fully-connected layers, as represented by FC in Fig. 3. The hyperbolic tangent (Tanh) function is chosen as the activation function in every layer of the networks since its output range $(-1$ to $1)$ suits the shape of the range of vehicle acceleration commands in this car-following problem. The DNN setup is summarized in Table 1. Most of the parameters in the DRL algorithm follow the default settings in the selected library.

2) OBSERVATION AND ACTION SPACES

Before formulating the observation vector which would really be applied to train the agents, this section first introduces the true underlying state vector s to outline the information which is used for the decision-making process of the DRL problem. The state vector contains the distance gap with the ith leader $g_{i,k}$, speed of the ego vehicle v_k, relative speed with the ith leader $\Delta v_{i,k}$, jerk j_k and can be written as

$$s_{i,k} = \left(g_{i,k}, v_k, \Delta v_{i,k}, j_k\right). \tag{1}$$

The distance gap, speed, and relative speed information enable the agent (the controller of the ego-vehicle) to understand the current state and allows it to determine actions which can achieve the desired state, while the jerk helps the agent to maintain a comfortable driving maneuver. The computation of distance gap $g_{i,k}$ using raw data collected from radar is described by

$$g_{i,k} = d_{i,k-\tau^s} - (i-1)\cdot l - i\cdot d_{\min}, \tag{2}$$

where $d_{i,k-\tau^s}$ is the net distance between the rear of the ith leader and the front bumper of the ego vehicle at time step $k-\tau^s$; $d_{\min}$ denotes the minimum distance gap; τ^s represents the sensor delay; l is the vehicle length used in this study. When training the policies, no sensor delay is considered so that the agent learns the correct behavior for the corresponding state observation without having to consider the uncertain influence of the delayed information. The vehicle length l is assumed to be 4 m uniformly in this study.

For other elements in the state vector, the speed of the ego-vehicle v_k and relative speed with its leader $\Delta v_{i,k}$ are information which can be directly obtained from the radar sensor. The jerk can be calculated by

$$j_k = \frac{a_k - a_{k-1}}{\Delta t}, \tag{3}$$

where a_k and a_{k-1} denote the vehicle acceleration values at the current and previous time step, respectively. Δt is the length of a single time step, which is set to 0.1 s in this study.

As mentioned, to train the controllers which can handle noisy measurements, the observation o would replace the states s, as described by

$$o_{i,k} = \left(g'_{i,k}, v_k, \Delta v'_{i,k}, j_k\right), \tag{4}$$

where $g'_{i,k}$ represents the distance gap with a random noise error, while $\Delta v'_{i,k}$ is the relative speed calculated from the speed of the leader which also contains random error.

Random errors ϵ are added to the distance gap and relative speed measurements to represent the measurement noise. The error terms follow two different independent zero-mean Gaussian distributions with standard deviations σ_{g_i} and σ_{v_i}, respectively, as shown by

$$g'_{i,k} = g_{i,k} + \epsilon \sim \left(0, \sigma^2_{g_i}\right) \tag{5}$$

and

$$\Delta v'_{i,k} = \Delta v_{i,k} + \epsilon \sim \left(0, \sigma^2_{v_i}\right). \tag{6}$$

In this study, the values of σ_{g_i} and σ_{v_i} are scenario-specific parameters, which will be determined based on the experimental scenarios.

In the defined problem, internal measurements of the ego-vehicle which can be obtained from interoceptive sensors, such as its own speed v_k and acceleration a_k, are assumed to be accurate. Hence, the jerk j_k would also remain unchanged.

The action defined in the DRL framework is the control signal u of the ACC controllers ranging from $u^{\min}$ = -6 m/s^2 to $u^{\max}$ =3 m/s^2, which is also the range of the resulting vehicle acceleration a, taking reference from the range of aggressive driving behavior defined in Bae et al. [49]. In safety critical situations, an even larger uncomfortable deceleration value may be required. However, they are not discussed in this problem context.

3) REWARD FUNCTION

The reward function should be designed to train the controllers to maintain the desired time gaps with the two leaders ahead. The proposed ACC system follows a constant time gap policy.

The function consists of three components, time gap error, jerk, and correctness of the action, while each of them represents the spacing policy, driving comfort, and penalty for undesired actions, respectively. In Shladover [50], it was pointed out that limiting jerk amplitudes has a potential destabilizing effect on the longitudinal dynamics of following vehicles. Therefore, the reward function should be able to guide the system to achieve a certain level of balance between these two factors. A negative reward function

$$R_{i,k} = \alpha\cdot\frac{-|e_{i,k}|}{e_i^{\max}} + \beta\cdot\frac{-|j_k|}{j^{\max}} + \min\left(\frac{e_{i,k-1}-e_{i,k}}{e_i^{\max}}, 0\right) \tag{7}$$

is designed. A negative reward function implies that every element in the function is negative. In (7), $e_{i,k}$ represents the time gap error, which can be calculated by

$$e_{i,k} = tg_{i,k} - tg_i^* \tag{8}$$

where $tg_{i,k} = g_{i,k}/v_k$ is the time gap of the ego-vehicle with the ith leader at time step k, and $tg*_i$ is the desired time gap with the ith leader determined according to the given spacing policy. To explore the performance limit of the proposed system, tg_1^* is set to 1 s for the controller (controller 1) which is responsible to follow the first leader, while tg_2^* would be 2 s for controller 2 to follow the second leader. An even smaller time gap is considered to be too dangerous given the reaction time required by human drivers when they have to take-over the control of the vehicle when a safety-critical situation occurs. $e_i^{\max}$ and $j^{\max}$ are values specified to normalize the gap error term and the jerk term, respectively. In this training setup, $e_i^{\max} = tg_i^*/2$ and $j^{\max} = (u^{\max} - u^{\min})/3/\Delta t$. α and β are the weights for the first two elements, respectively. The weighting shows the hypothetical trade-off between the first two factors in the function. There is no weighting for the third element since it serves as a penalty term in the function. In this study, a weighting combination $\alpha = 3/4$ and $\beta = 1/4$ is determined after several trials to ensure a desired level of string stability and ride comfort.

Based on the negative reward function design, the policy would seek to attain zero reward as fast as possible in the episode so that the ego-vehicle can reach the stable state (desired time gap and zero relative speed). In addition, the continuous components in the function provide a gradient for the agent to understand whether it is getting closer to the desired state and find the correct direction in the search space.

A negative value $P = -100$ would be added as a large penalty term if the resulting state falls into certain regions in the state space which are unreasonable or may lead to disengagement of the ACC system, including collision ($tg_{1,k} \leq 0$), an extremely long time gap ($tg_{1,k} \geq tg_i^* + 5$), or a negative speed of the ego-vehicle ($v_k < 0$). The value of P should be small enough so as to assign penalty to the incorrect behavior, but not too small which may cause an unnecessary difficulty in the training process.

4) TRAINING TASK

It is important to note that the created training tasks for the recurrent policies should contain enough variability so that the LSTM layer can be properly trained to help infer the future state, and such that overfitting can be prevented. Therefore, a traffic disturbance pattern with different amplitudes is designed for each training episode.

Each episode simulates a 30 s car-following task, which is equivalent to 300 steps for the ACC system. At the beginning of each episode, the leader and ego-vehicle start from the equilibrium state in which the two vehicles have no relative speed ($\Delta v_{i,0} = 0$) and keep the desired time gap ($tg_{i,0} = tg_i^*$). The disturbance then occurs when the leader speeds up or slows down at a certain moment, which may be caused by the cut-in and cut-out behavior of a vehicle further downstream. The leader stays at the resulting speed for a certain duration and starts to recover back to the original speed by either accelerating or decelerating at a smaller constant rate than the disturbance. Such behavior leads to a shockwave which propagates upstream along the platoon. Fig. 4 shows an example of the acceleration and speed profiles of the leader for the training of controllers.

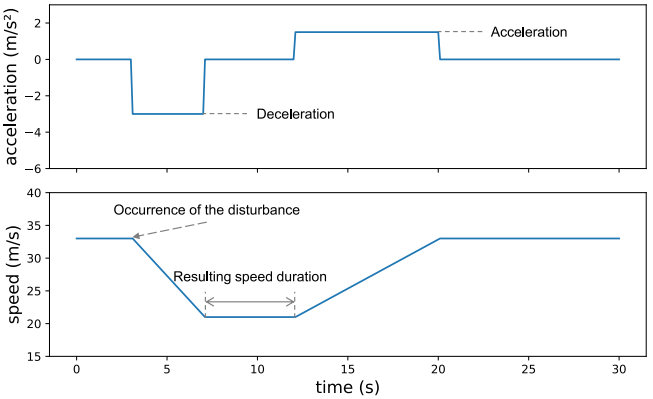

FIGURE 4. Example of the training tasks (initial speed = 35 m/s, deceleration = -2.4 m/s² from 5 s to 11 s, resulting speed duration = 6 s, and acceleration = 1.6 m/s²).

The range of parameters, including initial speed, disturbance duration, and accelerating/decelerating strength, etc., are described below:

- Initial speed of vehicles: [15 m/s, 35 m/s]
- Disturbance occurrence time: [2 s, 4 s]
- Disturbance duration: (0 s, 5 s]
- Disturbance acceleration/deceleration: [-4 m/s^2, 2 m/s^2]
- Resulting speed duration: [0.5 s, 8 s]

The acceleration/deceleration strength for speed-recovery ranges from 1/3 to 1 time of the original deceleration/acceleration strength during the disturbance. It is worth noting that the resulting speed caused by the disturbance would be bounded between 11 m/s (40 km/h) and 39 m/s (140 km/h), which can be regarded as the operational design domain of the proposed multi-leader ACC system.

The training episode would be terminated if one of the termination criteria is met, as the conditions mentioned in the reward function design. This prevents the agent from continuing the exploration in situations which are unlikely to happen and makes the training process more efficient.

5) SIMULATION AND TRAINING SETUP

A numerical simulation approach is adopted to represent the environment in the DRL method. Following Newton's law of motion, a vehicle motion model

$$\begin{cases} x_k = x_{k-1} + v_{k-1} \cdot \Delta t + \frac{1}{2} \cdot a_{k-1} \cdot \Delta t^2 \\ v_k = v_{k-1} + a_{k-1} \cdot \Delta t \\ a_k = a_{k-1} + (u_k - a_{k-1}) \cdot \Delta t / \tau^A \end{cases} \tag{9}$$

is formulated. In (9), x_k denotes the vehicle position, and τ^A is the actuator lag resulting from the lower-level ACC controller and the vehicle driveline, including the engine, throttle, and brake response. In this study, a 0.2 s actuator lag is implemented, as the value used in Xiao and Gao [51] and Wang et al. [52].

All the state information required by the controller can be derived from the updated position, speed, and acceleration of the leader and follower. These equations mathematically describe the movement of vehicles and the transition between states. Based on the action carried out and the simulated car-following behaviors, the reward function value can be computed.

Considering the large number of neurons and parameters in the neural network with an LSTM layer which need to be optimized, each agent is trained for 6 million steps. During the training process, the performance of the trained output policy is evaluated every 0.1 million steps by simulating 100 randomly generated episodes. The evaluated policy with the highest average reward would be stored as the optimal policy.

C. EXPERIMENTAL DESIGN

The experiments are carried out using the same numerical simulation framework described in the previous subsection. This subsection then describes the experiment scenarios. The car-following dynamics of a 20-vehicle platoon is simulated. The behavior of the first vehicle is pre-determined, while the other following vehicles act according to the acceleration command generated from the multi-leader ACC system.

1) DISTURBED LEADER BEHAVIOR

To test the string stability of the platoon, a disturbance needs to be generated to trigger the response of the following vehicles. From a traffic flow perspective, a fluctuation in the speed profile of the leading vehicle in the platoon is of interest since it is often the cause of the propagation of shockwaves on motorways. Therefore, this study considers a 50-s vehicle driving behavior with a speed fluctuation, which is similar to the disturbance case described in Li et al. [53].

Fig. 5 shows the designed speed profile for the first vehicle. The vehicle starts with a constant speed of 33 m/s for 3 seconds and conducts a strong deceleration with -3 m/s^2 for 4 s and maintains at the low speed 21 m/s for 5 s. Afterward, it spends 8 s speeding up with 1.5 m/s^2 to reach the original speed. This is a comparatively aggressive braking and accelerating maneuver, which helps evaluate the effectiveness of the proposed ACC design in this study.

2) MEASUREMENT NOISE

Sensor uncertainty is another important aspect in this study. An amount of sensor delay $\tau^s = 0.2$ s, which can degrade the system performance, is applied in all simulation scenarios. Noise is another frequently discussed source of uncertainty for sensor measurements. Measurement noise can be subject to the distance and speed of the detection target or the

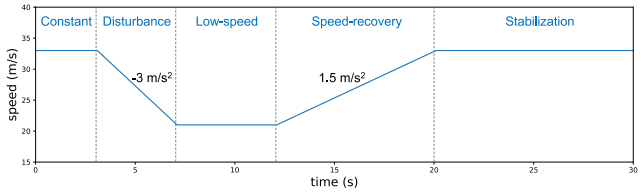

FIGURE 5. Speed profile of the first vehicle in the platoon in the experiment.

environment. However, the influence from these external factors is excluded in this study. Only random errors caused by the noisy measurements are considered. The errors of the distance gap and speed measurements are assumed to follow two independent zero-mean Gaussian distributions with standard deviations σ_{g_i} and σ_{v_i}, respectively. In this study, the accuracy of the first leader information is fixed at the level with $\sigma_{g_1} = 0.2$ m and $\sigma_{v_1} = 0.2$ m/s, which are slightly larger than the value tested in Zhou et al. [41].

The accuracy of the second leader measurements, however, is still an unexplored problem in ADAS applications. It is hypothesized that the second leader measurements would be more erroneous than the first leader measurements. Therefore, the system performance will be evaluated in four different levels of second leader measurement noise:

- N1: $\sigma_{g_2} = 0.5$ m and $\sigma_{v_2} = 0.5$ m/s
- N2: $\sigma_{g_2} = 1.0$ m and $\sigma_{v_2} = 1.0$ m/s
- N3: $\sigma_{g_2} = 1.5$ m and $\sigma_{v_2} = 1.5$ m/s
- N4: $\sigma_{g_2} = 2.0$ m and $\sigma_{v_2} = 2.0$ m/s

Each level has a standard deviation value for both the distance and speed measurements of the second leader. A radar sensor which has an even larger level of measurement noise than N4 would be considered unreliable for the multi-leader detection function and any other ADAS applications. Each scenario will be simulated 20 times to account for the stochasticity.

D. PERFORMANCE EVALUATION

To understand the effect of the proposed ACC design on the car-following dynamics in the platoon, this subsection describes the framework for performance evaluation and defines the performance indicator.

1) CAR-FOLLOWING BEHAVIOR MECHANISM

Before showing the performance indicators, this study seeks to explore the car-following behaviors of vehicles equipped with the proposed ACC systems by comparing the trajectories of the following vehicles with the hypothetical trajectories generated from Newell's car-following model [54]. The method was originally adopted by Laval and Leclercq [55] to investigate the behaviors of human driven vehicles. Li et al. [53] then also applied this method to analyze the behavior of ACC-equipped vehicles in empirical experiments.

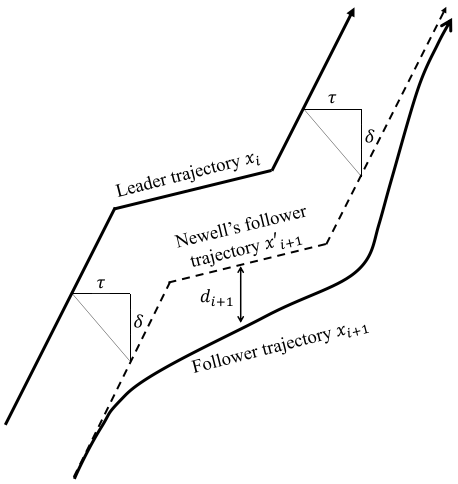

FIGURE 6. Deviation between the follower trajectory and the computed Newell's follower trajectory.

In Newell's car-following model, it is stated that the position of the following vehicle x'_{i+1} at time t is a distance δ upstream of the position of the preceding vehicle x_i at time $t-\tau$. The trajectories of the leader and the follower are identical except for a shift of space and time between them. The mathematical formulation of Newell's car-following model can be described by

$$x'_{i+1}(t) = x_i(t-\tau) - \delta \tag{10}$$

According to the constant time gap spacing policy adopted by the ACC system in this study, the time shift τ would be 1 s, and the distance shift δ would be 6 m (the jam headway calculated by summing the minimum distance gap $d^{\min}$ and vehicle length l).

Fig. 6 also presents how Newell's follower trajectory (black dashed line) is generated. The trajectory deviations (d_{i+1} in the figure) between the simulated and Newell's follower trajectories at every time step are computed to help discover the driving behavior response of the following vehicles to the preceding traffic disturbance. If there is a positive deviation, which is when the simulated trajectory stays below Newell's trajectory, it represents a relatively conservative driving maneuver. In contrast, a negative deviation indicates that the follower drives rather aggressively.

2) STRING STABILITY

To investigate the string stability performance, some indicators of this property should be defined to quantitatively evaluate the results of the conducted simulation experiments. For simplicity, this study only investigates the impact of disturbance on the speed of the following vehicles. It can be observed whether the disturbance is amplified or dissipated by looking at the propagation of the speed drop amplitude in the platoon. Therefore, the analysis first focuses on the response of the ACC system on the propagation of speed fluctuation. The speed drop amplitude of each vehicle is computed by the difference between the original speed and the lowest speed in the profile.

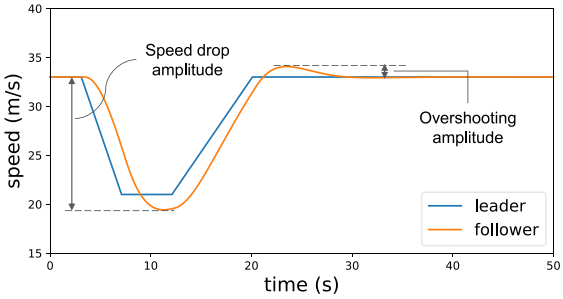

FIGURE 7. String stability indicators of a follower with local instability.

Additionally, after the following vehicle speeds up in the acceleration phase, the speed of it may slightly exceed the desired original speed. This phenomenon is referred to as *overshooting* in this study. It is also related to the string stability performance since a large overshooting amplitude can potentially lead to another traffic disturbance event. Hence, the overshooting amplitude, which is calculated by the difference between the original speed and the highest speed in the stabilization phase, is the second performance indicator. Both the decreasing speed drop amplitudes and decreasing overshooting amplitudes along the platoon indicate string stability. The two indicators should ideally decrease as fast as possible in the upstream direction after applying the proposed system. Fig. 7 provides a graphical example of the two string stability indicators shown in the speed profiles of a leader and an unstable follower.

3) RIDE COMFORT

The existence of measurement noise may not only affect the string stability performance but also result in uncomfortable driving maneuvers due to fluctuations in the acceleration profile. The implemented LSTM network is expected to serve as a state estimator which can smooth out the noise in both the distance and speed measurements so that the jerk at every time step can be reduced. In addition to ride comfort, reducing the jerk between consecutive time steps is also an important aspect for the vehicle driveline and mechanical systems. To evaluate the ride comfort performance of the proposed system when facing noisy measurements, the jerk amplitudes experienced by vehicles in the platoon are analyzed by looking at the probability distribution of jerk amplitudes of all vehicles in the simulation. By doing so, we examine whether the system produces uncomfortable driving maneuvers when facing the disturbance event and measurement noise.

The boundaries of comfortable and aggressive driving maneuvers have to be defined beforehand. According to Bae et al. [49], the jerk threshold for normal and comfortable driving behaviors can range from 0.3 m/s^3 to 0.9 m/s^3. Jerk amplitudes ranging from 0.9 m/s^3 to 2 m/s^3 are regarded as aggressive driving behaviors. A jerk amplitude larger than

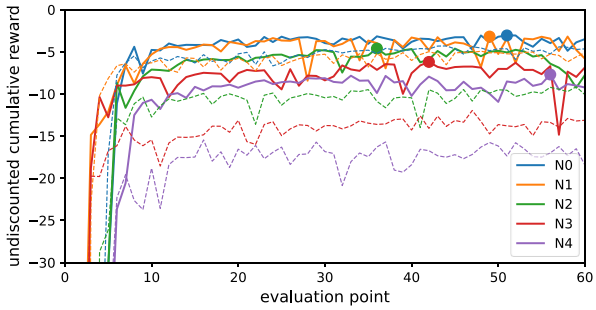

FIGURE 8. Average undiscounted cumulative reward at each evaluation point in the training process under each noise level (markers: the highest average reward evaluated; dashed curves: learning performance of non-recurrent policies).

$2\ \mathrm{m/s^3}$ is therefore considered abnormal and only occurs in emergency conditions.

IV. RESULTS

This section first shows the training outcome of the DRL policies. Following the proposed evaluation framework, the simulation output is then analyzed to understand the platoon performance. The advantage of multi-anticipative car-following behavior and the influence of measurement noise are also explored.

A. TRAINED DRL POLICIES

Before studying the performance of the simulated platoons, we first assess the training of the DRL policies under different levels of noise N0 to N4 (see Section III-C.2). The curves in Fig. 8 record the undiscounted cumulative reward at every evaluation point during the training process of each agent trained with a different level of measurement noise. As can be observed, the cumulative rewards under every noise level successfully reach a plateau after the tenth evaluation point, implying the convergence of the training process. On the other hand, the training performance is influenced by the uncertainty level. With higher level of noise, the highest undiscounted cumulative reward each trained policy can achieve decreases. In addition, the process with a higher level of noise leads to a more unstable training performance.

On the other hand, the figure demonstrates the positive effect of including the LSTM layer into the DNN structure. The dashed curves in the figure show the learning processes of each non-recurrent policy under each noise level. It is found that their maximum cumulative rewards are all smaller than those of the recurrent policies. The difference becomes even larger at higher noise levels. This implies the importance of making use of the power of memories to generate decisions under uncertainty.

B. SIMULATED PLATOON DYNAMICS

The simulation result of the one-leader system is first presented to serve as a benchmark for the multi-leader system. Fig. 9 shows the speed contour plot of the platoon which uses the one-leader ACC system when facing measurement noise level N0. The contour plot helps understand the propagation of the jam wave caused by the disturbance and provides a first impression of the system performance. As can be observed from the color in the low-speed phase, the lowest speed at the upstream of the platoon is higher than that of the first vehicle although the effect seems insignificant. This implies that the one-leader system can slightly mitigate the disturbance even when facing noisy measurements. Little overshooting phenomena are observed in the stabilization phase, which will be quantitatively examined in the next subsection.

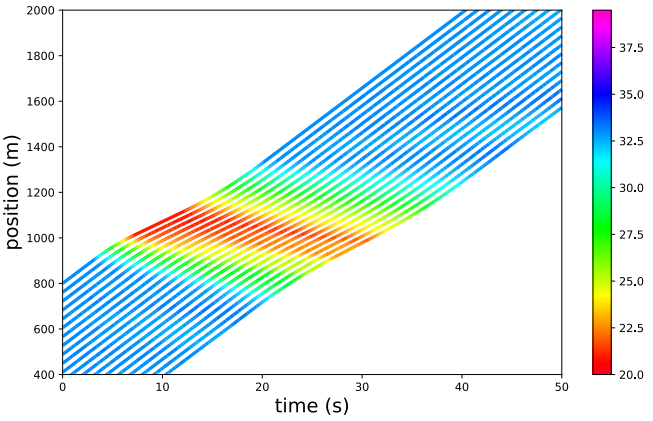

FIGURE 9. Speed contours of the platoon using the one-leader ACC system under measurement noise level N0.

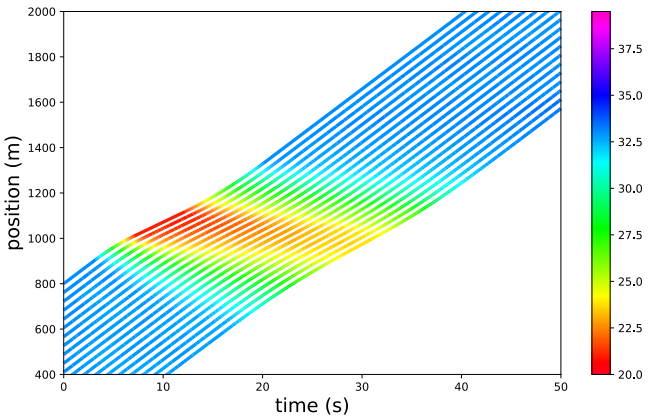

FIGURE 10. Speed contours of the platoon using the multi-leader ACC system under measurement noise level N1.

Figs. 10 and 11 then show the speed contour plots of the multi-leader systems in noise levels N1 and N4, respectively. The detailed performance indicators under each noise level are shown in Fig. 14 in the next subsection.

A significant improvement of car-following dynamics is shown in Fig. 10. The vehicle speed at the upstream of the platoon when coping with the disturbance is clearly larger than that of the platoon using the one-leader system. Compared to Fig. 9, it can also be observed that the deceleration phase (the green region) starts earlier for the following vehicles, and the overshooting phenomenon in the stabilization phase disappears. These all indicate that the followers are able to react to the preceding disturbance

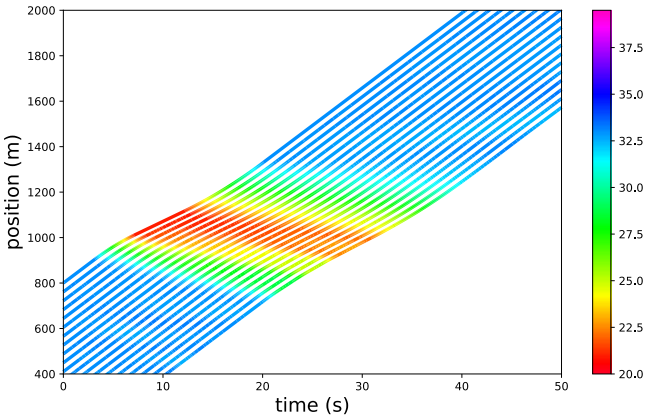

FIGURE 11. Speed contours of the platoon using the multi-leader ACC system under measurement noise level N4.

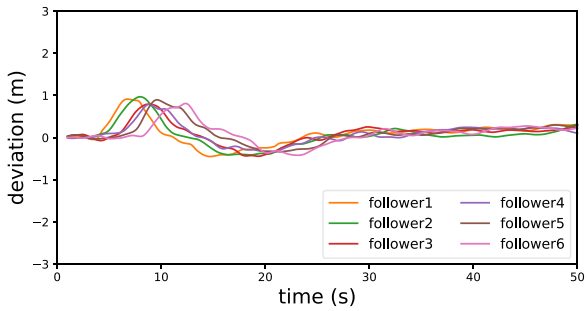

FIGURE 12. Trajectory deviation of the platoon using the one-leader ACC system under measurement noise level N0.

earlier thanks to the measurements collected from the second leaders. When using the multi-leader system in the platoon, the jam wave dissipates more rapidly, showing the merit of multi-anticipation.

The simulation result of the multi-leader system when facing measurement noise level N4 is shown in Fig. 11. Compared to the performance in the scenario with measurement noise level N1, the jam wave dissipation becomes slower. The speed contour plot shows a similar pattern to that of the one-leader system in Fig. 9. The early slow down behaviors are no longer observable in the deceleration phase of the trajectories. In addition, more speed irregularities and overshooting phenomena are found in the constant speed and stabilization phases, which also occur when using the one-leader system. The system performance degrades when there is a greater level of measurement noise although the jam wave can still be dissipated while propagating to the upstream.

Figs. 12 and 13 calculate the trajectory deviation of the first six followers compared to their Newell's trajectories in each noise level scenario. The positive deviations in Fig. 12 indicate that the following vehicles using the one-leader system slow down more than the Newell's trajectories in the deceleration phase. These relatively conservative driving behaviors mitigate the impact of the disturbance wave. The multi-leader system also exhibits similar behavior, as can be seen in Fig. 13. At noise level N1 for instance, the second, fourth, and sixth followers slow down even more than in the case of one-leader system. Such a behavior is resulted from the additional measurements collected from the second leader, which enable the vehicles to response to the disturbance earlier. However, it is also observed that this pattern gradually disappears as the measurement noise becomes larger. At level N4, the multi-leader system shows no significant difference from the one-leader system. The noisy measurements influence the ability of the controller to respond to the disturbance.

C. STRING STABILITY PERFORMANCE

The string stability performance is summarized by plotting the development of the indicators along the platoon in each scenario in Fig. 14. It shows the average speed drop and overshooting amplitudes experienced by each follower in all the simulation runs. The best performance is found when the platoon consists of vehicles equipped with the multi-leader system under measurement noise level N1. The speed drop amplitude decreases to around 9.3 m/s for the last follower in the platoon, while the overshooting amplitude stays within 0.3 m/s across the entire platoon. As can be seen from the speed drop amplitude curves, the value increases while the measurement noise becomes larger. At noise level N4, the curve almost overlaps with that of the one-leader system under noise level N0. This again implies that the benefit of utilizing the second leader measurements gradually disappears if the measurement noise becomes larger.

The curves of overshooting amplitudes exhibit a similar trend. The multi-leader system manages to reduce the overshooting amplitudes at the first three noise levels. However, at level N4, the amplitudes even become slightly larger than those of the one-leader system. Still, the overall overshooting amplitudes remain below 0.5 m/s, which is therefore deemed negligible.

D. RIDE COMFORT PERFORMANCE

The ride comfort performance of the proposed system is evaluated by examining the jerks experienced by following vehicles in the platoon at every time step. Noisy sensor measurements and disturbance can lead to fluctuating acceleration command. It is hence worth analyzing the jerk amplitudes experienced by vehicles in the platoon to see whether the system can still produce comfortable movements in such conditions.

Fig. 15 first shows the cumulative distributions of jerks in each noise level. Compared to the dotted curves which show the jerk distribution of the non-recurrent policies, the number of large jerk amplitudes experienced by vehicles equipped with the proposed ACC system using the trained recurrent policies are significantly reduced as the distributions are more concentrated at around 0 m/s^3. This verifies the effect of recurrency, which provides the controller with state estimation ability to reduce the impact of measurement noise.

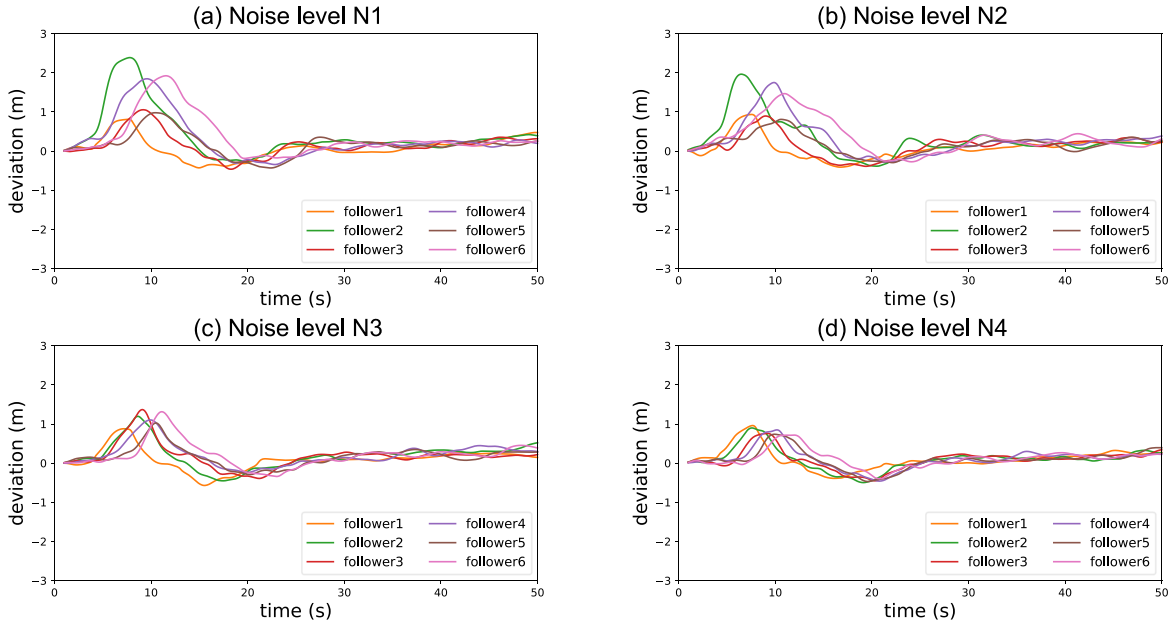

FIGURE 13. Trajectory deviation of the platoon using the multi-leader ACC system under measurement noise levels N1-4.

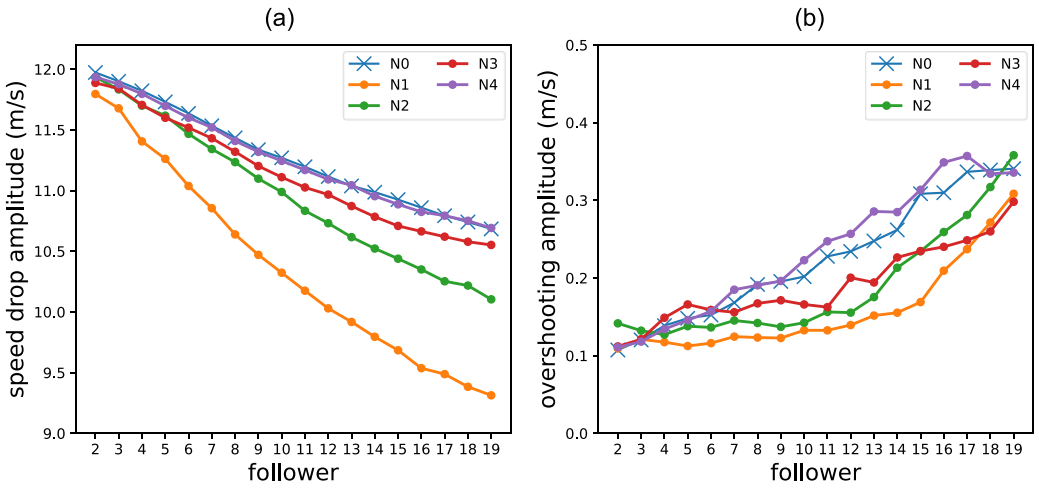

FIGURE 14. String stability indicators: (a) Average speed drop amplitude and (b) average overshooting amplitude of each following vehicle in the platoon.

Following the defined ride comfort thresholds, Fig. 16 plots the distributions of jerk amplitudes across different ride comfort levels in each noise level. One can see that the jerk amplitudes also follow a similar trend as discovered from the string stability indicators. Larger measurement noise results in more uncomfortable driving maneuvers. At levels N1, N2, and N3, the multi-leader system attains better ride comfort performance than the one-leader system under noise level N0. More than 90% of the driving maneuvers can stay within the comfortable region. This positive effect should be attributed to the second leader measurements, which help followers achieve multi-anticipation and hence prevent the need of conducting abrupt driving behaviors to recover back to the desired state after the disturbance. However, there are more aggressive driving maneuvers at level N4 than at other noise levels. The increased number of large jerk amplitudes can be either caused by the larger measurement noise or the degraded string stability, which requires the vehicles to conduct more uncomfortable driving maneuvers.

V. DISCUSSION

From the results of string stability performance, it is found that the one-leader system can already achieve a certain level of stability when looking at the slightly decreasing speed drop amplitude even though the string stability is not explicitly considered in the DRL setup. This also demonstrates the capability of the trained recurrent policy with the LSTM layer in handling state information uncertainty since a small level of measurement noise is already considered.

Compared to the one-leader system, the multi-leader ACC system shows better string stability performance when the measurement noise is relatively small. The speed

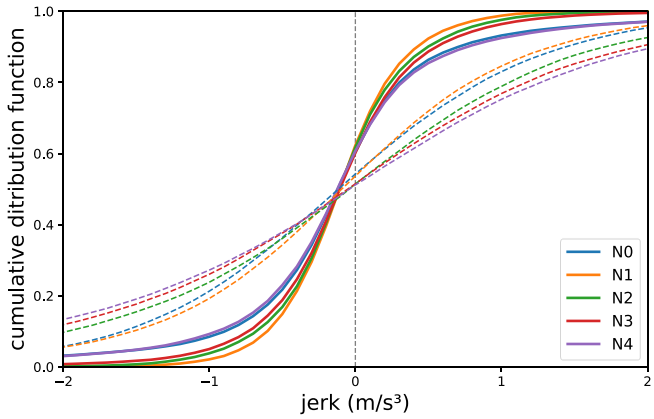

FIGURE 15. Cumulative distribution of jerks experienced by following vehicles (dashed curves: the ride comfort performance of non-recurrent policies).

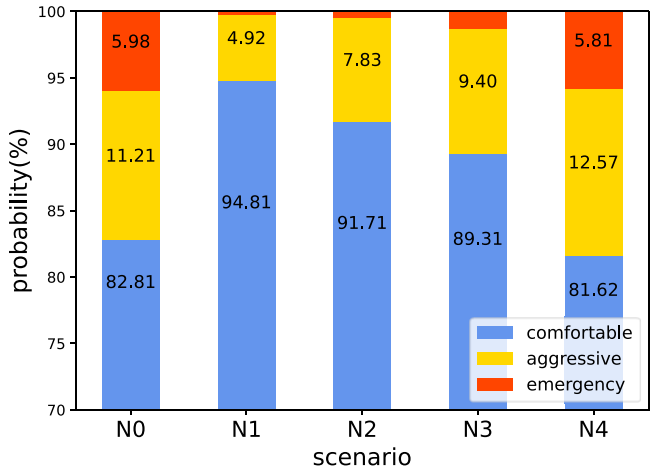

FIGURE 16. Probability distribution of jerk amplitudes in each ride comfort level.

drop amplitude decreases more rapidly when the second leader measurement is utilized, indicating the propagating disturbance can be damped out faster. The overshooting phenomenon is also mitigated when the followers can look at two leading vehicles at the same time. In addition, the ride comfort performance is improved when the following vehicles are equipped with the multi-leader system. The aggressive driving maneuvers of the following vehicles equipped with one-leader system mostly disappear. The following vehicles in the platoon have smoother trajectories when they can also perceive the second leader. The improvement on both string stability and ride comfort showcases the benefit of multi-anticipation.

On the other hand, different levels of measurement noise influence the string stability and ride comfort performances. The degradation occurs while the measurement noise increases. The benefit of multi-anticipation can still be observed at noise level N3. At level N4, however, the multi-leader system exhibits a similar level of performance to the one-leader system in both aspects. This also implies the performance limit of the trained controller in handling measurement noise. The uncertainty reduces the positive effect brought by multi-anticipation.

The simulation experiment considers a 20-vehicle platoon to showcase the effect of the proposed ACC system design. Judging from the string stability performance in Fig. 14, we expect that the platoon size can be further scaled as the speed drop amplitude has not yet reached a minimal plateau. However, it is important to be aware of the slightly increasing overshooting amplitude in case another disturbance is formed due to the excessive acceleration behavior. A longer simulation time horizon is also required for the performance assessment of a longer ACC platoon.

Moreover, Gaussian noise is the only source of measurement uncertainty considered in this study. It is worth noting that other types of uncertainty, such as false positive alarms or losing detection may also occur when using radar to detect the second leader. As mentioned, these characteristics are not considered in the study scope. For future work, this problem can be addressed by including other sources of measurement uncertainty in the training setup and numerical simulation so that the controllers possess the ability to tackle these situations.

VI. CONCLUSION

To enhance the string stability of a platoon consisting of ACC-equipped vehicles, this study applies DRL to design an ACC system utilizing noisy sensor measurements collected from multiple leading vehicles. The concept of recurrency is applied in the DRL formulation to handle state uncertainty caused by measurement noise. String stability and ride comfort performances of the proposed system are evaluated by a numerical simulation experiment. Relevant indicators are identified to help quantitatively analyze the simulation output.

The advantage of using the recurrent policies can first be seen from the learning curves which record the performance of the controllers during the training process. According to the higher cumulative rewards, it is demonstrated that the controllers which possess the ability of state estimation on their own outperform the controllers designed by the original DRL method. The evaluation results then show that the multi-leader ACC system can successfully improve both string stability and ride comfort performance when the measurement noise is relatively small. The achieved multi-anticipative car-following behavior damps out the propagating disturbance more rapidly compared to the case using the one-leader system. The number of aggressive or emergency driving maneuvers also becomes smaller when using the multi-leader system. The performance boundary is found at the highest tested measurement noise level N4, in which the standard deviations of the noise are 2 m and 2 m/s for distance and speed measurements, respectively. In such a condition, both string stability and ride comfort performances degrade to the similar level of a platoon using the one-leader system.

In this study, the inaccuracy of sensor measurements is considered both in the design and performance evaluation of the proposed multi-leader ACC system. The results and findings can contribute to the development of advanced sensor technologies and ACC systems with the capability of multi-anticipation and measurement uncertainty handling. Nonetheless, this study only investigates the influence of measurement noise. It is believed that the learning-based approach has the potential to be extended to cope with cases with other types of measurement uncertainty, which other control methods may not be able to tackle. Another aspect which is worth further consideration is the influence of platoon heterogeneity. In scenarios with different penetration rates of the multi-leader system, how much improvement we can observe from the platoon or traffic flow should be quantitatively assessed.

REFERENCES

[1] S. C. Calvert, W. J. Schakel, and J. W. C. van Lint,"Will automated vehicles negatively impact traffic flow?" *J. Adv. Transp.*, vol. 2017, pp. 1–17, Sep. 2017.

[2] G. Marsden, M. McDonald, and M. Brackstone, "Towards an understanding of adaptive cruise control," *Transp. Res. Part C, Emerg. Technol.*, vol. 9, no. 1, pp. 33–51, Feb. 2001.

[3] L. Xiao and F. Gao, "A comprehensive review of the development of adaptive cruise control systems," *Veh. Syst. Dyn.*, vol. 48, no. 10, pp. 1167–1192, Oct. 2010.

[4] J. VanderWerf, S. Shladover, N. Kourjanskaia, M. Miller, and H. Krishnan, "Modeling effects of driver control assistance systems on traffic," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1748, no. 1, pp. 167–174, Jan. 2001.

[5] A. Spiliopoulou, D. Manolis, F. Vandorou, and M. Papageorgiou, "Adaptive cruise control operation for improved motorway traffic flow," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 22, pp. 24–35, Dec. 2018.

[6] M. Makridis et al., "Empirical study on the properties of adaptive cruise control systems and their impact on traffic flow and string stability," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 4, pp. 471–484, Apr. 2020.

[7] T. Li, D. Chen, H. Zhou, Y. Xie, and J. Laval, "Fundamental diagrams of commercial adaptive cruise control: Worldwide experimental evidence," *Transp. Res. Part C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103458.

[8] V. Milanés and S. E. Shladover, "Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data," *Transp. Res. Part C, Emerg. Technol.*, vol. 48, pp. 285–300, Nov. 2014.

[9] V. L. Knoop, M. Wang, I. Wilmink, D. M. Hoedemaeker, M. Maaskant, and E.-J. Van der Meer, "Platoon of SAE level-2 automated vehicles on public roads: Setup, traffic interactions, and stability," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 9, pp. 311–322, Sep. 2019.

[10] G. Gunter et al., "Are commercially implemented adaptive cruise control systems string stable?" *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6992–7003, Nov. 2021.

[11] B. Ciuffo et al., "Requiem on the positive effects of commercial adaptive cruise control on motorway traffic and recommendations for future automated driving systems," *Transp. Res. Part C, Emerg. Technol.*, vol. 130, Sep. 2021, Art. no. 103305.

[12] S. P. Hoogendoorn, S. Ossen, and M. Schreuder, "Empirics of multianticipative car-following behavior," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1965, no. 1, pp. 112–120, Jan. 2006.

[13] K. Hasebe, A. Nakayama, and Y. Sugiyama, "Dynamical model of a cooperative driving system for freeway traffic," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 68, no. 2, Aug. 2003, Art. no. 26102.

[14] I. Wilmink, G. Klunder, and B. van Arem, "Traffic flow effects of integrated full-range speed assistance (IRSA)," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 1204–1210.

[15] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control," *Transp. Res. Part C, Emerg. Technol.*, vol. 40, pp. 290–311, Mar. 2014.

[16] R. Pueboobpaphan and B. Van Arem, "Driver and vehicle characteristics and platoon and traffic flow stability," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2189, no. 1, pp. 89–97, Jan. 2010.

[17] W. J. Schakel, B. van Arem, and B. D. Netten, "Effects of cooperative adaptive cruise control on traffic flow stability," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, 2010, pp. 759–764.

[18] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2324, no. 1, pp. 63–70, Jan. 2012.

[19] S. E. Shladover, C. Nowakowski, X.-Y. Lu, and R. Ferlis, "Cooperative adaptive cruise control," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2489, no. 1, pp. 145–152, Jan. 2015.

[20] P. Bansal and K. M. Kockelman, "Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies," *Transp. Res. Part A, Policy Pract.*, vol. 95, pp. 49–63, Jan. 2017.

[21] R. Donà, K. Mattas, Y. He, G. Albano, and B. Ciuffo, "Multianticipation for string stable adaptive cruise control and increased motorway capacity without vehicle-to-vehicle communication," *Transp. Res. Part C, Emerg. Technol.*, vol. 140, Jul. 2022, Art. no. 103687.

[22] R. Donà, K. Mattas, G. Albano, and B. Ciuffo, "Multianticipative adaptive cruise control compared with connectivity-enhanced solutions: Simulation-based investigation in mixed traffic platoons," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2677, no. 8, pp. 573–587, Aug. 2023.

[23] J. Kopp, D. Kellner, A. Piroli, and K. Dietmayer, "Fast rule-based clutter detection in automotive radar data," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, 2021, pp. 3010–3017.

[24] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, p. 4208, May 2022.

[25] N. Scheiner et al., "Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 1–10.

[26] S. Hayashi, K. Saho, D. Isobe, and M. Masugi, "Pedestrian detection in blind area and motion classification based on rush-out risk using micro-doppler radar," *Sensors*, vol. 21, no. 10, p. 3388, May 2021.

[27] A. Palffy, J. F. P. Kooij, and D. M. Gavrila, "Detecting darting out pedestrians with occlusion aware sensor fusion of radar and stereo camera," *IEEE Trans. Intell. Veh.*, vol. 8, no. 2, pp. 1459–1472, Feb. 2023.

[28] L. Xiao, M. Wang, and B. van Arem, "Realistic car-following models for microscopic simulation of adaptive and cooperative adaptive cruise control vehicles," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2623, no. 1, pp. 1–9, Jan. 2017.

[29] D. Corona and B. De Schutter, "Adaptive cruise control for a SMART car: A comparison benchmark for MPC-PWA control methods," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 2, pp. 365–372, Mar. 2008.

[30] M. Wang, W. Daamen, S. Hoogendoorn, and B. Arem, "Potential impacts of ecological adaptive cruise control systems on traffic and environment," *IET Intell. Transp. Syst.*, vol. 8, no. 2, pp. 77–86, Mar. 2014.

[31] L. Ng, C. M. Clark, and J. P. Huissoon, "Reinforcement learning of adaptive longitudinal vehicle control for dynamic collaborative driving," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 907–912.

[32] C. Desjardins and B. Chaib-Draa, "Cooperative adaptive cruise control: A reinforcement learning approach,' *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1248–1260, Dec. 2011.

[33] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach," *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 114030.

[34] Y. Lin, J. McPhee, and N. L. Azad, "Comparison of deep reinforcement learning and model predictive control for adaptive cruise control," *IEEE Trans. Intell. Veh.*, vol. 6, no. 2, pp. 221–231, Jun. 2021.

[35] H. Shi, Y. Zhou, K. Wu, X. Wang, Y. Lin, and B. Ran, "Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment," *Transp. Res. Part C, Emerg. Technol.*, vol. 133, Dec. 2021, Art. no. 103421.

[36] H. Zhou et al., "Review of learning-based longitudinal motion planning for autonomous vehicles: Research gaps between self-driving and traffic congestion," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2676, no. 1, pp. 324–341, Jan. 2022.
[37] R. Hallouzi, V. Verdult, H. Hellendoorn, and J. Ploeg, "Experimental evaluation of a co-operative driving setup based on inter-vehicle communication," *IFAC Proc. Vol.*, vol. 37, no. 8, pp. 126–131, Jul. 2004.
[38] C. Wu, Y. Lin, and A. Eskandarian, "Cooperative adaptive cruise control with adaptive Kalman filter subject to temporary communication loss," *IEEE Access*, vol. 7, pp. 93558–93568, 2019.
[39] J. Mirwald, J. Ultsch, R. de Castro, and J. Brembeck, "Learning-based cooperative adaptive cruise control," *Actuators*, vol. 10, no. 11, p. 286, Oct. 2021.
[40] H. Shi, Y. Zhou, X. Wang, S. Fu, S. Gong, and B. Ran, "A deep reinforcement learning-based distributed connected automated vehicle control under communication failure," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 15, pp. 2033–2051, Dec. 2022.
[41] Y. Zhou, S. Ahn, M. Chitturi, and D. A. Noyce, "Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty," *Transp. Res. Part C, Emerg. Technol.*, vol. 83, pp. 61–76, Oct. 2017.
[42] B. Sakhdari, E. M. Shahrivar, and N. L. Azad, "Robust tube-based MPC for automotive adaptive cruise control design," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2017, pp. 1–6.
[43] J. Wei, J. M. Dolan, J. M. Snider, and B. Litkouhi, "A point-based MDP for robust single-lane autonomous driving behavior under uncertainties," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2011, pp. 2586–2592.
[44] Z. Qiao, K. Muelling, J. Dolan, P. Palanisamy, and P. Mudalige, "POMDP and hierarchical options MDP with continuous actions for autonomous driving at intersections," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 2377–2382.
[45] K. Mani, M. Kaushik, N. Singhania, and K. M. Krishna, "Learning adaptive driving behavior using recurrent deterministic policy gradients," in *Proc. IEEE Int. Conf. Robot. Biomim. (ROBIO)*, 2019, pp. 2092–2098.
[46] M. Zhou, X. Qu, and X. Li, "A recurrent neural network based microscopic car following model to predict traffic oscillation," *Transp. Res. Part C, Emerg. Technol.*, vol. 84, pp. 245–264, Nov. 2017.
[47] S. Albeaik, T. Wu, G. Vurimi, F.-C. Chou, X.-Y. Lu, and A. M. Bayen, "Deep truck cruise control," *Control Eng. Pract.*, vol. 121, Apr. 2022, Art. no. 105026.
[48] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, *arXiv:1707.06347*.
[49] 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, Aug. 2019.
[50] S. E. Shladover, "Longitudinal control of automated guideway transit vehicles within platoons," *J. Dyn. Syst. Meas. Control*, vol. 100, no. 4, pp. 302–310, Dec. 1978.
[51] L. Xiao and F. Gao, "Practical string stability of platoon of adaptive cruise control vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1184–1194, Dec. 2011.
[52] M. Wang, S. P. Hoogendoorn, W. Daamen, B. van Arem, B. Shyrokau, and R. Happee, "Delay-compensating strategy to enhance string stability of adaptive cruise controlled vehicles," *Transp. B, Transp. Dyn.*, vol. 6, no. 3, pp. 211–229, Jul. 2018.
[53] T. Li, D. Chen, H. Zhou, J. Laval, and Y. Xie, "Car-following behavior characteristics of adaptive cruise control vehicles based on empirical experiments," *Transp. Res. Part B, Methodol.*, vol. 147, pp. 67–91, May 2021.
[54] G. Newell, "A simplified car-following theory: A lower order model," *Transp. Res. Part B, Methodol.*, vol. 36, no. 3, pp. 195–205, Mar. 2002.
[55] J. A. Laval and L. Leclercq, "A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 368, no. 1928, pp. 4519–4541, Oct. 2010.

YING-CHUAN NI received the B.Sc. degree in engineering from National Taiwan University, Taipei, Taiwan, in 2019, and the M.Sc. degree in civil engineering from the Delft University of Technology, Delft, The Netherlands, in 2022. He is currently pursuing the Doctoral degree with ETH Zürich, Zürich, Switzerland. His research interests include traffic flow theory, macroscopic traffic modeling, microscopic traffic simulation, traffic control, and urban traffic management. His current research focuses on multimodal traffic modeling for congested urban networks.

VICTOR L. KNOOP received the M.Sc. degree in physics from Leiden University in 2005, and the Ph.D. degree from the Delft University of Technology in 2009 on the effects of incidents on driving behavior and traffic congestion. He worked amongst others with the University of Lyon, Imperial College London, and the University of California at Berkeley. Since 2012, he holds a professorship with the Transport and Planning Department, Delft University of Technology (currently Associate Professor, tenured since 2018). He is currently the Co-Director of the Traffic Dynamics Modelling and Control Laboratory. His main research interest lies in the interactions between vehicles and how these interactions create patterns in traffic.

JULIAN F. P. KOOIJ (Member, IEEE) received the Ph.D. degree in visual detection and path prediction for vulnerable road users from the University of Amsterdam, Amsterdam, The Netherlands, in 2015. Afterward, he joined the Delft University of Technology, first with the Computer Vision Lab, and later with the Intelligent Vehicles Group, where he is currently an Associate Professor, and the Co-Director of the 3-D Urban Understanding AI Lab. His research interests include deep representation learning and probabilistic models for multisensor localization, object detection, and forecasting of urban traffic.

BART VAN AREM (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in applied mathematics from the University of Twente, The Netherlands, in 1986 and 1990, respectively. From 1991 to 2009, he worked with TNO. From 2003 to 2012, he was a part-time Full Professor with the University of Twente. He was appointed as a Full Professor of Transport Modeling with the Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, TU Delft in 2009. He was the Head of the Department of Transport and Planning from 2010 to 2017 and served as the Director of the TU Delft Transport Institute from 2012 to 2021. He has been serving as a Pro Vice Rector Magnificus for Doctoral Affairs of TU Delft since 2021. He is the Founding Editor-in-Chief of the IEEE OPEN JOURNAL OF INTELLIGENT TRANSPORTATION SYSTEMS.