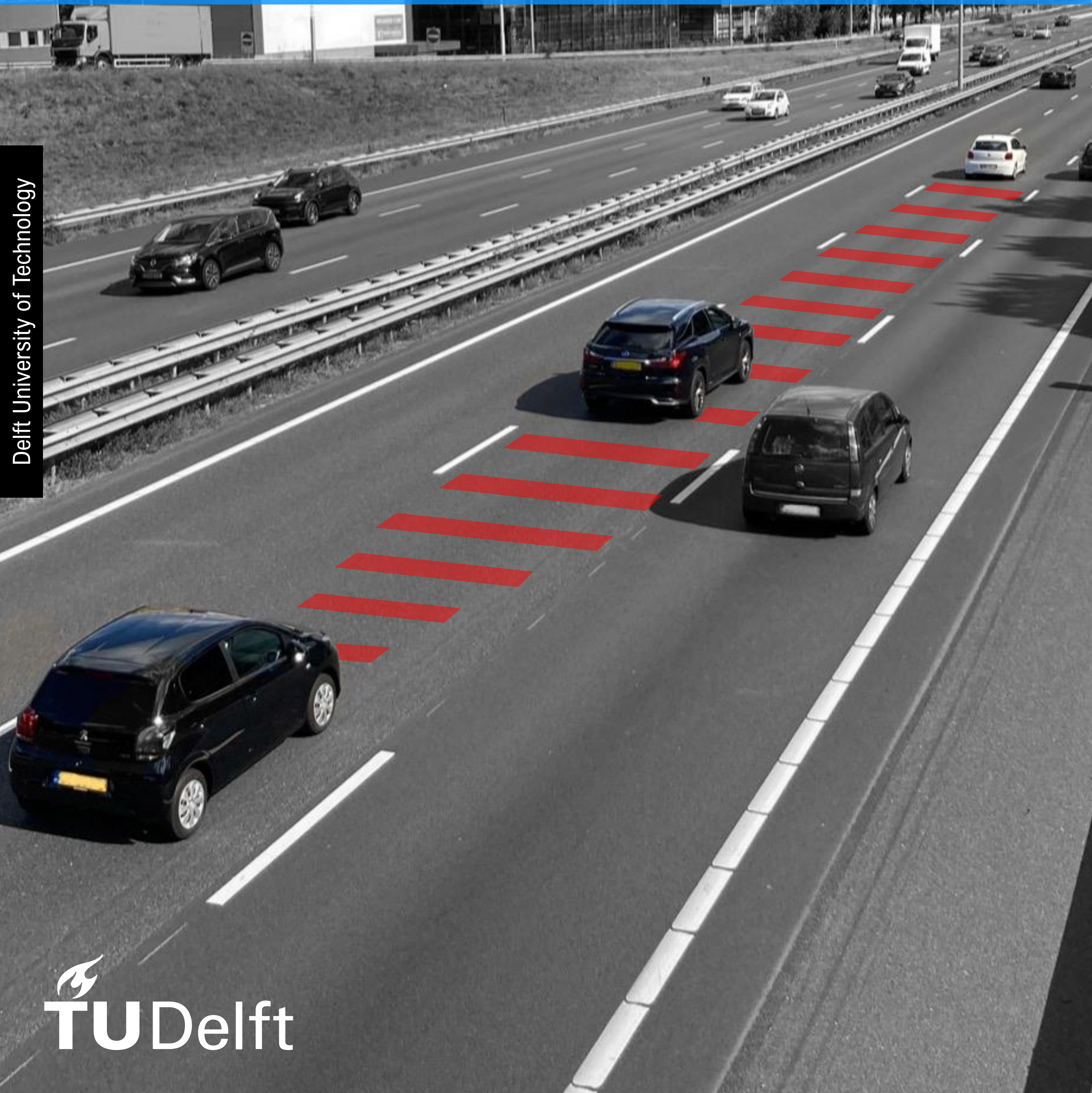
Multi-leader Adaptive Cruise Control Systems considering Sensor Measurement Uncertainties based on Deep Reinforcement Learning

MSc thesis Report

Ying-Chuan Ni



Delft University of Technology

TU Delft

Multi-leader Adaptive Cruise Control Systems considering Sensor Measurement Uncertainties based on Deep Reinforcement Learning

MSc Thesis Report

by

Ying-Chuan Ni

to obtain the degree of Master of Science in Civil Engineering

at Delft University of Technology,

to be defended publicly on August 25th, 2022 at 16:00h.

Student number:	5235391	
Project duration:	December 7th, 2021 – August 11th, 2022	
Thesis committee:	Prof. dr. ir. Bart van Arem,	CiTG, TU Delft, chair
	Dr. Victor Knoop,	CiTG, TU Delft, daily supervisor
	Dr. Julian Kooij,	3mE, TU Delft, external supervisor

The source code created in this thesis is published and can be found at
http://doi.org/10.4121/20436090.

Preface

Dear reader,

You are reading my master thesis, which concludes my two-year study for the MSc degree at TU Delft. Traffic flow, autonomous driving, and artificial intelligence are three fascinating areas to me. I am very glad that I could work on a topic which incorporates these three elements at the same time so that I could acquire the relevant knowledge and skills. This report shows you my dedication to this study and the great amount of effort I have made in the last eight months.

I would like to express my gratitude to the committee members of this thesis project. The first and most important thanks to Victor Knoop, my daily supervisor. He could always foresee many possible problems I would encounter and point out a lot of detail which I overlooked during the project. His helpful opinions guided me to explore and shape this study. The next thanks goes to Julian Kooij for being very supportive in the committee meetings and providing me with a lot of important comments with his expertise. Finally, I want to thank Bart van Arem for chairing this project. Every discussion we had really helped me move forward and enhance the quality of this work. It has been a unique and fabulous experience working with them.

Pursuing a master's at TU Delft is not an easy task. I started this master's study during the period of COVID-19, which made it even more challenging. It was really many people's participation which made this whole process possible. I would thereby like to thank my parents and my sister, who allowed me to study here and checked in on me through regular phone calls. Next, a big thanks to my girlfriend. We studied and lived here in the Netherlands together and helped each other with all the difficulties. Without her company, I could not have had such a great time here. Of course, I would not forget how joyful it was spending time hanging out with all my friends here at Delft. Furthermore, I would like to thank all the TDMaC Lab members. I enjoyed those presentations and highs-and-lows sharing. They brightened up my life at TU Delft. Last but not least, I appreciate all my friends in Taiwan who have been in contact with me in the past two years. Your presence has always been supporting me a lot. Perhaps a special thanks to myself for choosing TU Delft two years ago. Looking back on all the things I learned, culture I experienced, and people I met here, I genuinely think this is a very meaningful decision in my life.

The master degree and thesis together represent a milestone in my career. However, they do not signify the end of my journey of either being a student or doing research. I will carry on this journey with my passion and ambition to become a prominent expert in this study field. Thank you very much. Hope you enjoy reading the thesis.

Ying-Chuan Ni
The Hague, August 2022

Abstract

Adaptive Cruise Control (ACC) relieves human drivers' tasks by taking over the control of the throttle and braking of the vehicles automatically. However, it has been demonstrated in many empirical studies that current production ACC systems fail to guarantee string stability. It is believed that if vehicles can take the longitudinal dynamics further downstream into account and react to the propagating disturbance earlier, the string stability in the platoon may be improved. Instead of relying on inter-vehicle communication technologies, the ego-vehicle should be able to detect the second leading vehicle by leveraging the power of on-board sensors. Still, the second leader measurements can be highly erroneous. Therefore, it is important to consider the entailed measurement uncertainties when designing and evaluating such ACC systems. This study proposes several ACC systems which possess the property of multi-anticipation and uncertainty handling.

The possible sensor technology which can collect the second leader measurements is first investigated. Based on the considered setup, the measurement uncertainties are modelled to reflect the real-world conditions. The ACC system architecture and control system design method are then proposed. Deep reinforcement learning is applied for the controller design in light of its great potential in describing the complex non-linear control task and handling the uncertainties. Kalman filters and recurrent policies with a Long-Short-Term-Memory network are applied to cope with uncertain measurements. The first method estimates the state information before feeding it back to the controller agent, while the latter incorporates the state estimator into the controller to actively consider the uncertainties while making decisions.

A numerical simulation approach is adopted to theoretically assess the performance of the proposed ACC systems. A traffic disturbance event and multiple levels of measurement noise are considered in the experiment. To analyze the performance in terms of string stability and ride comfort and understand the car-following behavior mechanism resulted from the proposed systems, a quantitative analysis framework is developed.

The evaluation results demonstrate the applied learning-based approach succeeds to train ACC control policies which can ensure string stability. It is also found that the multi-anticipation ability significantly improves the string stability and ride comfort performance. In the scenarios with measurement noise, systems using the tuned Kalman filters exhibit the ideal level of string stability performance. However, ride comfort cannot be guaranteed in scenarios with large measurement noise. On the other hand, systems using recurrent policies can better ensure ride comfort performance while maintaining string stability at certain levels. Based on the results, the performance limits of the proposed ACC systems in the handling of measurement uncertainties are explored. In addition, with the different policy training setups, the trade-off between these two performance aspects is shown.

The findings of this study are anticipated to trigger the development of advanced multi-leader ACC system by automakers, sensor manufacturers, and traffic engineers. Future work can be directed to an enhanced controller design. Robustness of the systems with respect to other sources of measurement uncertainties, more types of traffic disturbance, and platoon heterogeneity is worth further design consideration and analysis.

Contents

List of Tables

List of Figures

1 Introduction

This chapter introduces the background of the research topic and the problem definition. Several research questions are proposed according to the research problem defined. The scope and overall structure of the study are also described at the end.

1.1. Background

Autonomous vehicles (AV), or self-driving vehicles, have gained a lot of attention from the truck industry, public transport agencies, and private car users in recent years. It is believed that AVs can improve the road safety by removing human errors, provide mobility to people with inconvenience, increase the traffic efficiency by optimizing driving behaviors, and hence reduce the vehicle emissions (European Commission, 2018). Many studies have tried to predict the market penetration rate of AVs and timeline of AV-related development. Milakis et al. (2017) stated that the fully automated vehicles are expected to be available between 2025 and 2045. However, various advanced driver assistance systems (ADAS), each representing a level of automation, have already been widely implemented on commercial vehicles nowadays.

Typical functionalities in ADAS include Adaptive Cruise Control (ACC), lane keeping assistance, forward collision warning and avoidance, automated emergency braking, and parking assistance, etc. This study focuses on the design of ACC, which is one of the most commonly-discussed applications in ADAS. It adapts the vehicle acceleration according to the desired spacing policy by using the measurements collected by on-board sensors in the driving environment. The development of this vehicle automation functionality achieves the Society of Automotive Engineers (SAE) level 1 automation. Various ACC systems have already been widely-implemented on commercial vehicles nowadays. Calvert et al. (2017) 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 speed of the vehicles automatically. According to the measurements collected from on-board sensors, ACC systems reduce the gap deviation between the preceding vehicle and the ego-vehicle from the desired spacing policy.

Early ACC systems have already shown their positive effect on improving ride comfort and reducing fuel consumption (Marsden et al., 2001; Xiao and Gao, 2010). However, when looking at the system from a traffic engineering perspective, the automated control over the vehicle longitudinal dynamics provided by ACC systems can greatly influence the traffic flow efficiency and stability. The effect of ACC systems on traffic flow performance has received a lot of attention from both the automotive industry and the research community of intelligent vehicles (T. Li et al., 2022; Makridis et al., 2020; Spiliopoulou et al., 2018; VanderWerf et al., 2002). In particular, the results in van Arem et al. (1996) and Calvert et al. (2017) both showed that the presence of AVs does not necessarily bring positive effect on traffic flow and road capacities in scenarios with high traffic demand or certain ACC penetration rates.

When multiple vehicles equipped with ACC systems drive along each other by following the behaviors of their predecessors, an ACC vehicle platoon is formed. It was demonstrated in many field experiments and empirical studies that platoons consist of vehicles equipped with current commercial ACC systems could not ensure string stability, indicating that the disturbance caused by the preceding vehicle would be amplified as it propagates upstream along the platoon (Ciuffo et al., 2021; Gunter et al., 2021; Knoop et al., 2019). Marsden et al. (2001) pointed out a specific case which could result in such a platoon instability. It was found that when a vehicle cut-in to the head of the platoon, a deceleration wave could propagate along the platoon and hence formed a shockwave when the first leading vehicle applies a braking. The factors leading to the amplification of shockwave disturbance, which is the so-called string instability, could be the reaction delay of following vehicles, e.g. actuator time lag and sensor delay, determined spacing policy, and parameters in the ACC controller (T. Li et al., 2021; Makridis et al., 2020; J. Zhou and Peng, 2005). Moreover, the problem of string instability not only induces shockwaves which degrade the traffic flow efficiency but also leads to increased energy consumption and unsafe traffic situations in extreme cases. The vehicles at the tail of an unstable platoon would even experience ACC disengagement or come to a complete stand still when its braking motion is too obtrusive. With the increasing number of ACC-equipped vehicles on public roads, this problem needs to be addressed.

To mitigate the amplification of traffic disturbances and improve the string stability of AV platoons, the benefit of cooperative ACC (CACC) systems which employ wireless inter-vehicle communication technologies, the so-called Vehicle-to-Vehicle (V2V) communication, has been frequently discussed and investigated in recent years. When individual vehicles can obtain and utilize the in-car information of its direct preceding vehicle, e.g. determined vehicle acceleration in the next time step, or even information of numerous preceding vehicles, the string stability of the platoon can be guaranteed. The improved traffic flow performance brought by CACC systems has been demonstrated by many studies using theoretical analysis and simulation approaches (Ploeg et al., 2014; Pueboobpaphan and van Arem, 2010; Schakel et al., 2010; Shladover et al., 2012).

As mentioned above, major research endeavor has evolved from ACC to CACC systems for the sake of string stability performance. However, the employment of CACC systems heavily relies on communication technologies to achieve information-sharing between vehicles (Shladover et al., 2015). To ensure string stability, a certain penetration rate of connected and autonomous vehicles (CAVs) is required, which is still rather difficult to achieve nowadays. In addition, there are many concerns regarding the adoption of wireless communication technologies, such as its unreliability, privacy issues, and the risk of cyber attacks. The consequence after communication breakdown is also one of the critical problems which should be considered. Hence, there are still many difficulties for practitioners to tackle before CACC systems can be broadly accepted and implemented.

1.2. Problem statement

According to the aforementioned difficulties and concerns, we seek to improve the string stability performance of ACC systems by exploiting the capability of sensors instead of using communication technologies to achieve any level of cooperative behaviors. If the measurements of leaders further ahead can be collected from on-board sensors, the vehicle can respond to the downstream car-following dynamics earlier and more accurately to prevent from overreacting to the disturbance. This also resembles the behavior of human drivers who look at more than one vehicle ahead to adapt their car-following behaviors, which was demonstrated by Hoogendoorn et al. (2006) and Ossen (2008). This kind of driving behavior, which enhances drivers' situation awareness, is called multi-anticipation. For AVs, Gorter (2015) also emphasized the importance of designing an ACC system which can look at more than one predecessor so that the traffic safety can be improved even when the ACC-equipped vehicles are driving at a relatively short headway.

A few studies have already designed such kind of CACC systems which possess the property of the multi-anticipation by looking at multiple predecessors to control the longitudinal motion of the ego-vehicle (Dollar et al., 2021; Hasebe et al., 2003; Wang, Daamen, Hoogendoorn, et al., 2014a; Wilmink et al., 2007). Given the current development of sensor technologies, we believe that this kind of autonomous driving ability should be technically achievable even without the help of communication technologies. ACC-equipped vehicles should be able to collect information from not only the direct leader but also those occluded leaders further downstream by using on-board sensors. Information related to how the state-of-the-art sensor technologies can achieve a non-line-of-sight detection and perceive a completely occluded object will be explained in section 2.4.

After implementing this detection functionality into AVs, its ACC system can then incorporate and utilize the measurements of the downstream leaders to determine the acceleration command in the next time step. In the first part of this study, we focus on the design of the controllers of the ACC systems with the assumption of accurate sensor measurements. However, different from CACC systems using vehicle communication technologies to obtain information, leveraging sensors to collect measurements from leaders further downstream may entail relatively high sensor measurement uncertainties. These uncertainties may affect string stability and ride comfort. The design of ACC systems which are aware of the uncertainties and can properly react to them is the other focal point in this study. The influence of the proposed systems on the platoon performance in terms of string stability and ride comfort will be assessed.

1.3. Research questions

This study aims to propose new ACC systems to control the car-following behavior of AVs and assess their influence on string stability and ride comfort performance in the platoon. Several research questions are developed and categorized into six groups. The research methodology should help answer these questions.

The first and second groups of research questions focus on the design of multi-leader ACC systems considering sensor measurement uncertainties. The second group again emphasizes the importance of modelling and handling of measurement uncertainties. To comprehensively evaluate the performance of the proposed systems, the third, fourth, and fifth question groups are developed. Each of them tries to explore one of the aspects of the system performance. Continuing from the third and fourth question groups regarding string stability and ride comfort performance, the last question group considers the influence of measurement uncertainties. It seeks to explore the performance limit of the proposed multi-leader ACC system designs and provide a reference for the required level of sensor measurement accuracy for the multi-leader detection functionality.

Design-related questions:

- What kind of control method for ACC systems has the potential to outperform other types of controllers in terms of string stability and the handling of measurement uncertainties? How to design the ACC controllers using the selected control method? Which factors can and should be considered in the control system design?
- What is the proper way to model and simulate the measurement uncertainties so that the real-world autonomous driving conditions can be reproduced? When it is known that the sensor measurements are erroneous, what kind of methods can be used by the ACC systems to cope with the uncertainties??

Performance-related questions:

- What is the criteria of string stability in this study? Which indicators can be used to appropriately evaluate the string stability performance of the proposed ACC systems?
- While aiming at preserving string stability, can the system still maintain a certain level of ride comfort for the platoon? How to quantitatively analyze the ride comfort performance of the proposed systems?
- What is the benefit of multi-anticipation for the ACC vehicle platoon? How to explore the positive effect of the proposed multi-leader ACC system compared to the one-leader system?
- What is the influence of the considered measurement uncertainties on the ACC system performance? What is the measurement uncertainty boundary for the proposed systems within which the desired performance in terms of string stability and ride comfort can still be preserved?

1.4. Research scope

The problem domain and the limitations of this study regarding the ACC system design, sensor technology setup, vehicle motion model, and simulated driving environment, are described in this section.

Vehicles equipped with an ACC system usually have a hierarchical architecture which consists of an upper-level controller and a lower-level controller. The upper-level provides the desired acceleration, while the lower-level mechanically controls the throttle and brake command. H. Zhou et al. (2021) emphasized the importance of the lower-level ACC controller for string stability performance. However, this study focuses on the upper-level controller, which determines the vehicle acceleration in the next decision time step. Therefore, it is assumed that the lower-level controller and vehicle internal driveline can react precisely according to the vehicle acceleration command from the upper-level. In addition, other external factors, including aerodynamics, rolling resistance, and road geometry, are not considered in this context.

Given the current development of the automotive exteroceptive sensors, AVs are expected to be able to detect at least two leaders ahead, which are the direct preceding vehicle and the pre-preceding vehicle (herein after referred to as the first and second leaders, respectively). Figure 1.1 illustrates the possible detection functionality setup when using RADAR as the major sensor for the detection task. The blue radio waves represent the detection of the first leader, while the green waves propagate to the second leader through diffraction or reflection under the bottom of the first leader. In this study, following vehicles can only collect preceding information via on-board sensors. Cooperative behaviors or centralized platoon control which requires the utilization of inter-vehicle communication technologies is not considered in the defined scope.

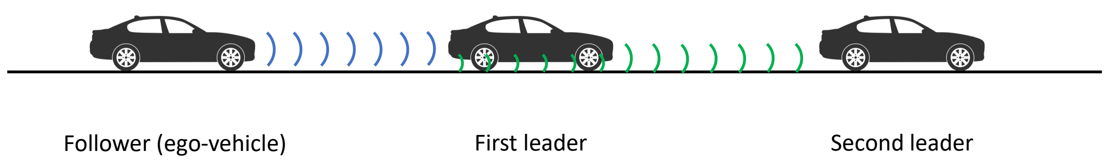


Figure 1.1: Illustration of multi-leader detection functionality using a RADAR

This study seeks to analyze the string stability of a single lane car-following problem on motorways, which indicates that only a single vehicle platoon is considered. Multi-lane driving behaviors, such as lane-changing or merging/diverging behaviors of vehicles in the platoon, are excluded. Following this setting, false positive alarms in the detection task are not considered. This also removes the concern of the on-board sensor capability to distinguish whether the detected second leader is on the same lane with the ego-vehicle or the adjacent lanes, which is a large difficulty for the vehicle perception system when the detected object is far away or the surrounding driving environment is complicated.

1.5. Research outline

The remainder of this study is summarized as follows. Chapter 2 reviews articles regarding the ACC/CACC system development, the definition and analysis approach of the string stability of a vehicle platoon, sensor technology development, and the modelling and handling of measurement uncertainties. Chapter 3 then introduces the proposed control system architecture and explains the methods and tools applied to design the whole control system in detail.

In chapter 4, several experimental scenarios which can be used to properly test the performance of the proposed ACC systems are created. Although various ACC or CACC controllers utilizing information from multiple predecessors have been proposed, there is still little understanding of the collective performance and induced effect of a vehicle longitudinal control system with the multi-anticipation capability. Therefore, chapter 4 also aims to propose a quantitative analysis framework to explore the string stability, ride comfort performance, and behavior mechanism of the vehicles equipped with the ACC systems when facing a traffic disturbance. Chapter 5 then presents and compares the evaluation results of each system in the experiment. Figure 1.2 provides an overview for this study.

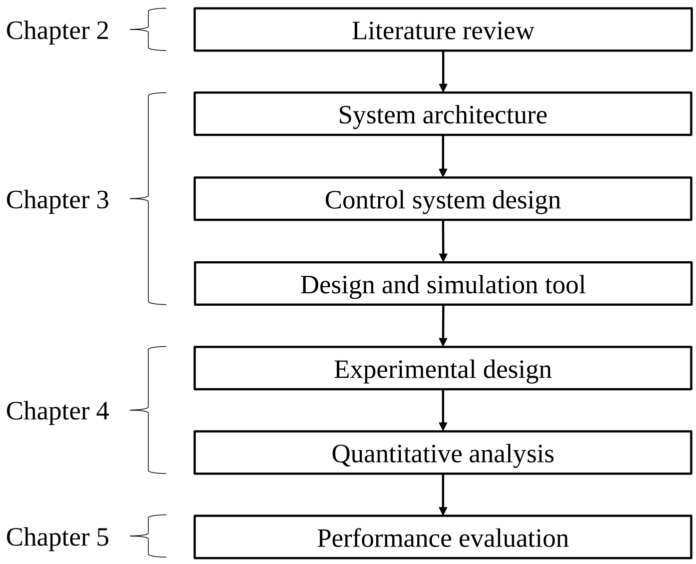


Figure 1.2: Research outline

At the end of the report, chapter 6 first summarizes all the findings based on the performance evaluation results. The results are also compared to the previous studies and the initial hypothesis. The limitations of the methodology in this study are also discussed in this chapter. Chapter 7 then concludes the study. It again recapitulates the major research findings and answers those proposed research questions. After giving the conclusion of the whole study, implications from both the scientific and practical perspectives and recommendations for relevant and possible future work are drawn to deliver messages to other interested researchers.

2
Literature Review

The research focuses on the design of multi-leader ACC systems and its potential effect on string stability and ride comfort. This chapter first reviews the literature regarding the state-of-the-art ACC systems development and modelling. In the second section, the concept of string stability and its different analysis approaches are introduced. The third section describes the benefit of multi-anticipative car-following behaviors for traffic and the ACC/CACC systems proposed in the past which possess such property. The last two sections discuss the suitable sensor technology setup for the multi-leader ACC driving task by reviewing the development of commonly-used automotive sensors and the modelling and handling of sensor measurement uncertainties in past studies. Based on the previous findings, the summary section at the end of this chapter determines the overall methodology for this study.

2.1. Adaptive cruise control

ACC has been one of the most popular research topics regarding ADAS in recent years. The earliest development of ACC systems can be viewed as an extension of conventional cruise control systems. In order to enhance driving comfort and convenience, numerous research projects and experiments were conducted by experts and vehicle manufacturers in both the US and Europe around 1990's to facilitate the development of ACC systems (Marsden et al., 2001; Xiao and Gao, 2010). After more than 30 years of development, automakers have now started to extend the availability of ACC systems from premium vehicles to middle-class commercial vehicles.

Typical ACC controllers are mathematically modelled as a state feedback controller in the research field. The goal of the ACC controller is to maintain the distance between vehicles according to a specified spacing policy. The controller determines the vehicle acceleration (control signal) and sends it to the lower-level controller. The internal state of the ego-vehicle and surrounding information would be collected by on-board sensors and feedback to the controller as the next control input, which forms the closed-loop structure.

CACC systems are the evolution of ACC systems with the help of inter-vehicle communication technologies. Because of its great potential in ensuring driving safety and efficiency, their development is gaining more attention than that of the ACC system in recent years. To follow the state-of-the-art ACC-related development, this section also includes many studies regarding the design and assessment of CACC.

With the improvement of online optimization algorithms and the increasing popularity of machine learning in the field of control system, ACC systems which apply these various control methods have also been proposed. The systems can, therefore, be categorized by the control methods adopted. This section covers three major types of control methods which were used for the ACC controller design.

2.1.1. Linear and non-linear state feedback control

The most studied ACC controllers use linear state feedback control logic. Typically, the system state includes the range error, which is the difference between the current physical gap and the desired gap based on the spacing policy, and the range rate, which is the relative speed between the preceding vehicle and the ego vehicle. The distance and speed measurements are collected from on-board sensors and fed back to the controller to determine the next vehicle acceleration after every time step. The feedback gains in the linear control logic are specified to represent the sensitivity of the two components in the system state to the next vehicle acceleration. Its formulation is pretty much similar to a proportional-derivative (PD) controller. In VanderWerf et al. (2001) and VanderWerf et al. (2002), the vehicle acceleration in the car-following mode can be calculated by the control law

$$a_{i,k} = k_1 \cdot e_{i,k} + k_2 \cdot (v_{i-1,k-1} - v_{i,k-1}), \tag{2.1}$$

where $e_{i,k}$ is the gap error between vehicle i and vehicle $i-1$ at time step k, as calculated by

$$e_{i,k} = x_{i-1,k-1} - x_{i,k-1} - t_d \cdot v_{i,k-1}. \tag{2.2}$$

In Eq 2.2, $x_{i-1,k-1}$, $x_{i,k-1}$, $v_{i-1,k-1}$, and $v_{i,k-1}$ are the position and speed of the preceding vehicle and ego-vehicle in the previous time step. t_d is the desired time gap based on the spacing policy specified. k_1 and k_2 represent the feedback gains of the gap error and relative speed, respectively.

Stemmed from the above mentioned linear state feedback model, Xiao et al. (2017) proposed an empirical ACC model. The gap error in the controller is calculated by

$$e_{i,k} = x_{i-1,k-1} - x_{i,k-1} - d_0 - t_d \cdot v_{i,k-1}, \tag{2.3}$$

where d_0 represents an additional dynamic spacing margin determined based on the speed of the ego-vehicle to prevent rear-end collision, as given by

$$d_0 = \begin{cases} \text{vehicle length} & \text{if } v_{i,k-1} \geq 15\text{m/s} \\ \frac{75}{v_{i,k-1}} & \text{if } 10.8 \leq v_{i,k-1} \leq 15 \\ \text{vehicle length} + 2 & \text{if } v_{i,k-1} < 10.8. \end{cases} \tag{2.4}$$

Similar to this model, many other ACC car-following models in the literature are variants of this linear state feedback formulation.

CACC systems can achieve better performance than normal ACC systems by utilizing more information obtained from vehicle communication technologies. In van Arem et al. (2006), the proposed CACC systems modified the linear state feedback car-following model of the original ACC systems by adding the lead vehicle acceleration, an in-vehicle information, into the control law

$$a_{i,k} = k_0 \cdot a_{i-1,k-1} + k_1 \cdot e_{i,k} + k_2 \cdot (v_{i-1,k-1} - v_{i,k-1}). \tag{2.5}$$

The term $a_{i-1,k-1}$ in Eq 2.5 represents the lead vehicle acceleration. With this information, the controller can react to the front vehicle behavior more accurately and thus decrease the minimum desired time gap between vehicles, which increases the traffic flow capacity more significantly (van Arem et al., 2006; VanderWerf et al., 2001; VanderWerf et al., 2002).

In addition to the linear models, others have attempted to apply nonlinear car-following models, which are originally used for modelling human driven vehicles, to both ACC and CACC systems. Hasebe et al. (2003) proposed a cooperative driving control system by extending the Optimal Velocity Model (OVM) to include not only multiple preceding vehicles but also vehicles behind. The acceleration is determined by the difference between the current speed and an optimal speed calculated by a nonlinear function of headways with other vehicles. However, it is pointed out that the OVM cannot ensure collision-free, which makes it less applicable for autonomous vehicles. Kesting et al. (2008) used Intelligent Driver Model (IDM) to represent ACC vehicles. The benefit of using IDM over linear state feedback models is its ability to create a more human-like and comfortable driving maneuver. It was further enhanced to solve the hard-braking behavior when facing a cut-in situation in dense traffic (Kesting et al., 2010). The model was again updated to prevent the originally unrealistic behavior caused by a speed which exceeds the desired speed (Treiber and Kesting, 2013). The acceleration command generated by the enhanced IDM is calculated by

$$a_{IDM} = a\left[1 - \left(\frac{v}{v_0}\right)^{\delta} - \left(\frac{s_0 + max\left[0, vT + \frac{v\Delta v}{2\sqrt{ab}}\right]}{s}\right)\right], \tag{2.6}$$

where v represents the current speed of the ego-vehicle, v_0 is the desired speed, Δv calculates the error between the desired speed and current speed, s denotes the current distance gap, s_0 is the distance gap in standstill conditions, and T indicates the desired time gap. a and b are parameters representing the maximum acceleration and minimum comfortable deceleration, respectively. δ is the parameter for the free acceleration exponent, which is usually set to $\delta = 4$.

Two most commonly used ACC controller formulations in the literature were introduced above. In VanderWerf et al. (2001), the linear models were used to develop new simulation methods to test the effect of driver assistant systems on traffic flow dynamics. These linear state feedback models were further discussed and used in many types of ACC-related studies to investigate the string stability and traffic flow effect of the ACC system. For instance, continuing from their previous study, VanderWerf et al. (2002) estimated the highway capacity for different market penetration rate of ACC and CACC vehicles with linear controllers through a stochastic simulation approach. There are also several studies using IDM models as the ACC car-following models to evaluate the performance of ACC systems. Milanés and Shladover (2014) even implemented an IDM controller into the ACC system of two commercial vehicles to conduct a car-following experiment. Spiliopoulou et al. (2018) used IDM as the ACC car-following model to propose a real-time driving behavior adaptation control strategy which changes the time gap setting to improve the traffic flow efficiency. There are also many studies exploring the string stability of (C)ACC systems based on these models, which will be discussed in section 2.2.

Many studies tried to calibrate the parameters of these ACC car-following models by using empirical data to reproduce the behavior of commercial ACC vehicles to evaluate the state-of-the-art development. James et al. (2019) provided a comprehensive investigation of the linear and IDM ACC car-following models by first calibrating them according to data collected from a 2013 Cadillac SRX equipped with a production ACC system. A simulation approach using VISSIM was then adopted to evaluate the influence of penetration rate and following headway on macroscopic traffic flow performance. Blauw (2019) also tried to calibrate the ACC system of an Audi A4. It was found that the ACC system exhibits non-linear driving behavior, which cannot be represented by a linear model. Shang and Stern (2021) also followed this kind of framework and pointed out the contradicting results in terms of string stability and bottleneck capacity between using the model in the literature and using the calibrated ACC controller. To further enhance the accuracy of ACC models in reproducing the behaviors of commercial systems, He et al. (2022) aimed to augment the linear controllers, IDM, and Gipps' car-following model with physics-based extensions, including perception delay, non-linear dynamics, and acceleration constraints. The vehicle trajectory data collected from the field experiment conducted at AstaZero test track were used to calibrate the proposed augmented models. However, the results did not show much improvement in terms of modelling accuracy compared to those previously-proposed models before being augmented. In addition, it was found that the best model to represent the commercial ACC systems is independent of the brand of the vehicles.

2.1.2. Model predictive control

Model predictive control (MPC), or receding horizon control, is one of the most popular feedback control methods nowadays. In recent years, many studies have been investigating ACC controller designs based on MPC. By using an optimization approach, various objectives and requirements regarding ecology, efficiency, stability, safety, and limitation of vehicle dynamics, can be considered in the model simultaneously. The advantage of feedback control can also be shown by using a prediction model to account for the future system dynamics.

In MPC, the control signal is determined by solving a minimization problem with a cost function containing all the objectives and constraints of the system behavior over a finite time horizon. Corona and De Schutter (2008) proposed an MPC-based control method for ACC system of a vehicle. The cost function includes the number of gear switchings, control input variation, and deviation from the leader trajectory. The operational range of the speed, the largest position deviation with the leader, and comfort acceleration boundary are designed in the constraints. The simulation of vehicle motion considering the air drag and tire friction was also used as the prediction model in the MPC framework. The trade-off between the performance and computation requirements compared with other control methods was discussed. Naus et al. (2008) applied MPC to design the ACC Stop-&-Go functionality. The primary objective of the MPC-based controller was to follow the leader at a desired distance according to the constant headway while the traffic efficiency and comfort being the secondary objective. The optimization model formulation allows the tuning of the resulting controller by changing a weighting factor in a formulated performance function related to traffic efficiency and comfort. As highlighted in these two studies, the MPC-based ACC controller design possesses the ability to consider multiple objectives.

Exploiting this merit of optimal control method, the MPC-based ACC controller has also been widely adopted to achieve ecological objectives in these years. Wang, Daamen, Hoogendoorn, et al. (2014) designed an Ecological Adaptive Cruise Control system considering the CO_2 emission rate via a function of vehicle speed. The optimization problem was solved by a dynamic programming approach. It was found that the eco-driving strategy produces a smoother acceleration profile. At the macroscopic level, a trade-off between traffic efficiency and emission rate in free flow traffic conditions, while increasing traffic efficiency and reducing emission rate can bring similar outcome in congested situations. Wang, Daamen, Hoogendoorn, et al. (2014b) proposed a model which enables the controlled vehicle to consider the future behaviors of other adjacent vehicles. In the next part, Wang, Daamen, Hoogendoorn, et al. (2014a) further extended the MPC framework into a cooperative system setting, in which vehicles equipped with the MPC controller utilize V2V communication to increase the situation awareness and achieve a global performance. The work demonstrated the flexibility of MPC in formulating into various control strategies.

Other than those studies which incorporated ecological objectives into the cost function of the ACC model, there were also several studies applying MPC to handle the uncertainties lying in the received information and the system dynamics of the ego-vehicle itself. Moser et al. (2017) applied a stochastic MPC approach for the development of a CACC system. The ego-vehicle utilizes the information transmitted from the two predecessors ahead. A new prediction model which estimates the probability distribution of the future speed of the preceding vehicle is trained with real data. It intends to prevent errors in the prediction model from causing degraded comfort, harsh actions, and more fuel consumption. In addition, the CACC controller has a flexible spacing policy by using a maximum and minimum headway constraint in the optimization model instead of the commonly-used constant headway policy. The results showed that the proposed MPC with the stochastic prediction model outperforms the deterministic MPC approach. To develop a robust ACC controller which can solve the same issue, Sakhdari et al. (2017) also proposed a tube-based MPC approach, which maintains the system state in a certain region. The tube is derived from offline defined uncertainty boundaries computed from an additional linear feedback controller, which is then incorporated into the cost function of the MPC problem. The simulation results demonstrated the success of the proposed approach in handling the sensor delay and system disturbances. According to the contribution of these studies, the potential of MPC in handling the uncertainty in the ACC system is implied.

However, it is computationally expensive to solve an online optimization problem in the MPC framework given the problem complexity and the prediction horizon specified. A lot of research endeavor has been placed on resolving the computation requirement of this control method to favor the real-world application. For instance, Wang, Daamen, Hoogendoorn, et al. (2014b) adopted a faster numerical solution algorithm to solve the optimal control problem for real-time control. S. E. Li et al. (2014) also aimed to deal with the computational complexity problem of MPC for the real-world online application of ACC by relaxing the inequality constraints and reduced the number of control variables in the prediction horizon. The results from the simulation experiments and field tests showed reduced computation time and satisfactory controller performance. There is also some methods proposed to completely remove the computational concern of this optimal control method, which will be discussed in subsection 7.4.1.

2.1.3. Intelligent control

Intelligent control methods lie between classical control theoretical approaches and those approaches utilizing the concept of artificial intelligence, including neural network, fuzzy logic, and the general machine learning methods. It is usually considered for the design of complex dynamical systems whose behaviors are difficult to be described.

There were many studies using various intelligent control methods to design ACC controllers. Naranjo et al. (2003) proposed an ACC controller based on fuzzy logic. The controller and experiments were then extended to cope with extreme Stop&GO maneuvers (Milanes et al., 2012; Naranjo et al., 2006). Tsai et al. (2010) also developed a fuzzy longitudinal controller which can adapt to various car-following conditions. In the simulation experiments done by Ko and Lee (2007), it was even found that a fuzzy ACC controller can guarantee string stability. However, the drawback of fuzzy logic-based control is its lack of systematic design methodology, such as parameter tuning. Neural network gets rid of the problem by automating it through a learning-based approach (Godjevac, 1995).

Neural network (NN) is another intelligent control method which can be used to describe complex system dynamics and implement human intelligence and reasoning into machines. Bifulco et al. (2008) developed a human-like ACC by using an artificial neural network (ANN). It was found that ANN can reproduce the human driving behavior as accurately as a linear model. However, it is mentioned that ANN may suffer from over-fitting problems caused by the learning process. There were also a few studies using NNs to model human drivers' car-following behaviors although they were not aiming for developing an ACC system. For instance, Ohno (2001) used a driving simulator to collect experienced drivers' behavior characteristics when using ACC and built a NN-based driver model. The purpose was to conduct simulations using the developed model and test the performance of vehicles equipped with the ACC system.

On the other hand, reinforcement learning (RL) has become one of the most popular areas in machine learning nowadays. One of the earliest research application of RL on AVs can be found in Forbes (2002). A model-based reinforcement learning method to train a control policy for AVs to conduct car-following and lane-following tasks. Many studies using RL and its extension for autonomous driving have been proposed since then. Two decades later, Z. Zhu and Zhao (2021) provided a comprehensive review of the all the studies developing autonomous driving policies using related methods and discussed them from the system architecture, task-specific, and problem-driven perspectives. Several problems which need to be aware of when applying this kind of method for autonomous driving and the directions for future development were summarized at the end.

For the development which specifically focused on ACC systems using RL, Ng et al. (2008) applied Monte Carlo RL to develop an ACC controller, which updates the feedback gains in a parametric linear controller at every time instance. A certain level of stable platoon performance was also found in the numerical simulation results. Desjardins and Chaib-Draa (2011) used the policy gradient algorithm in RL to develop CACC controllers. A Stop&Go scenario was used as the learning task for the agent. In addition, a reward function was designed to direct the agent to execute the desired car-following behavior. To the best of our knowledge, this was the first study to apply a model-free RL approach to design a vehicle longitudinal speed controller.

To leverage the strength of both learning-based control method and deep learning with NN, imitation learning (IL) and deep reinforcement learning (DRL), which combines the two methods together, gains a lot of attention in these days. Because of the rapid advancement of learning algorithms, many studies have started to apply IL to train CC, ACC or CACC controllers. First, although the term "IL" and "AVs" were not mentioned in the study, the car-following model trained in Chong et al. (2013) is actually a human-like vehicle longitudinal controller developed using the concept of IL in disguise. In order to overcome the drawbacks of traditional car-following model calibration methods, a fuzzy rule-based neural network to simulate individual driving behavior was constructed in that study. The network was trained with real vehicle trajectory data through the actor-critic method. The NN-based approach intended to solve the problem of high dimensional state space in the microscopic traffic behavior modelling and cope with the driver behavior heterogeneity. As can be seen in this study, researchers have started to improve the performance of intelligent control method by incorporating more than one concepts and approaches. M. Zhu et al. (2018) also intended to train a human-like AV longitudinal controller using IL. Empirical driving data were used in the simulation environment for the training. It was found that the built deep deterministic policy gradient (DDPG) controller has a greater accuracy in terms of reproducing human drivers behaviors than other data-driven car-following models do. It also showed good generalization to different car-following scenarios. In M. Zhu et al. (2020), objectives regarding safety, efficiency, and comfort were further considered in the reward design. The proposed approach significantly outperforms MPC-based ACC controller in terms of computation speed.

Instead of reproducing the human-like behavior by IL, DRL approaches for ACC or logitudinal speed controller design seek to maximize a cumulative reward function value in each training episode which is designed with expert knowledge. Zhao et al. (2013) adopted a supervised actor-critic approach to develop an ACC controller. The actor network was pre-trained using a supervised approach to enhance the convergence speed of the training and the control policy performance. The study was also proposed before the concept of DRL has become mature. There were a lot of DRL-based ACC controller developed using DDPG as it is one of the most recently proposed DRL algorithms. To dampen traffic oscillations and reduce energy consumption, Qu et al. (2020) also to develop a car-following model based on DDPG for electric connected and automated vehicles. A platoon of 10 vehicles which drive on a ring road were simulated in the environment for training. Lin et al. (2020) compared the DRL-based and MPC-based ACC controllers. It was found that DRL may still suffer from the generalization problem of machine learning approaches when the state inputs are out of the range for training. However, it performs better than MPC when there are large modelling errors, such as control delay or unexpected system dynamics. This study provided different insights into the benefit of DRL-based ACC controller design rather than only mentioning the increased computation efficiency. Hart et al. (2021) divided the ACC car-following model into two modes and hence trained two policies, the free-driving and car-following policies. They were trained on leader speed profiles generated from a stochastic process to reflect vehicle dynamics in reality. It was found in the testing of a five-follower platoon that the models can achieve string stability without explicitly considering it in the policy training setup.

For vehicle longitudinal control utilizing in-vehicle information of predecessors, M. Li et al. (2020) proposed a DRL-based driving strategy to reduce the collision risks in traffic oscillations. The acceleration of the preceding vehicle is transmitted via vehicle communication technologies. A surrogate measurement of safety was used in the reward function. The strategy was implemented into microscopic traffic simulator to analyze the traffic flow and safety performance. Mirwald et al. (2021) developed a learning-based CACC controller which taks the string stability and communication loss into account. Not only the current acceleration, the preceding vehicle transmits its predicted future acceleration information to the following vehicle. The string stability was considered by including a constraint in the reward function which limits the acceleration of the ego-vehicle according to the acceleration of the preceding vehicle in a certain time window in the past. On the other hand, the communication loss was modelled by two levels of communication quality with a Markov chain. The results showed that the trained agent can ensure string stability to a certain extent and reduce the impact of communication loss. To handle communication failure, Shi et al. (2022) also incorporated the time- and spatial-varying information flow topology of V2V communication into the training environment to develop a distributed longitudinal control strategy for connected automated vehicles to dampen the traffic oscillation. Different from other decentralized CACC controllers which only use the information of the direct preceding vehicle, in-vehicle information of multiple downstream preceding vehicles can be transmitted via dedicated short range communication in the control scheme. Numerical simulation results showed that the proposed distributed control outperforms other strategies in terms of stabilizing the oscillation. At the next level of vehicle platoon control, there were also studies using the concept of multi-agent DRL to develop vehicle longitudinal speed controllers which can achieve cooperative platoon behavior with the help of vehicle communication technologies. To achieve multiple objectives for a CAV platoon in a mixed traffic environment, Shi et al. (2021) proposed a cooperative CAV control strategy based on a DRL algorithm. The traffic stream is decomposed into multiple vehicle subplatoons led by a human-driven vehicle to reduce the computation and communication burden. Each subplatoon has a control module which determines the immediate acceleration of every CAV in the subplatoon. To consider the string stability in the reward function, a control theoretical string stability criteria in the frequency domain is modified into the time domain. By locally optimizing each subplatoon, the global traffic performance can be improved. The effectiveness of the proposed strategy was demonstrated by simulation experiments using NGSIM data.

Nowadays, one of the most popular issue for intelligent control to tackle is the so-called decision-making and planning under uncertainty. As mentioned previously, Mirwald et al. (2021) and Shi et al. (2022) applied DRL to tackle the state information uncertainty stemming from the communication loss. In addition to DRL based on the typical Markov decision process (MDP), it is worth developing RL or DRL-based vehicle controller using the concept of partially observable Markov decision process (POMDP) so that the presence of noise or other uncertainty in the system state can be considered, as discussed by Forbes (2002) at the end of his study.

There were already several studies developing vehicle longitudinal controller or control strategy based on the concept of POMDP. J. Wei et al. (2011) described the single-lane AV driving behavior as a point-based Markov decision process to consider three types of uncertainties, sensor noise, perception constraints, and future behavior of the surrounding vehicles. A statistical acceleration prediction model based on the assumption that the acceleration would decay to zero was used to predict the leader behavior. Simulation and road test results both indicated the proposed model-based RL approach produce more robust and safe driving behavior than the state-of-the-art ACC systems when considering uncertainties resulted from the sensor capability. Albeaik et al. (2022) developed a cruise controller for the vehicle mechanical system of a heavy duty truck using DRL. The problem was defined as a POMDP to consider the unknown mechanical configuration and internal state of the truck. The trained control policy showed success in inferring the state variables and tracking. More studies regarding DRL based on POMDP and their applications for AVs will be introduced in subsection 2.5.2.

H. Zhou et al. (2022) provided an in-depth review for the longitudinal motion planning of AVs through learning-based methods, e.g., IL and (D)RL. The state-of-the-art achievement in the industry and endeavor which have been made by the research community were both discussed. It pointed out that automakers tend to focus on the safety performance of the longitudinal motion planning systems. Hence, particular attention was placed on the contribution of these methods for congestion mitigation in the article. It was claimed by the autor that (D)RL-based methods possess a greater potential for integrating traffic-related domain knowledge into the motion planning design. The importance of the ability to predict the future motion of the surrounding vehicles using recurrent NNs was also emphasized. The suggestions made in this article are highly consistent with the above-mentioned review results in this subsection and will become an important foundation for the methodology adopted in this study.

2.2. String stability of vehicular traffic

String stability, which is also known as asymptotic stability or platoon stability, has long been a popular research topic in the domain of traffic flow and autonomous vehicle longitudinal control. String stability in car-following dynamics discusses the propagation of disturbance from vehicles to vehicles within the platoon. The platoon is said to be string unstable if the disturbance amplitude becomes larger while propagating to the next vehicles in the upstream direction. On the contrary, if the disturbance gradually damps out while propagating, the platoon achieves a strict string stability (Pueboobpaphan and van Arem, 2010). The methodologies adopted to analyze string stability and design ACC controllers are discussed in this section.

For human driven vehicular traffic, the two main reasons of string instability are the time lag of acceleration command for adapting the speed caused by the vehicle longitudinal control system limitation and the reaction time of the driver. For traffic consisting of autonomous vehicles, the sensor or communication delay, which is similar to the reaction time of human drivers, becomes the main source of instability. In Treiber et al. (2007) and Kesting and Treiber (2008), the influence of the reaction time and adaptation time on the stability of traffic flow was explored. These factors were mathematically incorporated into the IDM to microscopically simulate the affected traffic dynamics. It was found that these two factors causes string instability in different wave-lengths of instability. Apart from the time delays, other parameters, such as the desired time gap in the numerical car-following model, or the net distance between vehicles also influence the string stability of a platoon. However, the effect of these factors and parameters is not the major focus in this study.

The string stability of a vehicle platoon can be investigated using different approaches. Pueboobpaphan and van Arem (2010) provided an detailed review of stability analysis methods for vehicular traffic. At the beginning, the theoretical linear stability analysis was the most commonly used method. In these studies, the transfer function of the error of spacing, speed, or acceleration between the following vehicle and its leader is then formulated based on the car-following model considered. If the transfer function value is less than one, the platoon is believed to be string stable. An example of the transfer function of the speed error can be written as

$$\Gamma(z) = \frac{\tilde{V}_i(z)}{\tilde{V}_{i-1}(z)}, \tag{2.7}$$

where $\tilde{V}_i(z)$ and $\tilde{V}_{i-1}(z)$ are Laplace transforms of the speed disturbance of the follower and the leader, respectively.

The first study using this approach can be dated back to 1950s. Chandler et al. (1958) and Herman et al. (1959) considered the propagation of speed and acceleration fluctuation in a human driven vehicle platoon using a relative speed control model caused by the first leading vehicle. A criterion regarding the sensitivity to relative speed and reaction time was deducted. This method was later on used by Shladover (1978) to analyze the vehicle-follower control system for the automated guideway transit vehicles. Swaroop et al. (1994) also introduced this method to compare the string stability of two different spacing policies for automatically controlled vehicles. Wilson and Ward (2011) provided an overview of the framework of the microscopic car-following models and corresponding linear stability analysis methods. However, this kind of method relies on several assumptions, including a homogeneous vehicle platoon and a steady and periodic oscillation. These assumptions limits the applicability of such method in stability analysis (Sun et al., 2020). To cope with this difficulty, Wang et al. (2017) proposed a third-order linearized vehicle dynamics model which can be used for a platoon with mixed vehicle classes and controller parameter settings and extended the transfer function formulation of string stability criteria for heterogeneous platoons. The predictive power of this method was demonstrated by comparing with a systematic simulation approach. Montanino and Punzo (2021) also proposed a method to analyze the string stability with drivers' and vehicles heterogeneity. A condition for weak string stability was defined. The study also emphasized that the uncertain transfer function should be used to map the probability distributions of controller parameters to the defined condition. It was concluded that the string stability of a heterogeneous platoon is heavily governed by the platoon heterogeneity. The methods regarding string stability analysis for heterogeneous platoon were also discussed in Xiao and Gao (2011) and Shang and Stern (2021).

As adopted by Treiber et al. (2007) and Kesting and Treiber (2008), there are also many other studies using simulation approaches and looking at certain indicators, such as the amplitude of an instantaneous acceleration and the acceleration profile, to analyze the string stability effect. Simulation approaches overcome the limitation resulted from the assumptions made when using the theoretical analysis methods and can be flexibly adopted to evaluate the string stability in different types of traffic scenarios which is closer to reality. Studies which can be classified into this category include VanderWerf et al. (2001), Bareket et al. (2003), and Xiao et al. (2017).

The other approach to investigate the string stability is by conducting field experiments. Knoop et al. (2019) conducted an experiment asking seven SAE level-2 autonomous vehicles to drive as a platoon on public roads. The driving behavior and interactions between these vehicles were discussed. To analyze the platoon stability, the study chose a rather simplified and qualitative way by investigating the fluctuations in vehicle speed. The delayed, amplified, and strong response to the speed changes of leader could clearly be observed in the speed profiles, which showed that the tested platoon was unstable. Makridis et al. (2020) conducted an experiment with five ACC-equipped vehicles at the AstaZero test track. Recorded data with different perturbation magnitudes and equilibrium speeds also showed the instability of the platoon is visible even in small perturbation caused by the varying road gradient. Ciuffo et al. (2021) also analyzed the results of a test campaign held at ZalaZONE Automotive Proving Ground. A 10-vehicle platoon consists of human-driven vehicles and ACC-equipped vehicles from different automakers was tested. Both the weak and strict string stability was evaluated using the concept of $\mathscr{L}_p$ string stability. The results showed that string instability conditions occurred in every time gap setting. It was also found that the road geometry can induce and amplify the shockwaves. T. Li et al. (2021) setup an experiment using a three vehicle platoon. The lead vehicle produced several different driving profiles which reflect the typical traffic oscillation. The string stability was then analyzed by examining the response time, speed disturbance indicators, including oscillation growth and overshooting, and acceleration disturbance indicators, including deceleration and acceleration rates. The influence and correlation between these factors were also explored to characterize the ACC car-following behavior.

Many studies actually combined theoretical and empirical methods with a simulation approach to analyze the string stability of vehicular traffic. Milanés and Shladover (2014) tested three proposed ACC/ CACC models by implementing them on four Nissan Infiniti M56s. The experimental results of these controllers were also used to derive ACC/CACC car-following models for further microscopic simulation analysis. It was found from both the experimental and simulation results that a platoon of ACC vehicles is string unstable, while CACC systems can provide significant improvement. Gunter et al. (2019) conducted a series of car-following experiments to calibrate an optimal velocity relative velocity model. It was found that the model with the best goodness of fit are string unstable. Gunter et al. (2021) further extended the experiment by testing seven ACC-equipped vehicle models. The data were collected from a 1200-mile driving experiment. The theoretical analysis results again showed that none of them can guarantee string stability. Later on, these experimental data were used by Shang and Stern (2021) to calibrate the parameters in IDM and apply it for theoretical and simulation analysis of the string stability and traffic flow stability of traffic streams with ACC vehicles.

Studies mentioned above provided many methods to analyze string stability. There were many studies which aimed to consider string stability in the design of ACC controllers. Liang and Peng (1999) optimized the parameters in the linear ACC controller to guarantee string stable performance. The dynamics of the lower-level controller was considered. On the other hand, J. Zhou and Peng (2005) chose to ensure string stability and traffic flow stability by optimizing the spacing policy of the ACC system dynamically. The two stability criteria were formulated as constraints in the optimization problem. The simulation results showed that the traffic density can be significantly increased. To cope with the influence of sensor and actuator delay on string stability, Xiao and Gao (2011) also developed a sliding-mode ACC controller to produce string stable performance for both a homogeneous and heterogeneous platoon. Wang et al. (2018) adopted an MPC controller to compensate the influence of sensor delay by estimating the true state according to the previous state, historical information, and system dynamics, while the actuator lag was compensated by considering it in the prediction model to formulate an anticipatory control strategy. Simulation results showed that the MPC ACC controller is robust in scenarios with small sensor delays, while the impact of actuator lag is not as large as that of sensor delay.

2.3. Multi-anticipative car following models

This study is built on the belief that the string stability of vehicular traffic cam be improved by enabling drivers or vehicles to look at multiple vehicles ahead. This idea has been demonstrated in the literature. This section summarized these past studies using theoretical and empirical analysis to show the effect and existence of this behavior.

The concept of multi-anticipative car-following behavior in the real-world driving scenario has already been pointed out in Herman et al. (1959). In that study, a control scheme for the longitudinal behavior of a human driven vehicle involving two vehicles ahead was proposed by modifying the original model which only includes the influence of the first vehicle in front. The theoretical criteria for asymptotic stability of this control scheme was also formulated. Treiber et al. (2006) tried to include spatial anticipation for multiple vehicle ahead into a time-continuous car-following model. The acceleration is calculated by summing influence of every vehicle ahead considered based on the distance, speed, and speed difference. It was shown that the anticipation can compensate the reaction time and estimation error of human drivers, which could lead to destabilizing effects. The basic models proposed in the study can also be regarded as ACC controllers to understand the impact of autonomous vehicles on overall traffic. According to the understanding of the compensation mechanism discovered, Treiber et al. (2007) again highlighted the effect of anticipation by specifically looking at the stability performance. It was found that a proper anticipation behavior allows the reaction time to be even longer than the safety time gap.

Different from the previous studies which proposed modified car-following models based on theoretical knowledge, Hoogendoorn et al. (2006) carried out an investigation of multi-anticipative behaviors using empirical trajectory data. Several relatively simple car-following models were selected as the basis for the derivation of their multi-leader forms and calibrated using maximum likelihood estimation. It was concluded that considering the multi-anticipative stimuli improves the extent of which the model can represent the manual driving behavior. In addition, vehicles tend to be more sensitive to the relative speed with the second and third leaders than to the distance in between. More discussions on the influence of the multi-anticipative behavior were discussed in Ossen (2008). Heterogeneity exists in multi-anticipative behavior of different drivers. It was found that more than 50% of the drivers look further ahead than the direct leader, and even 20% considered more than two vehicles ahead. Furthermore, to which extent the multi-anticipative behavior influence the driving behavior depends on the action of both the direct leader and second leader ahead. The author also pointed out that the insights into this kind of behavior can help automakers produce a more appropriate design of ACC controllers for customers. To enhance the prediction accuracy of the multi-anticipative car-following behavior, Lu et al. (2015) used support vector regression method to train a model using NGSIM vehicle trajectory data. The analysis results also confirmed the existence of multi-anticipation and presented insights into the behavior.

The aforementioned multi-anticipative models were used to describe the car-following behavior of human driven vehicles. The purpose of them is to capture the real reaction of drivers on the road. For autonomous vehicles, these concepts are applied to improve traffic stability. Hasebe et al. (2003) extended the OVM to achieve cooperative longitudinal driving control considering the gaps between both preceding (forward-looking) and following (backward-looking) vehicles. Linear stability analysis method was applied to different set of parameters. It was found that by looking at two vehicles ahead and one following vehicle generates the best stability performance. Hallouzi et al. (2004) tested a cooperative driving setup which utilized inter-vehicle communication. The ego-vehicle looks at two preceding vehicles and maintains certain time gaps with them. The minimum acceleration is selected to ensure safety of the vehicle. In the CACC system developed by de Bruin et al. (2004), the distance and speed information of the vehicle further ahead serves as a correction term for the direct leader input information, which selects the vehicle with largest potential danger and modify the input information for the controller. To cope with the congestion forming at sags, Papacharalampous et al. (2015) also utilized the concept of multi-anticipation to design a CACC controller by adding a term in the control law to consider the influence of multiple vehicle downstream.

Accomplishing multi-anticipative car-following behavior through cooperative or communication systems allows the controlled vehicle to handle the situation that there are human driven predecessors between itself and the closest preceding vehicle which has the inter-vehicle communication functionality. Wilmink et al. (2007) proposed a CACC control mode which uses a mean speed difference with the predecessors considering the difficulty to obtain distance measurements when the system penetration rate is low. Lee et al. (2021) also proposed a CACC system which can still leverage the benefit of using communication technologies even when the direct predecessor is an unconnected vehicle.

Most of the studies using CACC systems with multi-anticipation mentioned above extended some developed car-following models or typical state feedback ACC controllers. Kreuzen (2012), Wang, Daamen, Hoogendoorn, et al. (2014a), and Dollar et al. (2021) developed MPC-based CACC systems with multi-anticipation. This kind of receding horizon control method allows the ego-vehicle to further predict the motion of multiple predecessors, which enhances the potential benefit of multi-anticipative behavior. By doing so, a high level of cooperative control strategy for the ego-vehicle and its followers can be achieved. The systems can also predict the behavior of the unconnected vehicles between the ego-vehicle and the first connected predecessor. In Wang, Daamen, Hoogendoorn, et al. (2014a), even a high level cooperative behavior which allows the CACC-equipped vehicle to control the behavior of the human-driven following vehicle was proposed. These studies demonstrated the potential of MPC-based CACC controller to achieve multi-anticipation considering platoon heterogeneity and low penetration rate of connected vehicles.

To our knowledge, there was little effort in both the industry and the research community studying or introducing the multi-anticipation functionality of ACC systems without using communication technologies. The first and only to develop a vehicle which has the ability to detect more than one leader was Tesla Autopilot v8.0 with its advanced RADAR system. The sensor setup which can achieve this kind of functionality will be discussed in section 2.4. However, the purpose of it is to provide early emergency braking for the ego-vehicle when one of the leaders ahead apply a hard braking. To what extent this advanced RADAR system contribute to the ACC system is unknown for the public. Following this functionality, Donà et al. (2022) recently published their work on developing and evaluating the string stability of an ACC system with multi-anticipation ability through RADAR sensing. The system uses the most commonly used linear state feedback controller with modified inputs which integrate the second leader measurements and considers their relative importance. The typical theoretical linear stability analysis approach was adopted to assess the string stability of the linear system and provide a reference for parameters tuning. The problem definition and consideration are quite similar with this study, as mentioned in chapter 1. However, the desired time gap setting was slightly larger than the discovered performance limit of the linear ACC system. In addition, the real-world sensor capability was not considered in the study. Whether the proposed system design can be applied in reality and the improvement it can bring to the overall traffic flow performance require more investigation.

2.4. Automotive sensor technology development

On-board sensors which are currently implemented on AVs include Inertia Measurement Units (IMU), Global Positioning System (GPS), Radio Detection and Ranging (RADAR), Light Detection and Ranging (LiDAR), cameras, and ultrasonic sensors, etc. The exteroceptive sensors which are used in production ACC systems to collect measurements from other vehicles include RADAR, LiDAR, and cameras (Rosique et al., 2019; Van Brummelen et al., 2018; Vargas et al., 2021).

It is important to take several practical factors, such as required information and reliability, into account when choosing or designing a suitable sensor setup for a specific AV task. As mentioned in section 1.4, the proposed ACC systems in this study rely on a multi-leader detection functionality which enables the ego-vehicle to perceive the second leader ahead of it. To detect the second leader which may be completely occluded by the first leader, the on-board sensor setup should be able to conduct non-line-of-sight detection task. Active sensors which uses light beam, such as LiDAR, and vision-based camera sensors are therefore not suitable for this functionality since it is impossible for the beam to penetrate through objects which are mostly occluded. On the contrary, RADAR sensors possess the capability of detecting completely occluded objects with the diffraction and reflection characteristics of radio waves.

Typical RADAR sensors used by automotive for ACC systems are Millimeter-wave (MMW) RADAR (Hasch et al., 2012; T. Zhou et al., 2020). By using the time of flight (ToF) between the emitted waves and the echo, a point cloud is generated at every time instance. The ToF can also be used to estimate the distance (range) between the ego-vehicle and the target vehicle. The speed (range rate) measurements are estimated by the observed Doppler frequency shift.

The idea and effectiveness of using RADAR to detect completely occluded or non-line-of-sight (NLOS) objects for automotive have already been realized by a commercial vehicle model. Tesla Autopilot v8.0 is equipped with an advanced RADAR system which has a more detailed point cloud and can detect more than one vehicles ahead, as mentioned in the section 2.3. It was stated that the RADAR signal can bounce under the direct predecessor to reach the vehicle further ahead. For studies related to NLOS detection, Scheiner et al. (2020) pointed out that detecting the NLOS objects requires recovering them from reflected signals, which is mostly fainted and considered to be the noise. A NLOS detection approach using Doppler RADAR sensor was developed by jointly detecting and tracking the NLOS objects. To detect rush-out pedestrians which may be in the blind area or occluded, Hayashi et al. (2021) also developed a detection and motion classification approach using micro-doppler RADAR. It was found that the radio wave can propagate by being diffracted or reflected from the bottom space of the occluding vehicle. The conducted experiments showed that the Doppler RADAR and simple clustering approach can achieve a high accuracy of behavior detection and classification. These studies both revealed the possibility of NLOS or occluded detection using RADAR sensor.

In addition to its ability to detect occluded object, the merits of using RADAR include long detection range, direct speed measurements, and low cost. RADAR can also conduct reliable detection under all weather since the radio wave has higher penetrability which makes them applicable in adverse visibility conditions (Vargas et al., 2021). These features enable it to outperform other exteroceptive sensors in driving environments with such sensing requirements. Although the measurement accuracy of RADAR may be slightly lower than LiDAR, it is still the most suited sensor for this specific task. Therefore, a MMW long range RADAR is selected for the multi-leader detection functionality in the proposed ACC systems.

Many studies nowadays have highlighted the importance of RADAR and seek to facilitate its development and applications in the automotive industry (Abdu et al., 2021). Despite all the advantages mentioned above, there is still a concern of RADAR sensor regarding the possibility and frequency of false positive signals when receiving reflections from other objects which are not of interest for the ADAS applications caused by the bouncing of the signals. The multiple radio wave reflections from these objects resulted in false alarms (Rosique et al., 2019; Van Brummelen et al., 2018). Moreover, RADAR-based sensor setup may only achieve an object-level perception task, while AVs nowadays usually requires a high-level perception to understand the extent of the object. Berthold et al. (2017) investigated the characteristics of RADAR point cloud data and the potential of using them for vehicle contour estimation. Palffy et al. (2022) also seek to utilize the elevation measurement from a 3+1D RADAR for multi-class road user detection. The results showed that the additional elevation information and subsequent RADAR data bring its performance closer to the level of using LiDAR data.

However, to really counter the issue of false positive and increase the accuracy of RADAR detection for the second leader measurements, the multi-leader ACC systems are expected to rely on a high-level perception sensor setup, which is not required by original ACC systems. Using a proper sensor fusion technique to combine the information from those vision-based sensors is a suitable method to achieve the multi-leader detection functionality (Liu et al., 2021; S. Wu et al., 2009). By integrating multiple calibrated sensors, the ego-vehicle has a better knowledge of whether the information collected from RADAR is indeed useful for the multi-leader ACC system. Since the focus of this study is a single-lane car-following problem, it is assumed that the on-board RADAR coupled with other sensors implemented on the ego-vehicle possess the ability to identify whether a detected object is on the same lane with itself by using the sensor fusion approach mentioned above. Therefore, the false positive detection signals are not considered in the defined problem context, as introduced in section 1.4.

2.5. Measurement uncertainties

This section discussed the method for modelling and handling of measurement uncertainties done by past studies related to autonomous driving or ACC/CACC systems in particular.

2.5.1. Modelling of measurement uncertainties

Sensor measurement uncertainties critically influence both the development of ADAS applications and validation of their performance. To obtain a reliable results, the sensor measurement must be simulated in a proper way which can reflect a realistic driving environment. Past studies applied many methods to simulate and model the sensor. For studies using data to statistically extract or learn the pattern of the error, Hanke et al. (2016) introduced a classification hierarchy according to the type of sensor measurements. It helped derive statistical models of the measurement error by using reference data of the corresponding sensor. To model the sequential measurements generated from LiDAR, Zec et al. (2018) proposed a hidden Markov model to represent the stochastic process of the measurement errors by training through large amount of data. Mitra et al. (2018) also integrated the error of camera bounding box detection by applying a non-linear auto-regressive method. The method helped the objection detection module to be aware of the erroneous information and improved the AV decision-making.

Apart from those data-driven modelling methods, there were also several studies following certain assumptions of the stochasticity of the measurement uncertainties, including noise, false negative, and false positive. In Saxena et al. (2019), it was simply assumed that measurement noise follows Gaussian distributions. When the false negative signal occurs, the noise variance becomes larger than in the normal condition to reflect the influence of the missing data point. To take the occurrence of false positive and false negative into account, Bock et al. (2018) tested a multi-sensor driver assistant system by simulating the probability of false negative and false positive signals at every time instance using discrete time Markov chains. Three cases with different levels of error complexity in terms of the dependency between successive sensor detection and correlation between different sensors are considered. Piazzoni et al. (2021) also simulated the false negative signals as a Markov chain with a specified steady state probability and a mean sojourn time in the detection and non-detection state. At a more detailed level, Elmquist and Negrut (2020) gave an overview of how the sensor models process the signals to generate measurements and how the sensor errors are incorporated. By comparing these sensor models and the experiment conducted, the required accuracy of sensor models used for the simulation of AVs were concluded.

Instead of previous studies which mainly discussed the influence of sensing and perception errors on the safety performance of general AVs, there were several studies specifically focusing on the influence of measurement uncertainties on developing and evaluating ACC/CACC systems for both safety and traffic stability. To close the gap between simulation and reality and explore the effect of measurement errors, Wang, Daamen, Hoogendoorn, et al. (2014), Y. Zhou et al. (2017), and Donà et al. (2022) all included Gaussian white noise in the range and range rate measurements collected from the considered on-board sensor. On the other hand, the uncertainties generated by the communication loss were more frequently discussed for CACC systems (Hallouzi et al., 2004; C. Wu et al., 2019).

2.5.2. Handling of measurement uncertainties

To reduce the effect of measurement uncertainty on decision-making of the vehicle, a suitable tracking algorithm can be used. Kalman filter (KF) is one of the most commonly employed methods for vehicle tracking in the past. Nowadays, many advanced filtering methods, such as adaptive KF or particle filter, have been proposed to overcome the limitation of KF. Floudas et al. (2005) tested the performance of KF and particle filter in terms of position and velocity estimation using RADAR measurements. It was found that KF-based techniques performs better than particle filter due to the comparatively short processing time requirement considering the developed computation power at that time. On the other hand, to achieve accurate tracking og the position of the preceding vehicle, Aldrich and Wickramarathne (2018) proposed a KF method for a LiDAR-based tracker since it is known that RADAR sensor measurements are known to be more noisy than other types of sensors. Kim and Park (2020) also proposed an extended KF (EKF) by fusing the RADAR and LiDAR measurements. A reliability function considering the distance as a factor of the sensing performance was used to determine the Kalman gain at every time step. The real-world experiment showed improved tracking accuracy after using the reliability function in the EKF framework.

There were several studies specifically using KF as the state estimator in the control loop of ACC/CACC systems. Y. Zhou et al. (2017) incorporated the KF into the developed MPC-based ACC and CACC controllers. The optimization problem was decomposed into a linearly constrained linear quadratic problem and a linear quadratic estimation problem. An adaptive KF was used to estimate the covariance of process dynamics. For the handling of temporary communication loss of a CACC system, Hallouzi et al. (2004) used an extended KF to fuse the position, velocity, and acceleration information of the preceding vehicle collected from the on-board GPS sensor via inter-vehicle communication. The state estimator were used to interpolate the position of the vehicles when there is no input information from the GPS. C. Wu et al. (2019) also adopted an adaptive KF to estimate the acceleration of the preceding vehicle and prevent the CACC system from degrading to an ACC system due to communication loss. The covariance of the process dynamics was calculated by a probability density function of the target acceleration value. Both the simulation and experiments showed a reduced gap error compared to the system using a normal KF.

Other than using traditional filtering approach to track the motion of the object, much research has started to use recurrent neural networks (RNNs) to predict and estimate the target motion. Chenna et al. (2004) compared the performance of KF and RNNs for state estimation and tracking problems. An RNN architecture with one hidden layer was built. The simulation results indicated that RNNs can produce state estimation and tracking performance which closely match that of a basic KF. However, the potential of RNNs for cases violating the linear and Gaussian assumptions was pointed out. These network models can be trained with time series data to learn the probability pattern of the next output state. Later on, a lot of studies regarding using Long Short Term Memory (LSTM) networks, a type of RNNs, to predict or track the motion trajectory of other surrounding vehicles have been proposed (Akita and Mita, 2019; Deo and Trivedi, 2018). Chandramouli (2021) developed several LSTM models to predict the longitudinal and lateral trajectory of the target vehicle. The highD dataset was used in both the model training and evaluation. The models with different attention mechanisms were trained with both filtered and unfiltered data. It was found that training with filtered data performs better than that of unfiltered data in general. the Lane changing behaviors can also be predicted. Furthermore, for the prediction of vehicle behaviors in traffic oscillation which is believed to be more difficult to predict than normal car-following cases, M. Zhou et al. (2017) proposed an RNN-based car-following model. The RNN-based model outperformed the classical IDM when considering the driver characteristic heterogeneity. There were also studies aiming to integrate the merit of RNNs into the design of KFs (Revach et al., 2022; Zheng et al., 2019). However, this kind of idea has not been applied particularly for trajectory prediction or tracking tasks in ADAS.

Previous studies seek to resolve the uncertainties by applying various state estimation models and approaches. Nowadays, much research has been exploring the possibility of handling the uncertainties at the decision-making level of AVs. RL based on POMDPs is a method which suits this kind of problem properly. The concept of using POMDPs for the decision-making and planning of AVs was also mentioned in Forbes (2002). In model-based RL based on POMDPs, a belief state, which is the probability distribution of the underlying state, is estimated in a feedforward structure (McAllister and Rasmussen, 2017). On the other hand, model-free DRL solves POMDPs by using RNNs to integrate the historical states and actions. Bakker (2001) is one of the earliest to propose the idea of using LSTM for model-free RL based on POMDPs. The purpose and advantage of using LSTM is its ability to utilize the memory to solve a problem which requires long-term dependency. There were several studies which integrated RNNs or LSTM networks into DRL networks to solve the POMDP problems. Hausknecht and Stone (2015) first added recurrency into a deep Q network to create a deep recurrent Q network (DRQN). The first convolution layer in the agent was replaced by a LSTM layer to deal with partial observability by utilizing the historical information. Several studies also tried to extend other DRL algorithms for control task with continuous action space by implementing LSTM networks into network structure, such as the recurrent deterministic policy gradient (RDPG) proposed by Heess et al. (2015) and the LSTM-twin-delayed deep deterministic policy gradient (LSTM-TD3) proposed by Meng et al. (2021).

Real-world traffic- and autonomous driving-related problems often contain a lot of information uncertainties in the environment. Therefore, the applications of DRL based on POMDPs in these fields of research can already be found in several relevant studies. For instance, Qiao et al. (2018) proposed a POMDP policy network with LSTM to tackle the problem of an AV traversing a non-controlled urban intersection. Mani et al. (2019) also applied RDPG algorithm to train an AV agent how to utilize the temporal information to navigate in a dense traffic environment by adapting to the slow speed or overtaking other vehicles. As discussed in subsection 2.1.3, there were several studies using DRL to design vehicle cruise controllers considering uncertainty. However, these studies did not explicitly use RNNs to estimate the belief state, indicating the RL problems were not formulated as POMDPs. In Mirwald et al. (2021) for instance, the actions taken in the past two time steps were included in the state vector to represent the history. By doing so, the agent has better access to the information in the environment. Albeaik et al. (2022) applied a recurrent deep learning model to deal with the unknown vehicle configuration and state information.

However, the detailed network architecture was not stated in the study.

Solving a model-based or model-free RL based on POMDPs requires the agent to plan or learn with an unknown transition dynamics of the hidden state. To overcome this additional difficulty, several studies aimed to integrate the concept of a filtering method, such as KF or particle filter, into POMDPs (Erez and Smart, 2010; Ma et al., 2020). Although these methods have not yet been applied to the filed of autonomous driving, they also symbolized the research endeavor to solve a decision-making problem under uncertainty using DRL.

2.6. Summary

This section summarizes all the recent research development regarding the research problem to provide proper reasons and considerations for several decisions which will be made for the methodology in this study.

There were many studies designing ACC controllers using different control methods for various objectives. Stability performance of the system remains to be one of the major challenges for ACC systems. The multi-anticipative car-following behavior provides a possibility to improve string stability as it has already been applied in several proposed CACC system designs. Based on the automotive RADAR sensors development nowadays, it is believed that advanced RADAR sensors are able to detect multiple leaders ahead even if they are completely occluded. A commercial vehicle in the market and a relevant study published recently (Donà et al., 2022) have demonstrated the feasibility of the idea.

This study aims to propose a new control strategy which has the property of multi-anticipation for ACC systems. The systems also need to be robust to measurement uncertainties since the second leader measurements collected from the on-board RADAR sensor may be highly erroneous. To overcome the limitation of a typical linear controller in describing the non-linearity of the complex dynamical behavior and the difficulties in handling sensor measurement uncertainties, MPC or learning-based control approaches possess better capability than typical linear state feedback controllers. According to studies comparing these two approaches, it is believed that learning-based control methods, such as DRL, may perform better than MPC in terms of computation time requirement and uncertainty handling. In addition, a few ACC-related studies have extended the controller design methods from DRL based on MDPs with full observability to DRL based on POMDPs to deal with the state estimation under uncertainty and time-series prediction. Therefore, DRL-based ACC controllers which can handle measurement uncertainties using RNNs will be developed in this study.

To analyze string stability of the developed intelligent control system considering measurement uncertainties, the control theoretical analysis approach may not be suitable for the problem. Instead, a numerical simulation-based approach is then adopted so that its flexibility for different kinds of disturbance scenarios can be utilized.

Regarding measurement uncertainties for ACC systems, past studies mostly considered the presence of noise in sensor measurements and actuator behaviors. Following this consideration, the noisy measurement will first be included in both the design and evaluation of the proposed systems. The author still wants to point out that losing detection (false negative signal) of the second leader may exist due to the undiscovered RADAR sensor capability. This is similar to the problem of communication loss, one of the major issues encountered by CACC systems. Hence, it is also worth investigating the problem of false negative for the proposed multi-leader ACC systems. However, for simplicity, this study does not touch upon these types of measurement uncertainties.

According to the findings from literature review, the methodology for both the system design and performance evaluation in this study will then follow.

3

Control System Design

The first section in this chapter introduces the basic architecture of the system, which is a simplified ACC control system, and provide an overview of the different ACC systems which will be designed following the proposed architecture. The remainder of this chapter describes how to design the ACC controllers and the state estimators, which are the two main components in the system architecture in this study. As mentioned in section 2.6, this study applies DRL to develop the ACC controllers. The general concept and elements in DRL are then elaborated in section 3.2. In the third section, the structure of the policy networks are introduced. The fourth section then explains how the ACC car-following problem is formulated for the training of controller agents with the assumption of accurate measurements. Section 3.5 describes the concept and standard procedure of Kalman filtering, which is one of the state estimation methods adopted by the proposed systems. Section 3.6 explains how to modify the content in section 3.4 to train recurrent policies for ACC controllers which can directly utilize the erroneous measurements and handle the uncertainties. The last section then introduces the simulation approach and tools which will be used to train the controller agents.

3.1. System architecture

Figure 3.1 presents the architecture of the ACC systems proposed in this study. Typical ACC systems can only utilize the distance gap and relative speed measurements with the direct preceding vehicle. However, the multi-leader ACC systems proposed in this study have two upper-level controllers. According to the specified spacing policy, controller 1 uses the position and speed information of the first leader and generate the acceleration command for the ego-vehicle to follow the first leader, while controller 2 is designed for the ego-vehicle to follow the second leader (the pre-preceding vehicle). The minimal acceleration command generated by these two controllers would be the real vehicle acceleration control input in the next time step, as did by the system proposed by Hallouzi et al. (2004).

The "Plant" block in Figure 3.1 contains the vehicle motion model applied to simulate the car-following dynamics in the driving environment according to the determined acceleration command, which will be described in detail in section 3.7. As mentioned in section 1.4, other sub-systems, functionalities, and external factors, e.g., the lower-level ACC controller, internal driveline, and road slope, which can affect the vehicle longitudinal motion are not considered in the defined problem context. A disturbance can occur in the driving environment and deviate the system from the equilibrium state, which is the desired spacing in the car-following problem. In this study, the disturbance would be a speed fluctuation of the leading vehicles. When facing disturbance in the environment, the controllers would then seek to stabilize the system and guide the ego-vehicle back to the desired spacing.

The on-board sensor is another important component in the whole system. It collects and processes exteroceptive measurements as input for the controllers. As mentioned in section 2.4 in this study, RADAR sensor is selected for this ACC system due to its suitability for the non-line-of-sight multi-leader detection functionality. Besides, the state estimators will also be designed to cope with erroneous measurements for the controllers.

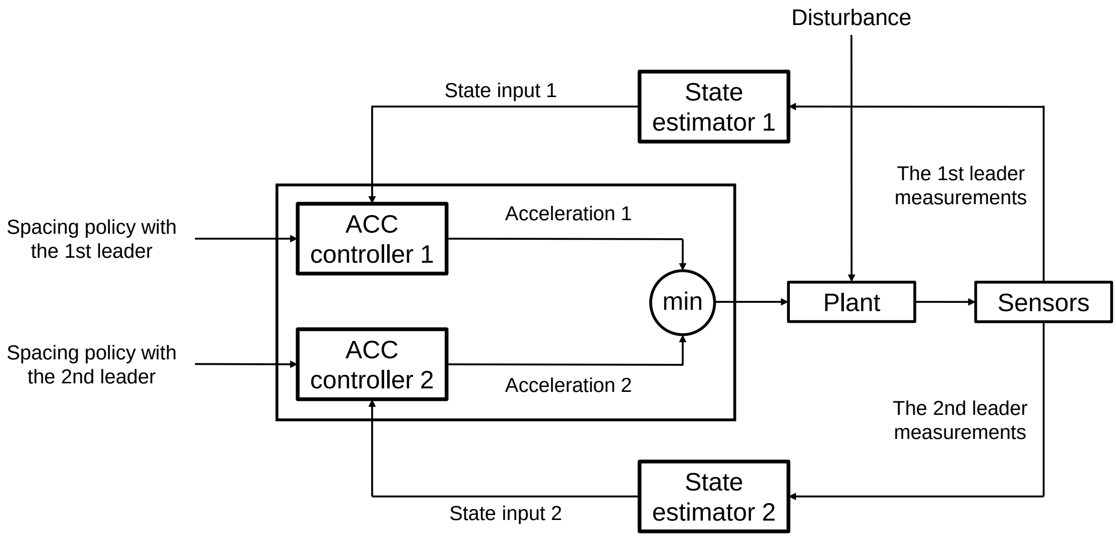


Figure 3.1: The architecture (block diagram) of the multi-leader ACC system

When designing ACC controllers, it is often assumed that the sensor and state estimator implemented in the system together possess the ability to provide accurate input information. This kind of consideration may be acceptable for the first leader measurements given the state-of-the-art sensor technology development. However, the second leader measurements may be relatively erroneous given that the ability of sensors to accurately detect the second leader ahead is still unknown especially when it is occluded to a large extent. Enabling the designed ACC controllers to handle the measurement uncertainties is one of the main research focuses in this study. Therefore, not only ACC controllers which take actions based on the assumption of perfect information but also controllers which can directly cope with inaccurate measurements will be designed. Hence, three types of ACC systems are developed:

- ACC systems containing controller(s) designed with accurate sensor measurements (section 3.4)

- KF-ACC systems containing the same controller(s) as those designed for the ACC systems and KF(s) as the state estimator(s) (section 3.5)

- LSTM-ACC systems containing controller(s) designed with uncertain sensor measurements in several levels (section 3.6)

The first type of systems is equipped with controllers designed based on the assumption of accurate measurements. To handle measurement uncertainties, the second type of systems uses the same controllers designed previously and additionally adopt Kalman filters to estimate the state information before using it by the controllers.

Different from the KF-ACC systems whose controllers can only passively make decisions based on the input information filtered by the state estimator, the third type of systems is LSTM-ACC systems which possess the ability to take actions by actively considering the measurement uncertainties. The controllers in this type of systems are trained with uncertain measurements. An LSTM network layer is implemented into these controllers directly to serve as the hidden state estimator. By doing so, the DRL policies used in these controllers become the so-called recurrent policies which are based on POMDPs. The improvements or potential drawbacks of this kind of controller design method will be discussed in chapters 5 and 7.

There will be two different systems with different number of leaders in each category mentioned above. It is also important to note that there will be multiple two-leader KF-ACC systems and two-leader LSTM-ACC systems, while each of them is designed or tuned for a specific level of measurement uncertainties. Furthermore, this study aims to explore the potential trade-off between string stability and ride comfort. To create this kind of diversity among the proposed systems, each system introduced in this section has two versions with different weighting setups on the gap error and jerk terms in the reward function design during the training of the controllers, which will later on be introduced in subsection 3.4.2. Table 3.1 below again summarizes the design of the twelve systems proposed.

Table 3.1: Overview of the proposed ACC systems

System	# leader(s)	Weightings	Controller(s)	State estimator(s)
ACC systems	1	0.5/0.5	Trained with accurate measurements	None
		0.9/0.1		
	2	0.5/0.5		
		0.9/0.1		
KF-ACC systems	1	0.5/0.5	Same controllers as in the ACC systems	Kalman filters
		0.9/0.1		
	2	0.5/0.5		
		0.9/0.1		
LSTM-ACC systems	1	0.5/0.5	Trained with uncertain measurements	LSTM networks
		0.9/0.1		
	2	0.5/0.5		
		0.9/0.1		

3.2. Introduction to deep reinforcement learning

RL is classified as a machine learning approach which teaches agents how to solve a sequential decision-making problem through trial-and-errors. Based on the reward or penalty obtained from the environment, the learning agent explores how to make the appropriate decisions and take actions in particular states. DRL stems from the combination of RL and the concept of deep learning. Besides the field of vehicle automation, it has also been applied to many studies related to traffic signal control and ramp metering in recent years (Han et al., 2022; H. Wei et al., 2021). This section intends to provide a brief overview of the fundamentals of DRL and introduce the selected algorithm.

3.2.1. Markov decision process and Bellman's equation

In RL, agents learn by interacting with the environment. The process of RL can be mathematically described as an MDP. An MDP can also be understood as a Markov chain with rewards and decisions (actions). A Markov chain describes the stochastic transitions between states S based on a defined transition probability matrix P. The transitions possess the Markov property, meaning the future state is independent of the given past states. In a Markov reward process (MRP), rewards R are assigned to provide judgements on the transitions. When the set of agent's decisions or actions A are further considered in the MRPs, they become the so-called MDPs. An MDP can be described as a tuple (S, A, P, R, γ). To determine the action at each given state, a policy π comes into play. π describes the probability distribution over actions at the current state where the agent is.

The essence of RL is the training of the optimal policy π^* by considering not only the immediate reward but also the future rewards the agent can receive. Hence, the objective is to maximize the expected cumulative reward, which is known as the return G, by conducting a sequence of actions. For each reward in the sequence, a discount is given to consider their importance by using a discount factor $\gamma \in (0,1)$. A larger discount factor means a larger importance and influence the future states are for the agent. The calculation of the return G can be written as

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \tag{3.1}$$

To derive the optimal policy which generates the largest return, it is important to know the value of each state or state-action pair. Therefore, the value function of a state and the action-value function (Q-function) of a state-action pair should be formulated. Value function represents the expected return which the agent can gain by starting from a given state s and acting according to a given policy π throughout the whole trajectory, while Q-function indicates the expected return an agent can gain after taking a specific action a at a given state s and again following a determined policy π. Bellman's equation provides a mathematical way to formulate the value function

$$V^{\pi}(s) = \underset{\substack{a \sim \pi \\ s' \sim P}}{E} \left[R_{t+1} + \gamma V^{\pi}(s') \right] \tag{3.2}$$

and the action-value function

$$Q^{\pi}(s,a) = \underset{s' \sim P}{E} \left[R_{t+1} + \gamma \underset{a' \sim \pi}{E} \left[Q^{\pi}(s',a') \right] \right]. \tag{3.3}$$

In these equations, the next state (s') is sampled from the defined transition probability distribution (P); the current action (a) and the next action (a') are both sampled from the current policy (π). Eq 3.3 is formulated according to an on-policy RL algorithm, indicating that the policy used to make decisions is the same as the one which is being optimized.

Instead of the value function or Q-function which shows the absolute quality of a state or state-action pair, some RL algorithms need to know how good a specific action is than other actions based on the given policy at the current state. An advantage function of taking the action a at state s can be formulated for this kind of algorithms. There are many different ways to describe an advantage function, while

$$A^{\pi}(s,a) = Q^{\pi}(s,a) - V^{\pi}(s) \tag{3.4}$$

is one of the most straightforward example.

More formulations of the advantage function for various RL algorithms can be found in Schulman et al. (2015). Other in-depth explanations about the DRL algorithm used in this study will be introduced in subsection 3.2.4.

3.2.2. Partially observable Markov decision process

The RL problem introduced above can be suitably applied when it is assumed that the agent has complete access to the perfect information in the environment. The environment is called a fully observable environment when this is the case. However, the environment is not always fully observable for the agent in real-world scenarios. Partial observability (PO) refers to the situation that the agent can only receive part of the state information or noisy values from the environment. In a PO condition, the underlying state in the environment is not fully observable for the agent. Therefore, it is difficult to make decisions only based on the current state it observes.

To overcome the limitation of MDPs in handling information uncertainty, POMDPs are developed. In a normal MDP, the process satisfies the Markov property, meaning that the next state only depends on the current state and action conducted. However, a POMDP, which can be described as a tuple $(S, A, P, R, \Omega, O, \gamma)$, is built upon the modification of the original MDP. Figure 3.2 illustrates the process of a POMDP. Instead of using the state, a belief state b is determined by a state estimator (the SE in Figure 3.2) according to the current observation and the history, which includes all past observations and actions taken. Ω

represents the set of observations the agent can perceive; while O is the observation function which gives the probability distribution of possible observations based on the given resulting state.

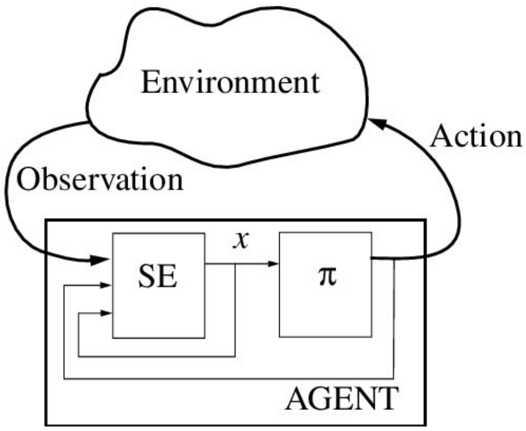


Figure 3.2: Process of a POMDP (Littman, 1996)

As shown in the figure, the history of process needs to be known by the agent so that it can determine the belief state. Hence, to derive the optimal policy, it is important to provide the agent with a "memory" of the past. The method used to cope with the PO condition in the training process in this problem context will be introduced in section 3.3.

3.2.3. Deep neural network agent

Different from traditional RL methods, DRL methods use a deep neural network (DNN) as a function approximator to represent the mapping between the states and actions decided. Figure 3.3 illustrates the interaction of a DRL agent and the environment. The DNN takes the state information and reward sent from the environment as the input and outputs a probability distribution of actions. One of the advantages of combining RL and DL over traditional RL is the ability to handle information in a high-dimensional state space, e.g., images and raw data from sensors. However, it also suffers some drawbacks of neural networks, such as data-efficiency and difficulty to be interpreted.

3.2.4. Reinforcement learning algorithm

Deriving the optimal policy network relies on a proper RL algorithm. RL algorithms are branched into model-based and model-free algorithms. In a model-based RL approach, the agent has access to a known or learned model of the environment (transition probability distributions and rewards). The primary task of the agent is to plan the optimal sequence of actions by using a predictive model. Although model-based RL approaches have better performance in terms of sample efficiency, they relies on a representative model which can help predict the outcome in the future. Such methods cannot work well when an accurate predictive model of the environment is not available. On the other hand, a model-free RL approach learns by directly interacting with the environment. Without any

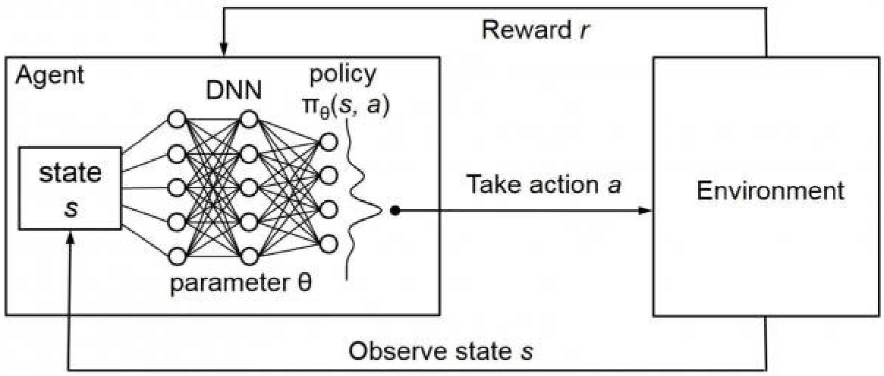


Figure 3.3: Conceptualization of deep reinforcement learning (Mao et al., 2016)

given knowledge from a model, this kind of methods requires more samplings to explore and interact with the environment. Nowadays, the model-free methods are receiving much more attention than model-based methods due to its easiness to implement and applicability to a wide range of problems. This study also uses a model-free RL method to train the optimal policy of the ACC controllers.

RL approaches can also be further divided into two categories, policy optimization and Q-learning, according to whether the action to be taken in the learning process is determined based on the most recent version of the policy which is optimized. In policy optimization methods, the policy being evaluated and improved is the same as the one which is being used to make decisions. Hence, this category is also called on-policy algorithms. Some examples include the classical Policy Gradient, Actor-to-Critic (A2C), Trust Region Policy Optimization (TRPO), and Proximal Policy Optimization (PPO). Whereas those Q-learning methods, such as Deep Q Network (DQN) and its variants, are often called off-policy algorithms because the updated policy is independent of the behavior policy. Each type of RL algorithms has its own pros and cons. On-policy algorithms are believed to be more stable, while off-policy algorithms are more sample efficient. To compensate the trade-off between these two categories, there are several algorithms which are proposed by borrowing the strengths of each side, including Deep Deterministic Policy Gradient (DDPG), Twin-delayed DDPG (TD3), and Soft Actor-Critic (SAC).

Given the relatively simple search space of action defined in the problem, this study selects PPO, an on-policy method which can be used for both discrete and continuous action spaces, as the RL algorithm to train the optimal policy network. It is proposed by Schulman et al. (2017) with the merit of being easier to implement than TRPO. PPO is an Actor-Critic method, implying that it has two neural networks, the actor and the critic. The actor stands for the policy network which determines the next action, while the critic estimates the value function of the state.

PPO updates the hyperparameter θ of the policy network via gradient ascent, as can be written as

$$\theta_{t+1} = \arg\max_{\theta} \underset{s_t,a_t\sim\pi_{\theta,t}}{E} \left[L^{CLIP}(s_t, a_t, \theta_t, \theta)\right], \tag{3.5}$$

where L^{CLIP} denotes the objective function, which is calculated by

$$L^{CLIP}(s_t, a_t, \theta_t, \theta) = \min\left(r_t(\theta)\hat{A}_t, clip\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right), \tag{3.6}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_t}(a_t|s_t)}$ is the ratio between the probability of choosing action a_t at s_t using the new policy π_θ and that using the current policy π_{θ_t}. If $r_t(\theta) > 1$, action a_t would be selected more often in the new policy than in the current policy, and vice versa. $\hat{A}_t$ represents the estimated advantage function of taking action a_t at state s_t.

However, when the action selection probability in the current policy $\pi_\theta(a_t|s_t)$ is low, the objective function becomes unstable due to the large probability ratio $r_t(\theta)$. The clipping function used in the objective function prevents the instability of the learning process by using a parameter ϵ to restrain the step size of the update.

For the critic network, the algorithm aims to minimize the error between the estimated value and the target value by updating the neural network parameters ϕ. The gradient descent method can then be applied, as written by

$$\phi_{t+1} = \arg\min_{\phi} E\left[L^{VF}(s_t, a_t, \phi)\right]. \tag{3.7}$$

The loss function L^{VF}, which calculates the squared error between the estimated value function and the target return value, is formulated by

$$L^{VF}(s_t, a_t, \phi) = \left(V_\phi(s_t) - \hat{G}_t\right)^2, \tag{3.8}$$

where $V_\phi(s_t)$ is the estimated value function with parameters ϕ. $\hat{G}_t$ represents the target return value. The general process of the PPO algorithm is described in Algorithm 1.

Algorithm 1. Proximal policy optimization

Initialize the actor network parameters θ_0 and the critic network parameter ϕ_0
for $t = 1, 2, \ldots$ **do**
  Generate sequences of actions with T time steps based on policy π_{θ_t}
  Calculate estimated advantage function $\hat{A}_t$ based on the current critic network V_{ϕ_t}
  Calculate the objective function L_t^{CLIP} when using policy π_{θ_t}
  Optimize and update the actor network parameters θ_{t+1} via gradient ascent
  Calculate target return $\hat{G}_t$
  Calculate the loss function L_t^{VF} when using value function V_{ϕ_t}
  Optimize and update the critic network parameters ϕ_{t+1} via gradient descent
end for

To achieve the learning of partial observability, this study further combines RNN structure into the PPO algorithm, which will be introduced in section 3.3. The procedure of the training would also become more complicated and slightly different. However, this section does not introduce the algorithm in detail given that it is not the major focus and a DRL library will be used in this study.

To sum up, this study aims to apply a model-free DRL method to train the optimal policies for the ACC controllers in the multi-leader ACC system. The settings for the learning agents, environment, and training process will be discussed in the following sections.

3.3. Deep neural network structures and parameters

As mentioned in section 3.2, the ACC controllers of the ego-vehicle would be the agents in the DRL framework. Each controller learns a DNN to map the state input to the output action. The learning process of PPO, the selected RL algorithm, consists of two DNNs, the actor and the critic. Both networks contain two hidden fully-connected layers, while each of which has 64 neurons. Figures 3.4 shows the actor-critic network structures of the controllers trained with full observability. The hyperbolic tangent (Tanh) fucntion is used as the activation function in every layer of the networks because its output range (-1 to 1) suits the range of vehicle acceleration values in this car-following problem.

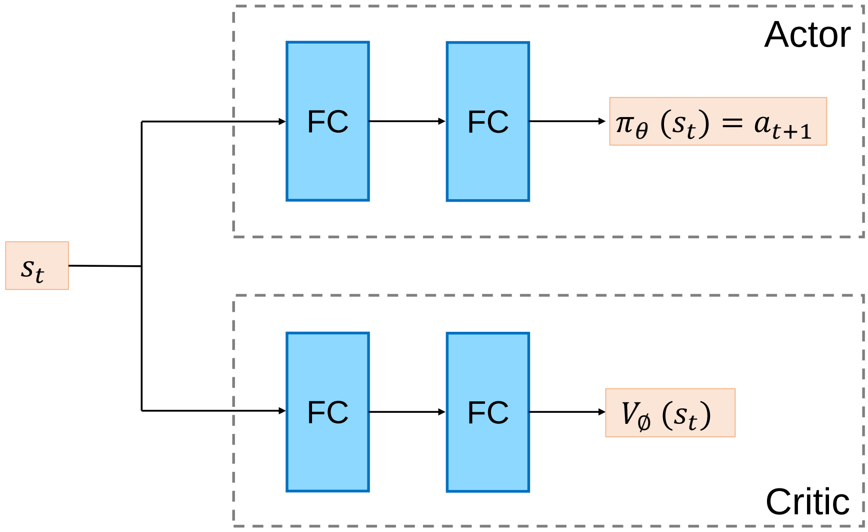


Figure 3.4: Policy proximal optimization actor-critic network structure of the ACC controllers

For the training of LSTM-ACC controllers in partial observable environments, a shared LSTM layer with 128 units is implemented before the actor and the critic networks to form recurrent policies. The LSTM network is a special type of RNNs which has a chain structure enabling the model to learn through a sequence of data in the time series. Different from ordinary RNNs, an LSTM network resolves the issue of long-term dependency by using three gating mechanisms, including a forget gate, an input gate, and an output gate, to control the flow of information. Figure 3.5 gives an example of an LSTM unit. By doing so, the DRL problem becomes a POMDP instead of the original MDP. The LSTM layer is expected to infer the belief state at every time step using the history, current state, and current action as the input. Figure 3.6 shows the structure of the recurrent policy network in LSTM-ACC controllers.

The DNN setups and parameters used in the PPO algorithm are summarized in Table 3.2. Most of the parameters follow the default settings in the selected DRL algorithm library and suggestions in the original paper.

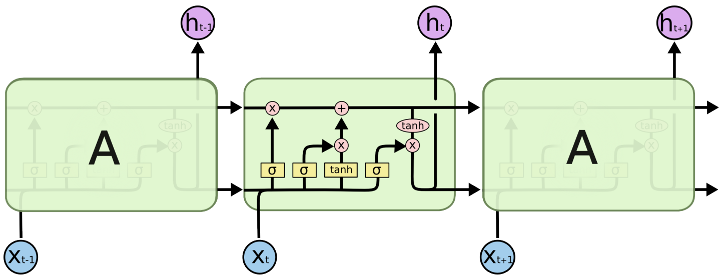


Figure 3.5: The internal structure of a Long-Short-Term Memory unit (Olah, 2015)

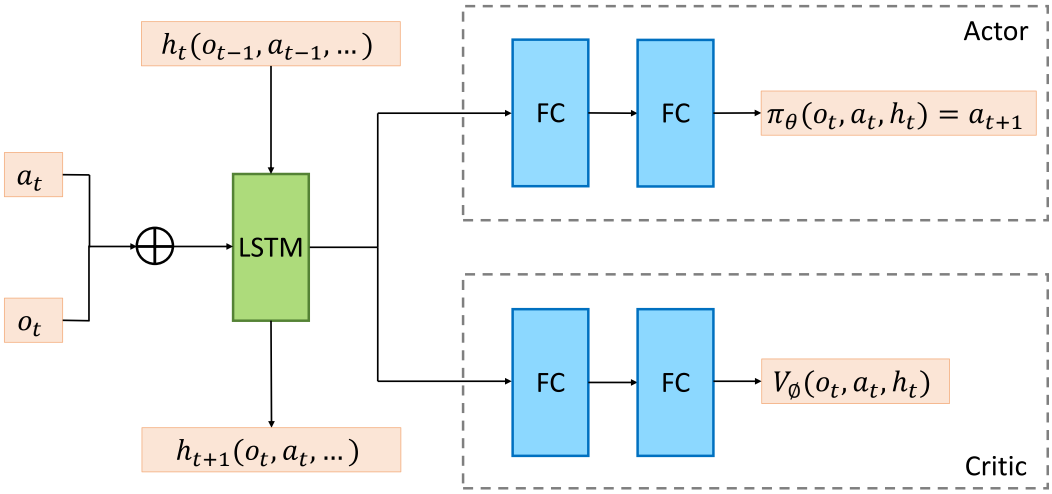


Figure 3.6: Recurrent policy proximal optimization actor-critic network structure of the LSTM-ACC controllers

Table 3.2: Deep neural network design and parameters in the learning algorithm

	Shared LSTM	Actor	Critic
# hidden layers	1	2	2
# units per layer	128	64	64
Activation function	—	Tanh	Tanh
Weights initialization	Orthogonal matrix		
Optimization algorithm	Adam		
Mini batch size	64		
Learning rate	0.0003		
# steps per update (T)	2048		
Discount factor (γ)	0.99		
Clip range parameter (ϵ)	0.2		

3.4. Policy training setup with accurate measurements

To apply DRL to develop the ACC controllers, several elements in the DRL policy have to be defined to suit the desired car-following task. In this section, the state and action spaces of the policy in the ACC controllers are formulated. Other settings regarding the training, including the reward function design and training tasks, are also described here.

3.4.1. State and action spaces

The state vector of the policy should contain enough information for the agent to make proper decisions. Several elements in the state vector are 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, while each component has a large enough state space to accommodate possible values. The state vectors of the ACC controllers can be written as

$$s_{i,k} = (g_{i,k}, v_k, \Delta v_{i,k}, j_k). \tag{3.9}$$

The distance gap, speed, and relative speed information enable the agent (ego-vehicle) to understand the current state which allows it to execute the desired car-following task, 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 the RADAR is described by

$$g_{i,k} = d_{i,k-\tau^S} - (i-1)\cdot l - i \cdot d_{min}, \tag{3.10}$$

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 jam (minimum) distance gap; τ^S represents the sensor delay; l is the vehicle length used 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

$$j_k = \frac{a_k - a_{k-1}}{\Delta t}, \tag{3.11}$$

where a_k and a_{k-1} denote the vehicle acceleration values at the current and previous time step, respectively.

The action defined in the DRL framework is the control input 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 vehicle acceleration a, taking reference from the range of aggressive driving behavior defined in Bae et al. (2019). These values are determined considering drivers' comfort and the ability of vehicle mechanical systems. In extreme safety critical situations, an even larger deceleration value may be required. However, it is not considered here in this problem context and the major purpose of the proposed ACC systems.

The values of the parameters used in the ACC controllers are summarized in Table 3.3. In the training of the policies, the sensor delay τ^S is equivalent to zero so that the agent learns the correct behavior for the corresponding state observation without considering the influence of the delayed information. The length of vehicles is assumed to be 4 m in this study. In reality, the length of the first leader should be estimated by using the contour of the detected vehicle or advanced image recognition technique.

Table 3.3: Parameter values used in the policy training

Parameter	**Value**
Sensor delay (τ^S)	0 s (training), 0.2 s (experiment)
Vehicle length (l)	4 m
Minimum distance gap (d_{min})	2 m
Time step size (Δt)	0.1 s
Std. of distance gap (σ_g)	scenario-specific
Std. of relative speed (σ_v)	scenario-specific
Maximum acceleration (u_{max})	3 m/s^2
Minimum acceleration (u_{min})	-6 m/s^2

3.4.2. Reward function

Many different spacing policies and their effect have been investigated in previous ACC studies. This study selects the constant time gap (CTG) policy, which is one of the most commonly-discussed spacing policies. The ACC controllers aim to help the ego-vehicle maintain the desired time gaps with its two leaders ahead.

The reward function in the designed DRL framework 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 (1978), it was pointed out that limiting the jerk has destabilizing effect on the longitudinal dynamics of the following vehicle. Therefore, the reward function should be able to guide the system to achieve a certain level of balance between these two factors. For ACC controller agents in this study, a negative reward function

$$R_{i,k} = \alpha \cdot \frac{|e_{i,k}|}{e_{max}} + \beta \cdot \frac{|j_k|}{j_{max}} + \min(\frac{e_{i,k-1} - e_{i,k}}{e_{max}}, 0). \tag{3.12}$$

is designed. A negative reward function implies that every element in the function is negative. In Eq 3.12, $e_{i,k}$ is the time gap error, which can be calculated by

$$e_{i,k} = tg_{i,k} - tg_i^*, \tag{3.13}$$

where $tg_{i,k} = \frac{g_{i,k}}{v_k}$ is the time gap of the ego-vehicle with the ith leader at k time step, and tg_i^* is the desired time gap with the ith leader determined according to the given spacing policy. e_{max} and j_{max} are values specified to normalize the gap error term and the jerk term, respectively. In this reward function design, $e_{max} = \frac{tg^*}{2}$ and $j_{max} = \frac{(u_{max}-u_{min})/3}{\Delta t}$. α

and β are the weights of the the first two elements, respectively. There is no weighting for the third element since it serve as a penalty term in the function. In this study, two weighting combinations are considered:

- Weighting combination 1: $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$
- Weighting combination 2: $\alpha = \frac{9}{10}$ and $\beta = \frac{1}{10}$

The two combinations aim to highlight the hypothetical trade-off between the two factors in the reward function and also help observe how the string stability and ride comfort performance are influenced by the different weighting setups.

Based on the negative reward design, the policy would seek to attain zero reward as fast as possible so that the ego-vehicle can reach the equilibrium and stable state (desired time gap and zero relative speed). In addition, the reward function should be able to provide a gradient for the agent to understand whether it is getting closer to the desired state or not. Therefore, every component in the function is continuous.

The proposed ACC systems in this study follow the constant time gap policy as the control goal in their controllers. The desired time gap (tg^*) in the policy of each ACC controller should be determined and used in the reward functions. Wang et al. (2017) showed that a linear state feedback ACC controller can preserve string stable performance but with a limited range of feedback gain parameters when driving at 1 s time gap with 0.2 s sensor delay and 0.2 s actuator lag in a homogeneous platoon. To investigate whether the proposed system can also achieve this level of performance, 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 slightly too dangerous given the reaction time required by human drivers when they have to take-over the control of the vehicle if a safety-critical situation occurs.

A negative value ($P = -100$) would be added as a 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 an collision ($tg_k < 0$), an extremely long time gap ($tg_k > tg^* + 5$), or a negative speed of the ego-vehicle ($v_k < 0$). The value of P should be sufficiently small so as to assign penalty to the incorrect behavior, but not too small which may cause a large difficulty for the training of the DNN agents.

3.4.3. Training task design for controllers with full observability

In the training process of ACC controllers, each episode simulates a 30 s car-following task, which is equivalent to 300 time steps for the ACC system. The simulation approach will be introduced in subsection 3.7.1. This study first setup a stationary environment (constant speed of the leader) and seek to train the optimal policies by creating different initial conditions in the environment for the agent (ego-vehicle) to tackle. In each episode, the leader keeps a constant speed, while the ego-vehicle starts with a randomized initial speed and time gap so that the agent can explore various car-following tasks. For instance, an initial state with a positive relative speed or a time gap larger than the desired value in the spacing policy trains the ego-vehicle how to approach a fast leader; an initial state with a negative relative speed or a time gap shorter than the desired value teaches the ego-vehicle how to properly adjust its speed profile to reach back to the equilibrium state.

In each episode, the speed of the leader is selected in a range between 15 m/s (54 km/h) and 35 m/s (126 km/h), which is approximately equivalent to the uncongested motorway speed range, as 120 km/h and 50 km/h are used as the design speed and the threshold speed between uncongested and congested traffic flow in Dijker et al. (1997). The applicability of the trained policy in situations outside the specified speed range will be further examined. All the detail regarding the initialization of each training episode is described as follows:

- Constant speed of the leader: [15 m/s, 35 m/s]
- Initial relative speed with the ith leader ($\Delta v_{i,0}$): [-3 m/s, 3 m/s]
- Initial time gap with the ith leader ($tg_{i,0}$): [$tg_i^* - 0.5$ s, $tg_i^* + 3$ s]

The combination of every initial condition explored during the training process contributes to the final output policy altogether. Note that a few studies using RL to develop ACC controllers have the ego-vehicle learn in a non-stationary environment, which means the leader may have a trajectory with an oscillating acceleration profile. This kind of method asks the ego-vehicle to learn the state transition probability and the control policy without knowing the exact future movement of the leader. In this kind of setting, it may be challenging for the agent to make the correct decision based on rewards obtained from the varying environment. In addition, determining sufficient and representative training tasks which ensure the applicability of the output policy is relatively difficult. Therefore, this kind of method is not preferred for the training of agents with full observability in this study. On the other hand, this kind of method may work in studies developing CACC systems, in which the ego-vehicle has a certain extent of knowledge of the next action (acceleration command) of the leader according to the in-vehicle information received through wireless vehicle communication technologies.

Each training episode would be terminated if the termination criteria is met, as the conditions mentioned in the previous subsection. This measure avoids the agent from continuing the exploration in situations which are unlikely to happen and make use of the training process efficiently.

3.5. State estimation using Kalman filter

When there is uncertainty in the sensor measurements, a KF can be used to derive an estimation of the true underlying state of the second leader measurements. It can also be understood as a method to track the movement of the leading vehicle. For actual RADAR measurements, an EKF should be applied to handle the non-linearity in the state transformation for measurements in polar coordinates, which ruins the Gaussian property. However, this study aims to explore the influence of uncertainty in range (distance gap) and range rate (relative speed) measurements for the multi-leader ACC application. Therefore, this study uses a linear KF instead of EKF for the ego-vehicle to track its leaders since the azimuth measurement is not of major interest in the defined problem context. For completeness, this section reiterates the steps and equations of the basic discrete linear KF for vehicle longitudinal motion tracking which can be used by the proposed ACC systems according to the material (Welch and Bishop, 2006).

In this problem, the state vector includes the distance gap between the leader and follower, relative speed, and relative acceleration. Hence, the estimated state of the ith leader at time step k is formulated as a vector

$$\hat{z}_{i,k} = \begin{bmatrix} g_{i,k} \\ \Delta v_{i,k} \\ \Delta a_{i,k} \end{bmatrix}. \tag{3.14}$$

The standard KF is a recursive process, which consists of two main steps, the prediction of state and update of the state estimation using measurements. Assumptions regarding the noise for the initial state, process dynamic, and measurements are determined beforehand. This method is believed to be a practical and state-of-the-art approach to deal with the potential sensor measurement uncertainties nowadays.

The steps and parameters in KF are describes as follows. The assumed process dynamic of the system is first formulated to produce the predicted state vector in the next time step. In this study, the leading vehicle motion is assumed to be a constant acceleration movement following Newton's law of motion. Therefore, the prediction of mean of the state vector in matrix form can be derived by

$$\hat{z}^-_{i,k} = F \cdot \hat{z}_{i,k-1}, \tag{3.15}$$

where $\hat{z}^-$ is the predicted state estimation, and F represents the state transition, which is calculated by

$$F = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.16}$$

The predicted error covariance matrix of state variables P^- can be calculated by

$$P^-_{i,k} = F \cdot P_{i,k-1} \cdot F^T + Q. \tag{3.17}$$

In Eq 3.17, Q represents the process noise covariance, which represents the uncertainty in the process dynamic. According to the assumption of the constant acceleration motion, there may be some noise in the leading vehicle acceleration. Hence, Q can be derived by projecting the random variance of the acceleration ${\sigma_a}^2$ on the process dynamic using the state transition matrix F, as shown in

$$Q = F \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & {\sigma_a}^2 \end{bmatrix} \cdot F^T = \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 & \Delta t \\ \frac{\Delta t^2}{2} & \Delta t & 1 \end{bmatrix} \cdot {\sigma_a}^2. \tag{3.18}$$

After predicting the mean and variance of the state vector, the output variables $\hat{y}^-$ need to be derived from the predicted state vector $\hat{z}^-_{i,k}$ using the observation matrix, as shown by

$$\hat{y}^-_{i,k} = H \cdot \hat{z}^-_{i,k}, \tag{3.19}$$

while

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{3.20}$$

Since RADAR provides distance and direct speed measurement, the observation matrix H is used to extract the predicted values of these two output variables from the predicted state vector.

The derived measurements in $\hat{y}^-$ will then be used to update the mean of the state vector. In the next step, the Kalman gain $G_{i,k}$, which is the weighting factor between the process dynamic and measurement, is calculated by

$$G_{i,k} = \frac{P^-_k \cdot H^T}{H \cdot P^-_k \cdot H^T + R}. \tag{3.21}$$

To achieve the lowest variance for the state, the optimal Kalman gain can be calculated as the ratio of the degree of uncertainty in our assumed process dynamic and the degree of uncertainty in the observation. In Eq 3.21, R denotes the measurement covariance.

After calculating the Kalman gain, the mean and variance of the state vector at current time step can be updated, as shown in

$$\hat{z}_{i,k} = \hat{z}^-_{i,k} + G_{i,k} \cdot (y_{i,k} - \hat{y}^-_{i,k}) \tag{3.22}$$

and

$$P_{i,k} = (1 - G_{i,k} \cdot H) \cdot P^-_{i,k}. \tag{3.23}$$

The updated mean of the state vector $\hat{z}_{i,k}$ can then be used in the DRL-ACC system. Both the mean and variance of the updated state would be used in the next time step to continue the estimation of the leader state information.

The initial error covariance $P_{i,0}$, process noise covariance Q, and measurement noise covariance R are parameters in the KF process. In this study, we assume the system has accurate information at the beginning. Therefore, $P_{i,0} = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$ for both $i = 1$ and $i = 2$. In addition, the noise covariance matrices stay constant throughout the whole tracking task. After several trials of tuning, the process noise covariance Q, which is derived from $\sigma_a = 0.5$ m/s^2, is found to have the best performance in terms of maintaining string stability considering the case of a typical traffic disturbance, which will be introduced in chapter 4. The measurement noise covariance matrices R would also follow the scenario settings, which will also be discussed later.

Note that this study adopts the conventional KF method. As mentioned in subsection 2.5.2, several studies have proposed tracking methods using adaptive KF, which changes the measurement and process noise covariance matrix dynamically during the task (Mehra, 1972). In this context, random noise of the leader acceleration may also become a time-dependent value σ_{a_k} which is changed dynamically so that the tracking task can adapt to changes in the target vehicle movement more quickly by using the adapted KF than it does when using the conventional KF method.

3.6. Recurrent policy training setup with uncertain measurements

To tackle the measurement uncertainty, the other method adopted in this study is the utilization of RNNs, as introduced in section 3.3. LSTM-ACC controllers can hence be formulated. Different from the previous ACC controllers which are trained with accurate measurements, these controllers are trained directly with uncertain measurements. Therefore, the input for the DNN agents has to be changed. Furthermore, a completely different setup is applied to design the training task.

3.6.1. Observation space

To train the LSTM-ACC controllers with noisy state information, the observation o at each time steps would replace the true underlying states s, as described by

$$o_{i,k} = (g'_{i,k}, v_k, \Delta v'_{i,k}, j_k), \tag{3.24}$$

where $g'_{i,k}$ represents the distance gap with random 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 RADAR range and range rate measurements to represent the measurement uncertainty. Since the RADAR sensor directly collects and calculates these two measurements through different approaches, the error terms of the distance gap and relative speed follow two independent zero-mean Gaussian distributions with standard deviations σ_{g_i} and σ_{v_i}, respectively, as shown by

$$g'_{i,k} = g_{i,k} + \varepsilon \sim N(0, \sigma_{g_i}^2) \tag{3.25}$$

and

$$v'_{i,k} = v_{i,k} + \varepsilon \sim N(0, \sigma_{v_i}^2). \tag{3.26}$$

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

3.6.2. Training task design for controllers with partial observability

For the training of LSTM-ACC controllers, the training tasks are different from those of the ACC controllers. The implementation of the LSTM enables the agent to predict the next state. In this problem context, the ACC controller of the ego-vehicle would be able to predict the distance gap and relative speed with the target leader. The original constant speed leader behavior setup designed for the training of ACC controllers with full observability cannot leverage the benefit of using LSTM and deteriorates its prediction ability. Therefore, instead of training the LSTM-ACC controllers in a stationary environment, this study seeks to train these recurrent policies in a non-stationary environment.

It is worth noting that RL in a non-stationary environment has been a challenging topic for the field of artificial intelligence (Alegre et al., 2021). It is difficult for the agent to understand the state transition dynamics if the environment changes too frequently or the changes diverse a lot. The approach of using the prediction ability of LSTM to overcome this potential difficulty adopted in this study is one class of the methods for RL to cope with non-stationarity.

When creating non-stationary driving behaviors for controllers to learn, traffic disturbances with different amplitudes of acceleration/deceleration are considered. At the beginning of each episode, the leader and ego-vehicle starts from the equilibrium state in which the two vehicles have a same speed and keep the desired time gap. The leader will then speeds up or slow down at a certain moment in the episode, which may be caused by the cut-in and cut-out behavior of a vehicle in front of the platoon leader. The leader stays at the resulting speed for a certain time length and starts to fix back to the original speed by either accelerating or decelerating with a smaller rate than the disturbance itself. Such a leader behavior leads to a shockwave which propagates upstream along the platoon. In this training setup, it is assumed that the disturbance and speed-fixing period have a constant acceleration. The disturbance of each episode is randomly generated following the ranges of values described below.

- Initial speed of both the leader and ego-vehicle: [15 m/s, 35 m/s]
- Starting point of the disturbance: [2 s, 4 s]
- Time length of the disturbance: (0 s, 5 s]

- Acceleration/deceleration of the disturbance (a_d): [-4 m/s^2, 2 m/s^2]
- Time length of the resulting high/low speed period: [0.5 s, 8 s]
- Acceleration/deceleration of the speed-fixing behavior:

$$\begin{cases} [-a_d \text{ m/s}^2, -\frac{a_d}{3} \text{ m/s}^2] & \text{for } a_d > 0 \\ [-\frac{a_d}{3} \text{ m/s}^2, -a_d \text{ m/s}^2] & \text{for } a_d < 0 \end{cases}$$

Figures 3.7 and 3.8 illustrate two examples of the leader behavior for the training of the LSTM-ACC controllers. The first example shows a traffic disturbance caused by an accelerating behavior of the leader, while the second one depicts a disturbance caused by a strong deceleration of leader. After the resulting high/low speed period, these vehicles try to reach the original speed with a determined acceleration or deceleration.

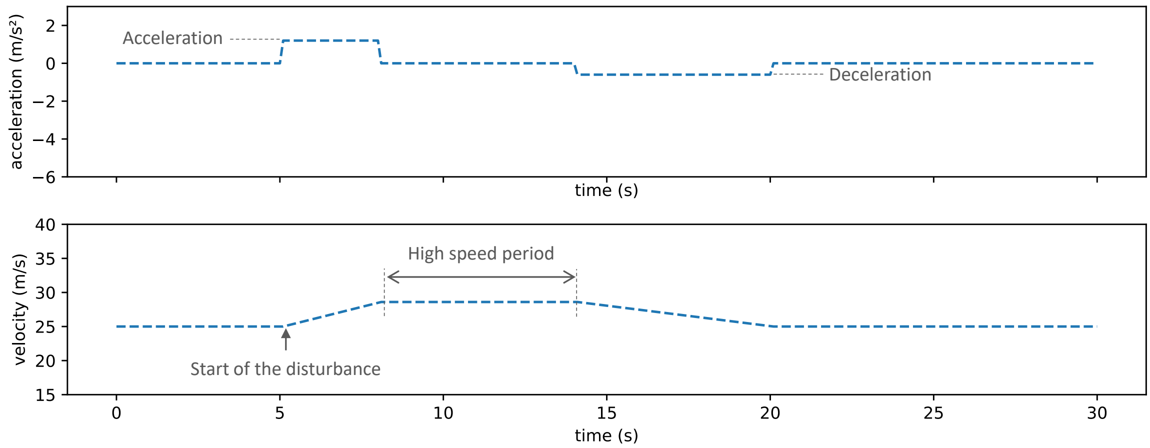


Figure 3.7: The first example of the leader behavior with initial speed = 25 m/s, accelerating disturbance with 1.2 m/s^2 occurring from 5 s to 8 s, 6 s high speed period, and decelerating with -0.6 m/s^2

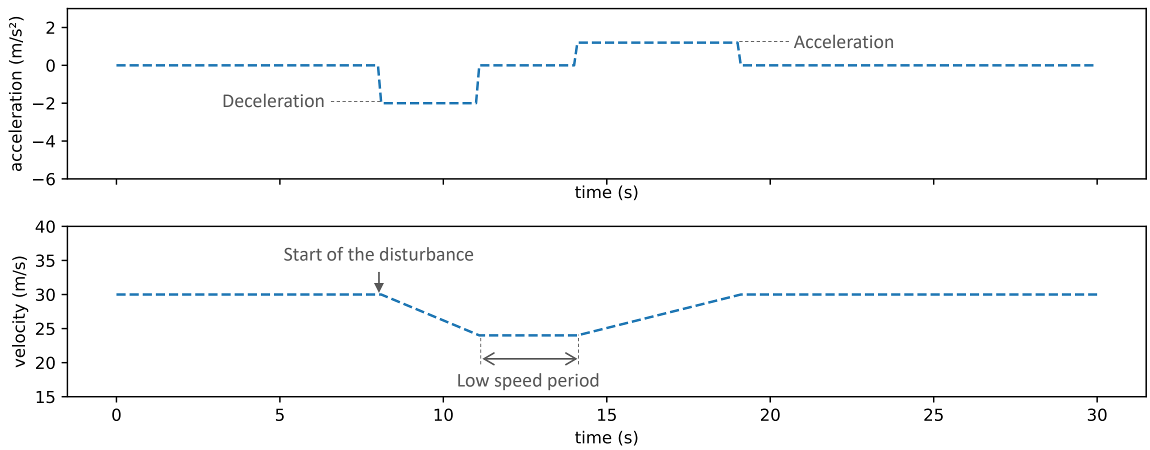


Figure 3.8: The second example of the leader behavior with initial speed = 30 m/s, decelerating disturbance with -2 m/s^2 occurring from 8 s to 12 s, 2 s low speed period, and accelerating with 1.6 m/s^2

The resulting speed caused by the oscillation would be bounded between 11 m/s (40 km/h) and 39 m/s (140 km/h) to match the operational design domain of the proposed multi-leader ACC systems specified in the training of ACC controllers. The acceleration or deceleration amplitude of the disturbance would be recalculated based on the bounded resulting speed and the randomly generated time length of the disturbance if the original resulting speed exceeds the boundary.

3.7. Simulation and training tools

The previous sections explain the theory behind the selected system design methods and the setup for the defined problem. In this section, the simulation approach and programming tools used for the design of the ACC controllers and the simulation experiment afterward will be introduced.

3.7.1. Numerical simulation approach

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

$$\begin{cases} x_k = x_{k-1} + v_{k-1}\Delta t + \frac{1}{2}a_{k-1}\Delta t^2 \\ v_k = v_{k-1} + a_{k-1}\Delta t \\ a_k = a_{k-1} + (u_k - a_{k-1})\Delta t/\tau^A \end{cases} \tag{3.27}$$

is formulated. In Eq 3.27, x_k is 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 adopted, as the value used in Wang et al. (2018) and Xiao and Gao (2011). The model updates the position, speed, and acceleration of every simulated vehicle according to the action carried out at every time step.

The state information introduced in section 3.4 can be derived from the updated position, speed, and acceleration of the leader and follower. For instance, the distance gap ($d_{i,k-\tau^S}$) is the difference between the positions $x_{k-\tau^S}$ of the ith leader and that of the ego-vehicle. These equations mathematically describe the movement of vehicles and the transition between states. Based on the action carried out and the car-following dynamics simulated, the reward function value can then be calculated.

3.7.2. Training tool and setup

The DRL training and numerical simulation (DRL environment) are both built using OpenAI Gym (Brockman et al., 2016) in Python 3.7.11, which is suited for Stable Baselines (Hill et al., 2018), the selected DRL algorithm library using Tensorflow 1.15.0 as the deep learning package. Note that the new version of Stable Baselines (Stable Baselines 3), which supports PyTorch, was already released (Raffin et al., 2021). However, the recurrent policy networks are not yet supported in Stable Baselines 3 for the time being of this thesis project. Therefore, Stable Baselines 2.10.2 is adopted. The training of 10000 time steps takes around one minute on a laptop which has a 4-core (8 thread) Intel Core i7 CPU running at 2.80 GHz with 12Mb cache and an elementary Nvidia GeForce MX350 GPU.

For the training of the DNN agents with full observability (accurate measurements), the total number of time steps is 1500000. On the other hand, those agents with partial observability for LSTM-ACC systems are trained for 6000000 time steps (four times of the ACC controllers) since the LSTM layer contains much more neurons and parameters to be optimized. Therefore, a longer training process is necessary. During the training process, the performance of the learned model is estimated every 100000 time steps by simulating 100 randomly generated episodes. The estimated policy which has the highest average reward would be stored as the optimal policy and used in the simulation experiments for performance evaluation.

4

Experimental Design and Analysis

After designing the ACC systems, this chapter introduces how the proposed systems will be evaluated. The simulation approach, which has already been used in the system design, is again explained. The second section describes the scenarios in the experiment. Section 4.3 then summarizes all the scenarios and develops a framework for the evaluation and comparison afterward. The quantitative analysis framework which will be applied to analyze the output from the numerical simulation is introduced in the last section.

4.1. Simulation setup

As described in section 3.7, a numerical simulation approach is formulated to be used as the environment in the DRL method. The same approach is extended to carry out the designed experiments in this study. In the numerical simulation, the longitudinal dynamics of a 20-vehicle homogeneous platoon is simulated. The behavior of the first vehicle is predetermined, and the other following vehicles act according to the acceleration command generated from the ACC system. The time step size (Δt) of the simulation is 0.1 s. The time length of the simulation experiment depends on the designed car-following task, which will be described in the next section. The position, speed, and acceleration of each following vehicle at every time step will be stored for further analysis.

4.2. Experimental scenarios

In this section, the experiment including several scenarios is designed to evaluate the performance of these ACC systems. The leader behavior and level of sensor measurement uncertainties are the two main control variables for the experiment, while this study first considers a specific leader behavior with disturbance for simplicity in the performance evaluation.

4.2.1. Leader behavior with traffic disturbance

The car-following dynamics of the following vehicles depends on the behavior of the first vehicle in the platoon. To test the string stability of the ACC systems, a disturbance needs to be added to the behavior of the first vehicle to investigate the response of the following vehicles. From a traffic flow perspective, a speed fluctuation behavior of the leading vehicle 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 type of speed fluctuation described in T. Li et al. (2021) as the disturbance. This kind of leader behavior is also used to train the DRL-LSTM-ACC controllers in this study, as mentioned in subsection 3.6.2.

Figure 4.1 gives an example of the speed profile of the first vehicle in the platoon. In the created driving behavior, the first vehicle starts with a constant stabilization speed 33 m/s for 3 seconds. It then conducts a strong deceleration with -3 m/s^2 for 4 s and maintains at the low resulting speed for 5 s. Later on, it spends 8 s speeding up with 1.5 m/s^2 to reach the original stabilization speed. This is a comparatively aggressive braking and accelerating maneuver for human drivers, which can better help demonstrate the effectiveness of the proposed ACC systems in this study.

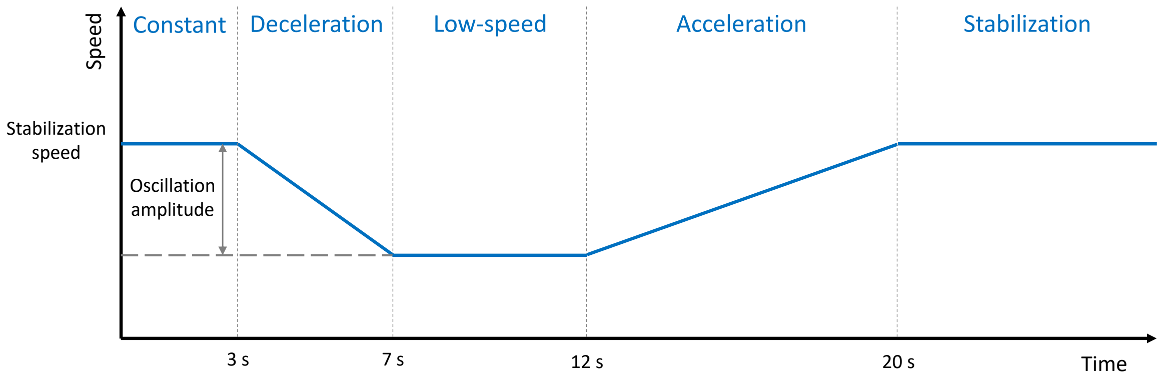


Figure 4.1: Oscillation speed profile of the first vehicle in the platoon

4.2.2. Sensor measurement uncertainties

Sensor measurement uncertainties are the other important aspect in this study. To explore the uncertainty boundary for string stability performance of the developed ACC systems, measurement noise is considered in the experiments.

Noise is one of the most common type of uncertainties for sensor measurements. It can come from the detection of different points of the leading vehicle, e.g. rear bumper or chassis of the vehicle. Different scattering radar point clouds lead to different estimations of the target vehicle contour and produce noisy range (distance) measurements. The second type of sources which can lead to measurement noise is the signal-to-noise (S/N) ratio, the ratio of reception power to noise power. Factors influencing the S/N ratio include the measurement time, frequency used, and the power of transmission and reception specified in the sensor. Measurement noise can also subject to other external factors related to the distance and speed of the target object or the environment. However, these external and environmental factors are excluded in this study due to the unavailability of data from which such features can be extracted.

Hence, random errors resulted from the noisy measurements mentioned above 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 m and σ_v m/s, respectively. The accuracy of the first leader information is fixed to $\sigma_{g_1} = 0.2$ m and $\sigma_{v_1} = 0.2$ m/s, which are determined according to the state-of-the-art automotive long-range RADAR accuracy summarized in Hasch et al. (2012) and Gamba (2020). This scenario which only considers the measurement uncertainties of the first leader is denoted by "N0" in this study, as it will also be applied in those two-leader systems.

The accuracy of the second leader measurements, however, is still unknown since it is a relatively new concept in the ADAS applications or ACC system development. For the simulation experiment and model training in this study, it is assumed that the second leader measurements would be more erroneous than the first leader measurements. According to Hasch et al. (2012), the standard deviations of the two measurements can be calculated based on the same S/N ratio. Hence, the standard deviations of the range and range rate measurements tend to increase together when the S/N ratio increases.

Given that the RADAR calibration and parameters remain unchanged, the S/N ratios vary in four different levels. Accordingly, the system performance will be evaluated in four different levels of measurement noise:

- N1: $\sigma_{g_2} = 0.5$ m and $\sigma_{v_2} = 0.5$ m/s
- N2: $\sigma_{g_2} = 1$ m and $\sigma_{v_2} = 1$ m/s
- N3: $\sigma_{g_2} = 1.5$ m and $\sigma_{v_2} = 1.5$ m/s
- N4: $\sigma_{g_2} = 2$ m and $\sigma_{v_2} = 2$ m/s

Each level has a standard deviation value for both the range and range rate measurements of the second leader. A RADAR sensor which has a larger level of measurement noise than level N4 would be regarded as an unreliable sensor for the multi-leader detection functionality and any other ADAS applications. Each scenario will be simulated 20 times to account for the randomness.

It is worth noting that there are indeed other sources of sensor measurement uncertainties which can be considered in this single-lane longitudinal driving case. For instance, the problem of losing detection (false negative) may often happen to the second leader detection task when the power of the received signal becomes too small and significantly affects the reliability of the multi-leader ACC system. However, this type of sensor measurement uncertainty creates another dimension to investigate and is hence left out in this study. More relevant discussion can be found in chapter 7.

4.3. Performance evaluation framework

To draw valuable research findings and answer the proposed research questions afterwards, a clear framework for the evaluation and comparison of the system performance is required. In chapter 5, the simulation and analysis results will be divided into four parts, which will be introduced as follows.

- Part 1 (section 5.2): This study emphasizes the importance and possibility of the design of multi-leader ACC systems. Therefore, the first part of the evaluation will focus on the the comparison between the performance of the one-leader ACC system and that of the two-leader ACC system in the scenario with accurate measurements. By doing so, the string stability, ride comfort performance, and car-following behavior mechanism of the two systems will first be investigated, as will be introduced in the next section. The benefit of multi-anticipation for the car-following dynamics in the platoon, which is also the fourth research question in this study, can also be discovered. The performance of these two systems under perfect information will become the baseline for the comparison with other systems in scenarios with uncertain measurements.

- Part 2 (section 5.3): In the second part of the evaluation, sensor measurement uncertainties are considered. It is important to understand the impact of different levels of uncertainties on string stability and ride comfort. Therefore, the ACC systems evaluated in Part 1 will then be simulated in scenarios with measurement uncertainties modelled as described in section 4.2.2. It is anticipated that the system performance will be degraded as the level of measurement uncertainties increases.

- Part 3 (section 5.4): The third part focuses on the effect of applying KFs as the state estimators in the ACC systems. Therefore, the KF-ACC systems using the same controllers of the previous ACC systems and the tuned KFs will be simulated to explore whether they can attain the same level of performance as the scenarios with accurate sensor measurements in Part 1 of the evaluation.

- Part 4 (section 5.5): The last part evaluates the performance of the LSTM-ACC systems which use recurrent policies to handle the noisy measurements. Besides of answering the question whether this type of systems can perform as good as the previously-proposed ACC systems in scenarios with accurate measurements, these systems will also be compared with the systems using KFs. It is worth investigating the difference between the effects of these two methods in terms of measurement uncertainty handling.

The structure in chapter 5 will follow this four-part performance evaluation framework. Table 4.1 provides an overview of each experiment and the whole evaluation framework. It should be noted that every system listed in the table has two different versions which are designed with different weighting combinations for the terms in the reward function of the DRL training setup.

Table 4.1: Overview of the system performance evaluation framework

Evaluation	Scenario(s)	System	# leader(s)	Weightings	Controller(s)	State estimator(s)
Part 1	Accurate measurements	ACC systems	1	0.5/0.5	Trained with accurate measurements	None
				0.9/0.1		
	Accurate measurements		2	0.5/0.5		
				0.9/0.1		
Part 2	Measurement uncertainty level N0		1	0.5/0.5		
				0.9/0.1		
	Measurement uncertainty levels N1-4		2	0.5/0.5		
				0.9/0.1		
Part 3	Measurement uncertainty level N0	KF-ACC systems	1	0.5/0.5	Same controllers as the ACC systems	Kalman filters
				0.9/0.1		
	Measurement uncertainty levels N1-4		2	0.5/0.5		
				0.9/0.1		
Part 4	Measurement uncertainty level N0	LSTM-ACC systems	1	0.5/0.5	Trained with uncertain measurements	LSTM networks
				0.9/0.1		
	Measurement uncertainty levels N1-4		2	0.5/0.5		
				0.9/0.1		

4.4. Quantitative Analysis

This section describes the quantitative analysis framework for the numerical simulation output. Initial hypotheses of the performance and effect of the proposed ACC systems on string stability, ride comfort, and overall car-following dynamics are required to develop this analysis framework.

4.4.1. String stability indicators

To investigate the string stability performance, some indicators which can represent such property should be defined to quantitatively evaluate the results of the simulation experiments conducted. For simplicity, this study only investigates the impact of disturbance on the speed of the following vehicles. First, the analysis focuses on the effect of the ACC system response on the propagation of speed fluctuation. The oscillation amplitude of each vehicle is computed by the difference between the stabilization speed (SS) and the lowest speed (LS) in the profile, while the oscillation growth amplitude (OG) is the difference between the oscillation amplitude of the first vehicle in the platoon and that of the follower considered . By looking at the propagation of the oscillation growth amplitude in the platoon, whether the disturbance is amplified or damped out can be observed.

Additionally, after the following vehicle speeds up in the acceleration phase, the speed of it may slightly exceed the SS. This phenomenon is known as the overshooting effect. It is also related to the string stability performance since a large overshooting amplitude can potentially lead to another disturbance and traffic oscillation. Hence, the overshooting amplitude (OS), which is calculated by the difference between the SS and the highest speed (HS) after the acceleration phase in the speed profile of the follower, would be the second indicator in the quantitative analysis framework. Both an increasing oscillation growth amplitude and an increasing overshooting amplitude along the platoon indicate a certain level of string instability. After applying the proposed systems, both indicators are expected to decrease as fast as possible in the upstream direction of the platoon.

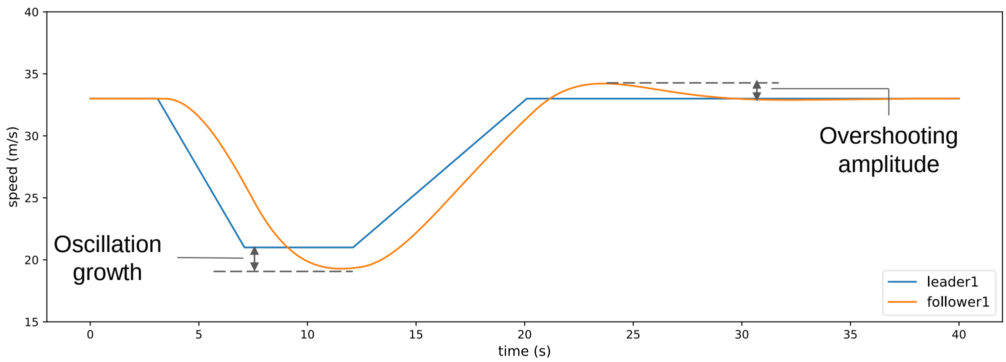


Figure 4.2: String stability indicators shown in the speed profiles of a following vehicle equipped with a non-optimized linear state feedback ACC controller

Figure 4.2 shows the speed profiles of an oscillating leader and a follower equipped with a non-optimized linear state feedback ACC controller ($k_1 = 0.2$ and $k_2 = 0.4$). It provides a graphical example of the two string stability indicators.

The string stability indicators of each following vehicle and their calculation methods are summarized:

- Oscillation growth amplitude (m/s): $OG_n = LS_n - LS_l$
- Overshooting amplitude (m/s): $OS_n = HS_n - SS$

Subscript l represents the first leading vehicle in the platoon, whose behavior is predetermined, while subscripts n denotes the number of the following vehicle within the platoon.

4.4.2. Ride comfort indicators

As mentioned in section 4.2, measurement noise is considered in the simulation experiment. The existence of noise may not only affect the string stability performance but also result in uncomfortable driving maneuvers (fluctuation in the acceleration profile). The state estimators, KF and the implemented LSTM network, are expected to help smooth out the noise in both the range and range rate 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. Although not being explicitly simulated in this study, commercial vehicles usually have a jerk limiting mechanism to preserve ride comfort and prevent severe damage on the hardware.

Hence, to evaluate the performance of the system when facing noisy measurements, the jerk amplitudes experienced by vehicles in the platoon will be recorded and analyzed. First, the probability distribution of jerk of all vehicles in all simulation runs will be calculated to discuss whether the system produces uncomfortable driving maneuvers when facing the disturbance and measurement uncertainty. The criteria and threshold of jerk are determined beforehand to define the boundary of comfortable and aggressive driving maneuvers. According to Bae et al. (2019), 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 2 m/s^3 is therefore considered abnormal and only occurs in emergency conditions. Figure 4.3 graphically shows the ride comfort levels and the threshold values.

By doing so, the impact of noisy measurements and the improvement brought by the adopted state estimation methods can be explored. It is also hypothesized that noisy measurements will definitely create large jerk amplitudes between consecutive time steps. A higher level of measurement uncertainties may further result in larger jerk amplitudes and hence more uncomfortable driving maneuvers.

As mentioned in subsection 3.4.2, jerk and stability may be two potentially contradicting performance indicators for a vehicle platoon. Whether this trade-off does exist in the

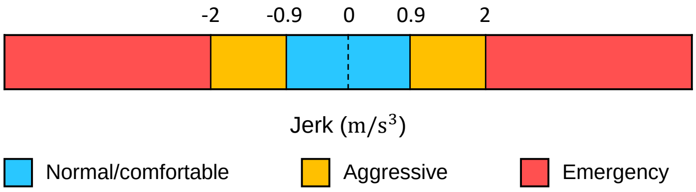


Figure 4.3: Definition of ride comfort levels based on jerk amplitudes

performance of the proposed systems and what kind of effect can the adopted state estimation methods or controllers bring for these two performance aspects will be discussed in chapter 5 by comparing the system performance before and after the implementations of the state estimators. The different weighting combinations used in the reward function of the DRL policy training setup may also provide insights into this aspect.

4.4.3. Car-following behavior mechanism

Other than investigating string stability and ride comfort by looking at these indicators, this study also seeks to explore the car-following behaviors of vehicles equipped with the proposed ACC systems by comparing the trajectories of the following vehicles with their hypothetical trajectories generated from Newell's car-following model (Newell, 2002). The method was originally adopted by Laval (2011) and Laval and Leclercq (2010) to investigate the behaviors of human driven vehicles. T. Li et al. (2021) also applied this method to analyze the behavior of ACC-equipped vehicles in empirical experiments.

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 trajectory of the leader and the follower should be identical except for a shift of space and time between them, as shown in Figure 4.4.

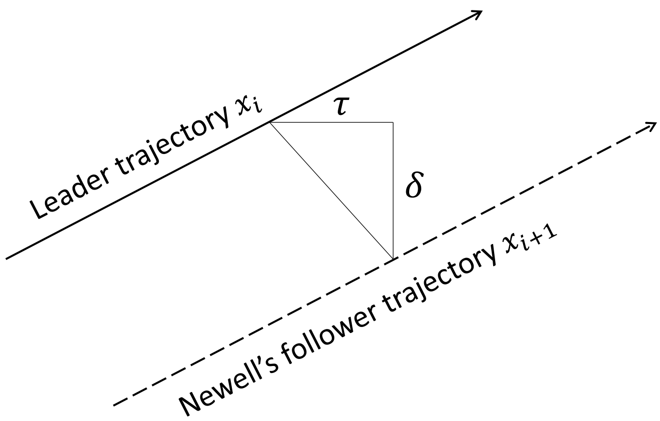


Figure 4.4: Follower trajectory generated based on Newell's car following model

Using the same notations, the mathematical formulation of Newell's car-following model can be described by

$$x_{i+1}(t) = x_i(t-\tau) - \delta. \tag{4.1}$$

According to the constant time gap spacing policy considered by the ACC systems in this study, the time shift τ would be 1 s, and the distance shift δ would be 6 m, which is also the jam headway calculated by summing the minimum distance gap (d_{min}) and vehicle length (l).

If the vehicle trajectory stays above the Newell's follower trajectory (black dashed lines in Figure 4.4), it represents an aggressive driving maneuver. In contrast, a trajectory below the Newell's trajectory indicates that the follower has a timid or conservative behavior. The position deviations of the following vehicles from the Newell's follower trajectories are calculated and recorded at every time step to help discover the behavioral response of the following ACC vehicles to the preceding traffic oscillation event.

5 Results

This chapter first presents the training performance of the ACC controllers and LSTM-ACC controllers. The learning processes of the agents under different weighting setups and different levels of uncertainties can be observed, which helps explain and verify the policy network training setup in chapter 3. The simulation output of the proposed systems in those scenarios defined in section 4.2 have been analyzed following the quantitative analysis framework proposed in section 4.4. The analysis results are then divided into different sections based on the systems considered. These sections provide preliminary and detailed discussions according to the observations from the analysis results.

5.1. Training performance

The curves in Figures 5.1 and 5.2 record the undiscounted cumulative reward and episode length at every evaluation point during the training process of the ACC controllers. The dots in the plot of cumulative reward represent the evaluated model with the highest average reward.

As can be seen in the plot of episode length on the right hand side of Figure 5.1, the learning agent of ACC controller 1 found the policy which successes to complete all the car-following tasks at the fifth evaluation point, while the fourteenth evaluated model is selected as the optimal policy. For the ACC controller 2, the algorithm managed to find the success policy within 100000 time steps of training , which is even before the first evaluation point. In general, the cumulative rewards of controller 2 are higher than those of controller 1 throughout the training process. The training of a successful policy is also easier for controller 2 than it is for controller 1. This is due to the setting in reward function design and the difference between their spacing policies. Since controller 2 has a larger desired time gap, the reward function values are often higher than those in controller 1. A larger time gap also allows it to prevent the training episode from being easily terminated due to the occurrence of a too large or too short time gap.

Figure 5.2 shows the training performance of the ACC controllers in the second weighting combination setup. It is found that the learning agent of the ACC controller 2 also has

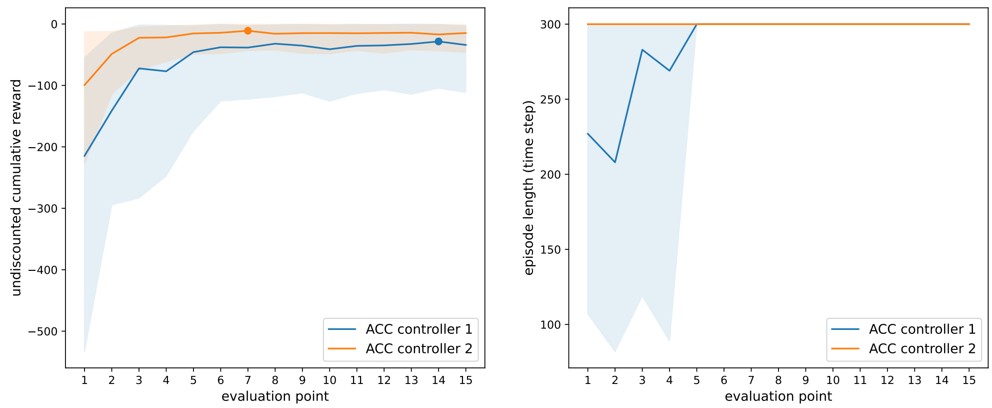


Figure 5.1: Undiscounted cumulative reward (left) and episode length (right) of the ACC controllers trained with the first weighting combination at every evaluation point in the training process

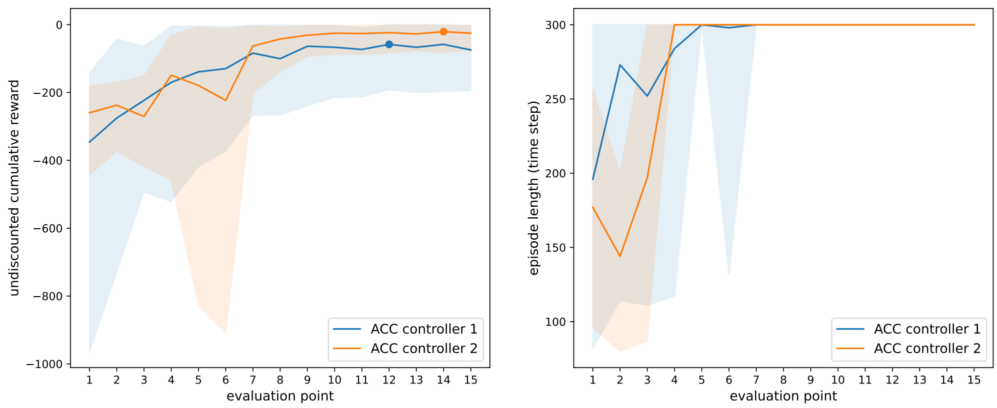


Figure 5.2: Undiscounted cumulative reward (left) and episode length (right) of the ACC controllers trained with the second weighting combination at every evaluation point in the training process

higher cumulative reward than the agent of ACC controller 1 at the end of the training process. However, there is no significant difference between the amounts of training needed for each agent to find the success policy. This implies that the second weighting combination setup, which focuses more on the gap-keeping performance than the ride comfort, makes it more difficult for the agent with full observability to learn how to complete the car-following task. More explorations are required for the agents in this reward function design.

On the other hand, the learning curves of recurrent policies show similar trend in the training process. As shown in the curve of episode lengths on the bottom of Figure 5.3, the algorithm found the policy which enables the LSTM-ACC controller 1 to finish every evaluation episode at the 15th evaluation point, while the trained policies of every LSTM-ACC controller 2 success to complete the entire simulation episode around the 5th point. For the training of the LSTM-ACC controller 2 under multiple levels of uncertainties, it is also found that the training performance is influenced by the level of uncertainties. In the

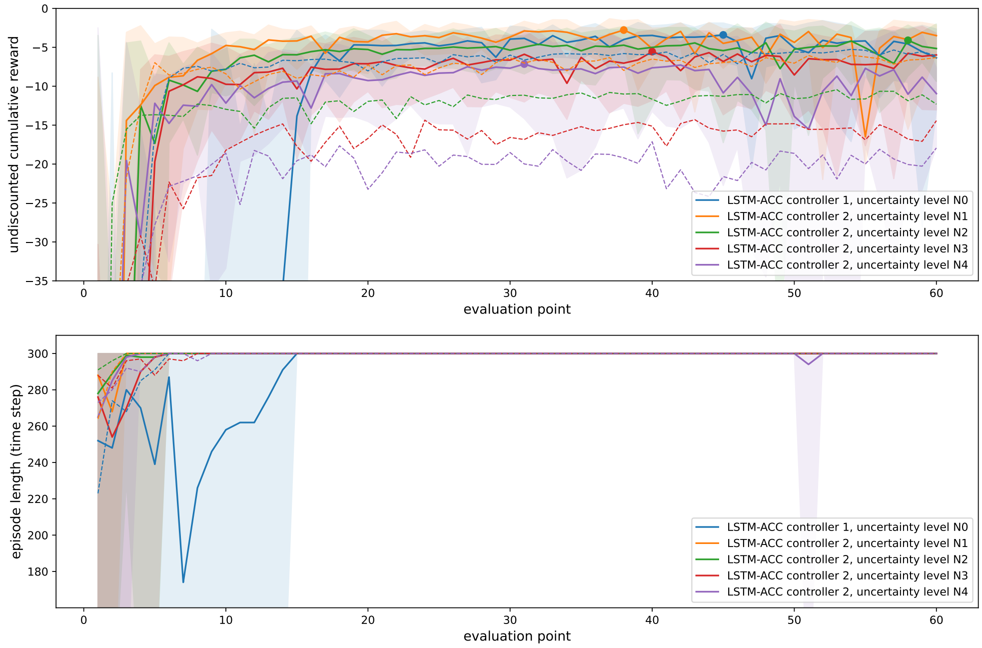


Figure 5.3: Undiscounted cumulative reward (top) and episode length (bottom) of the LSTM-ACC controllers trained with the first weighting combination at every evaluation point in the training process (dotted lines represent their counterparts without using the LSTM networks in the network architecture)

scenarios with lower level of uncertainties, the trained policy is able to achieve a higher undiscounted cumulative reward. In addition, the training with a higher level of uncertainties leads to more unstable training performance and larger variation of the average cumulative reward. For instance, the policy for LSTM-ACC controller 2 in measurement uncertainty level N4 fails to complete all the training episodes at the 51st evaluation point, as shown in both the curves of cumulative rewards and episode lengths, while this does not happen to the LSTM-ACC controller 2 in other uncertainty levels.

On the other hand, for the training of LSTM-ACC controllers which can handle measurement uncertainties, the positive effect of adding the recurrent layer into the DNN agent can also be demonstrated. The dotted learning curves in in Figure 5.3 represent the learning processes of four policies in the corresponding levels of uncertainties without using the LSTM network. According to the episode length curve, these policies may be able to find a success policy for the car-following task faster than their counterparts with recurrency since the comparatively simple DNN structure makes the training much simpler. However, it is found that the maximum cumulative rewards of these policies are all smaller than those of the recurrent policies. Moreover, the difference between the policy and the recurrent policy becomes larger at higher uncertainty levels. This implies the importance of making use of the history of states and actions and the power of memories to make decisions under uncertainty.

Figure 5.4 presents the training of LSTM-ACC controllers in the second weighting com-

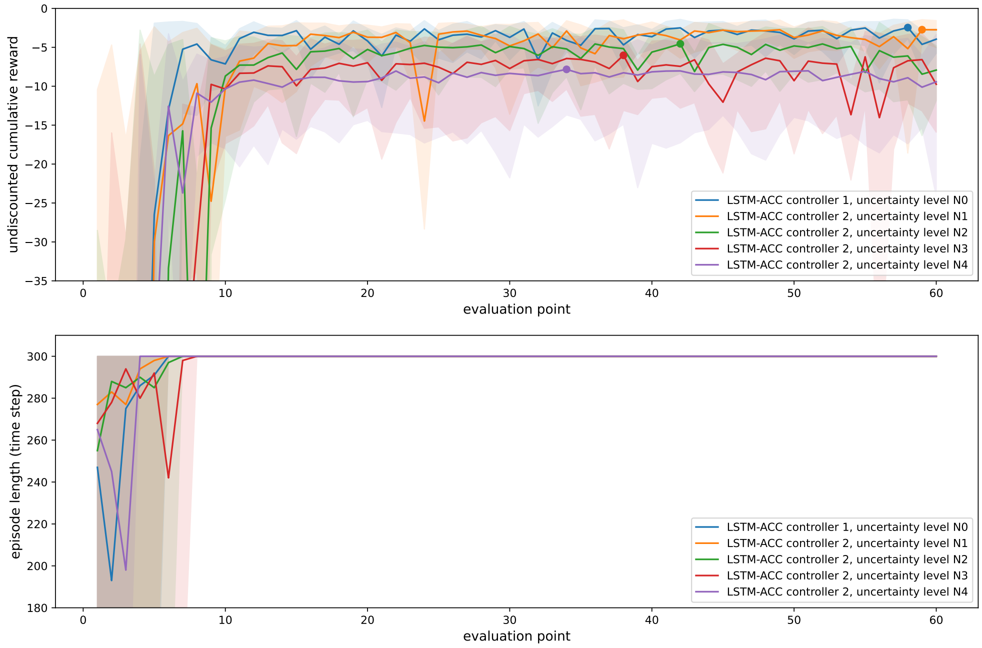


Figure 5.4: Undiscounted cumulative reward (top) and episode length (bottom) of the LSTM-ACC controllers trained with the second weighting combination at every evaluation point in the training process

bination setup. It also shows that the optimal cumulative reward is smaller when the measurement uncertainty level becomes higher. However, different from the trend observed in the training performance of ACC controllers, the second weighting combination setup makes it easier for the agent of LSTM-ACC controller 1 to find the successful policy than the first weighting combination setup does. The improvement may also be reflected in the final optimal policy performance, which will be discussed in subsection 5.5.1.

5.2. System performance with accurate measurements

This section shows the simulation and performance analysis results of the one-leader and two-leader ACC systems in scenarios with accurate measurements. The main purpose is to show the ability of DRL in designing ACC controllers and also explore the potential benefit of multi-anticipative car-following behavior of the two-leader DRL-ACC system.

5.2.1. Performance of one-leader ACC systems

The 20-vehicle platoon consisting of a human driven leader and nineteenth following vehicles equipped with the one-leader ACC system designed with the first weighting combination is first simulated in the created traffic oscillation case. Figure 5.5 shows the acceleration/speed/gap profiles of the first six vehicles in the platoon. As can be seen in the acceleration and speed profiles, it seems that the disturbances in the driving behaviors of the first five following vehicles are maintained at nearly the same level as the first vehicle (leader 1 in the figure). Although it is uncertain that whether the disturbance is damped out or not in the upstream of the platoon, it can be observed that the disturbance is not dramatically amplified. Furthermore, the time gap of each follower with their leaders does not deviate too much from the desired value (1 s) when encountering the disturbance and can be fixed back to the desired value in the stabilization phase. These two aspects demonstrate the success of DRL in developing the ACC controllers.

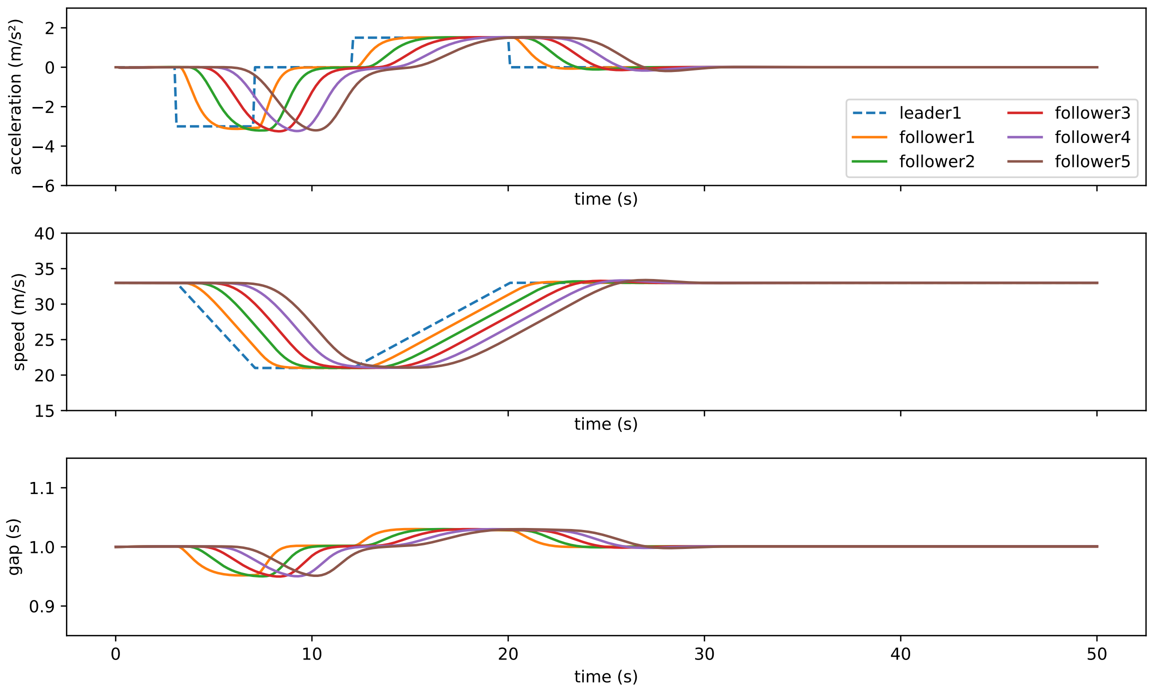


Figure 5.5: Acceleration, speed, and time gap profiles of the platoon using the one-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

The car-following dynamics of the whole platoon can be observed in Figure 5.6, which plots the trajectories and speed contours of every vehicle in the platoon. By looking at the speed contours, it is actually found that the minimum speed of the 20th vehicle seems to be slightly larger than the first vehicle according to the color of the speed contour although it is still difficult to observe any difference. This indicates that the oscillation amplitude is gradually reduced while propagating upstream. On the other hand, the speed profile in Figure 5.5 shows that the overshooting behaviors exist and are slightly amplified, as demonstrated by the dark blue line segments (around 34 m/s) in the stabilization phase of each trajectory in Figure 5.6. This will be further discussed by numerically investigating the string stability indicators in subsection 5.2.3.

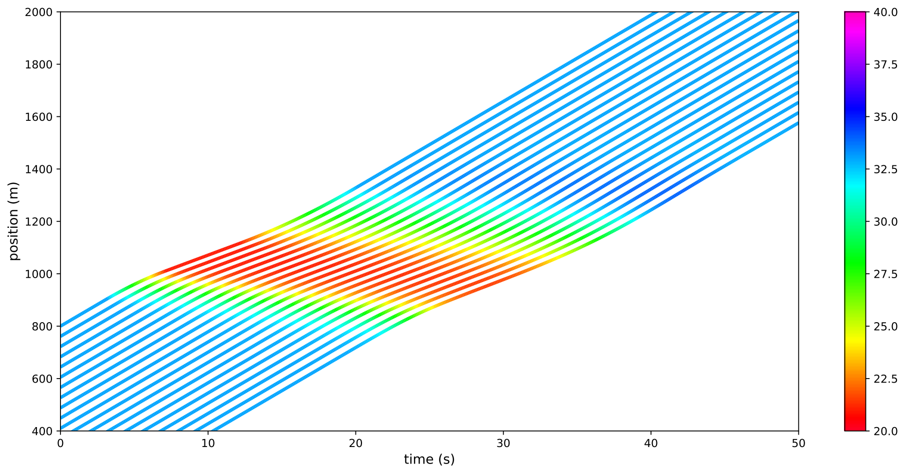


Figure 5.6: Vehicle trajectories and speed contours of the platoon consisting of vehicles equipped with the one-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

The car-following behavior mechanism of the system is analyzed by comparing the trajectories of the following vehicles to the trajectories generated by Newell's car-following model. Figure 5.7 highlights a segment of the vehicle trajectories and plots both the real trajectories and Newell's trajectories. It shows that the two trajectories are nearly overlapped. Figure 5.8 further calculates the position deviation between the real trajectories and Newell's trajectories of the first five following vehicles. As can be seen in the figure, there are only little variations between these two trajectories, which indicates that the vehicles equipped with the ACC system basically follow the longitudinal movement of the leader only with a space shift and time delay. There is no apparent timid or aggressive driving behavior found from the followers.

After the performance assessment of the one-leader ACC system designed with the first weighting combination, the platoon consisting of followers equipped with the system designed with the second weighting combination is then simulated. Figure 5.9 shows the acceleration/speed/gap profiles of the first five followers. It can be seen that the minimum deceleration of the followers becomes larger while propagating upstream, which demonstrates that the disturbance is gradually damped out. In addition, compared to the system trained with the first weighting combination, the gap errors of the followers equipped with the system trained with the second weighting combination are smaller. This clearly indicates the effect of the second weighting combination setup in the reward function which

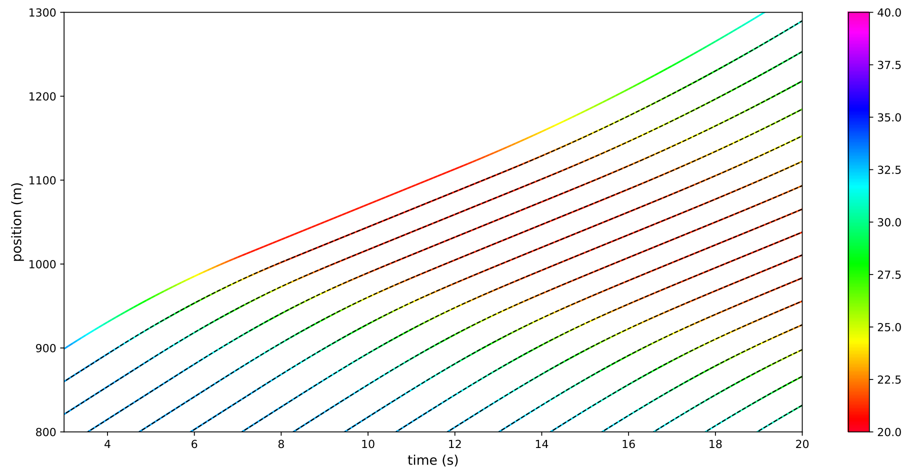


Figure 5.7: Real trajectories (colored lines) and Newell's trajectories (dotted lines) of the platoon using the one-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

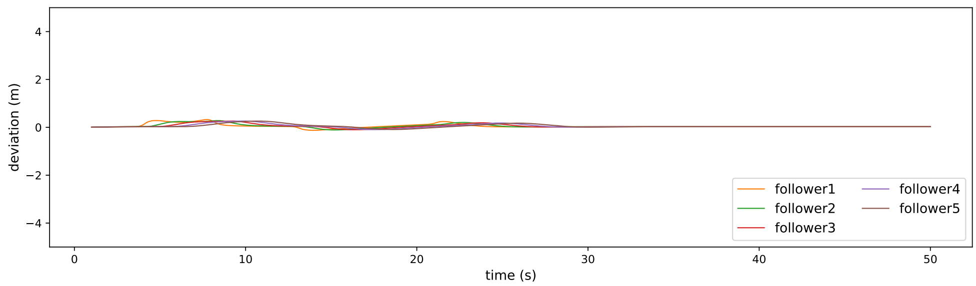


Figure 5.8: Position deviation with Newell's trajectories of the first five followers in the platoon using the one-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

favors the gap keeping performance more than reducing the jerk amplitudes.

The speed contours in Figure 5.10 also show that the minimum speed of the upstream vehicles becomes higher as the dark red segments becomes shorter or even fades out. Moreover, the overshooting behaviors (dark blue segments) can no longer be observed. These features all demonstrate the better string stability performance of the one-leader ACC system designed with the second weighting combination.

The car-following behavior mechanism of the one-leader ACC system designed with the second weighting combination is also analyzed by comparing the trajectories of the following vehicles to the trajectories generated by Newell's car-following model. Figure 5.11 also shows the position deviation between the two trajectories for each follower. It is shown that there are more significant variations between these two trajectories than the platoon using the previous ACC system although the deviations are basically smaller than 1 m. The positive deviations at the deceleration phase indicate that the followers slow down more than the theoretical behavior, meaning that the vehicles equipped with the one-leader ACC sys-

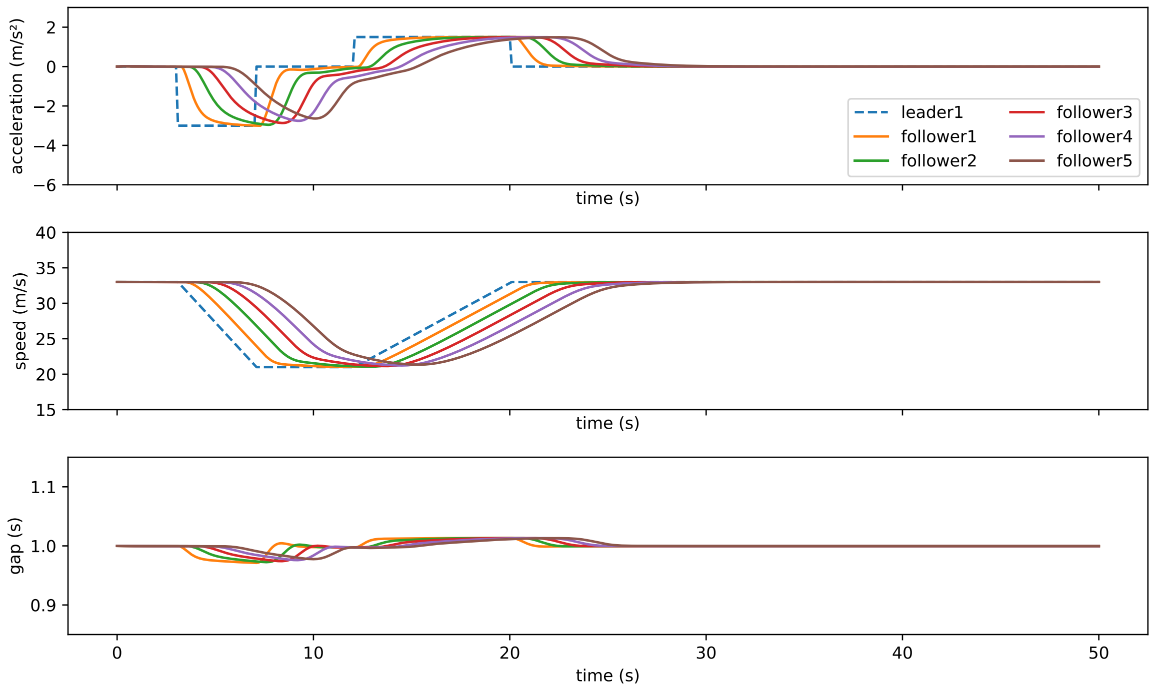


Figure 5.9: Acceleration, speed, and time gap profiles of the platoon using the one-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

tem designed with the second weighting combination tend to decelerate more than those vehicles equipped with the previous system. It can also be observed by looking at the speed contours in Figure 5.10 that the green segments at the deceleration phase start earlier and also become longer. This kind of relatively conservative driving behavior helps mitigate the propagation of the disturbance.

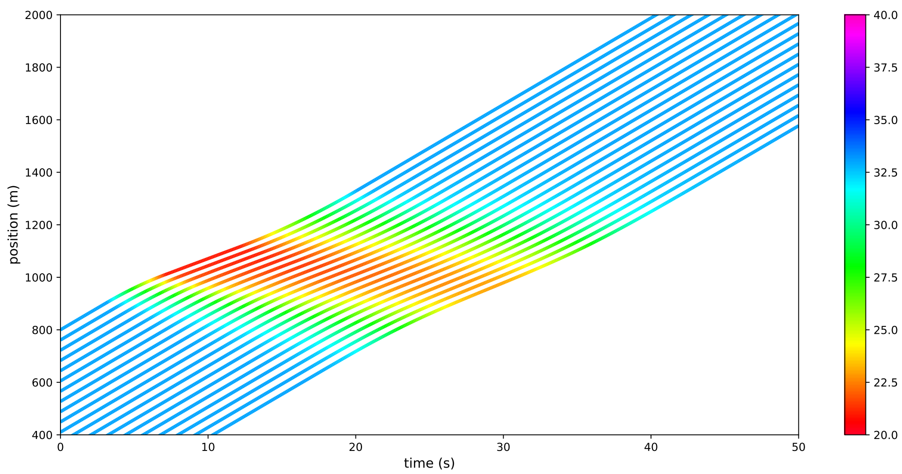


Figure 5.10: Vehicle trajectories and speed contours of the platoon consisting of vehicles equipped with the one-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

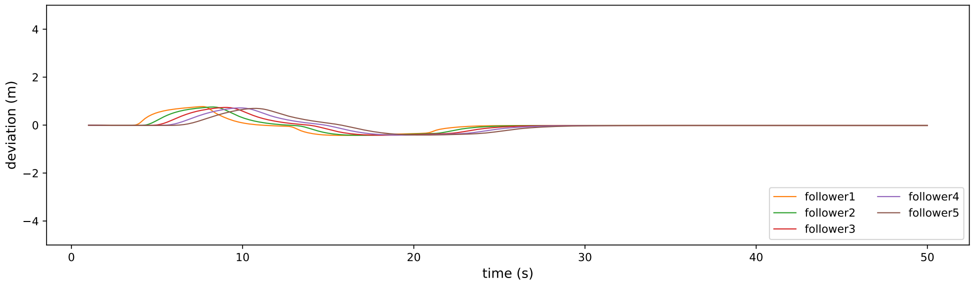


Figure 5.11: Position deviation with Newell's trajectories of the first five followers in the platoon using the one-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

To sum up, the one-leader ACC system designed with the first weighting combination can slightly damp out the disturbance although the effect is still difficult to be observed, which indicates a small level of string stable performance. However, little overshooting behavior exists at the stabilization phase for every following vehicle. On the other hand, the one-leader ACC system designed with the second weighting combination can significantly damp out the disturbance and prevent the overshooting phenomena, showing the better string stability performance.

5.2.2. Performance of two-leader ACC systems

The simulation results of the vehicle platoon using the two-leader ACC system are presented in this subsection. Figure 5.12 provides the first impression on the performance of the two-leader system designed with the first weighting combination. By looking at the acceleration profile, the followers have significantly higher minimum acceleration value in the deceleration phase than their leaders do when facing the disturbance. The same effect can be observed from the slope of the curves in the speed profile. In addition, it seems that the minimum speed of each follower becomes higher in the upstream of the platoon although it cannot be clearly shown by only looking at the first five followers.

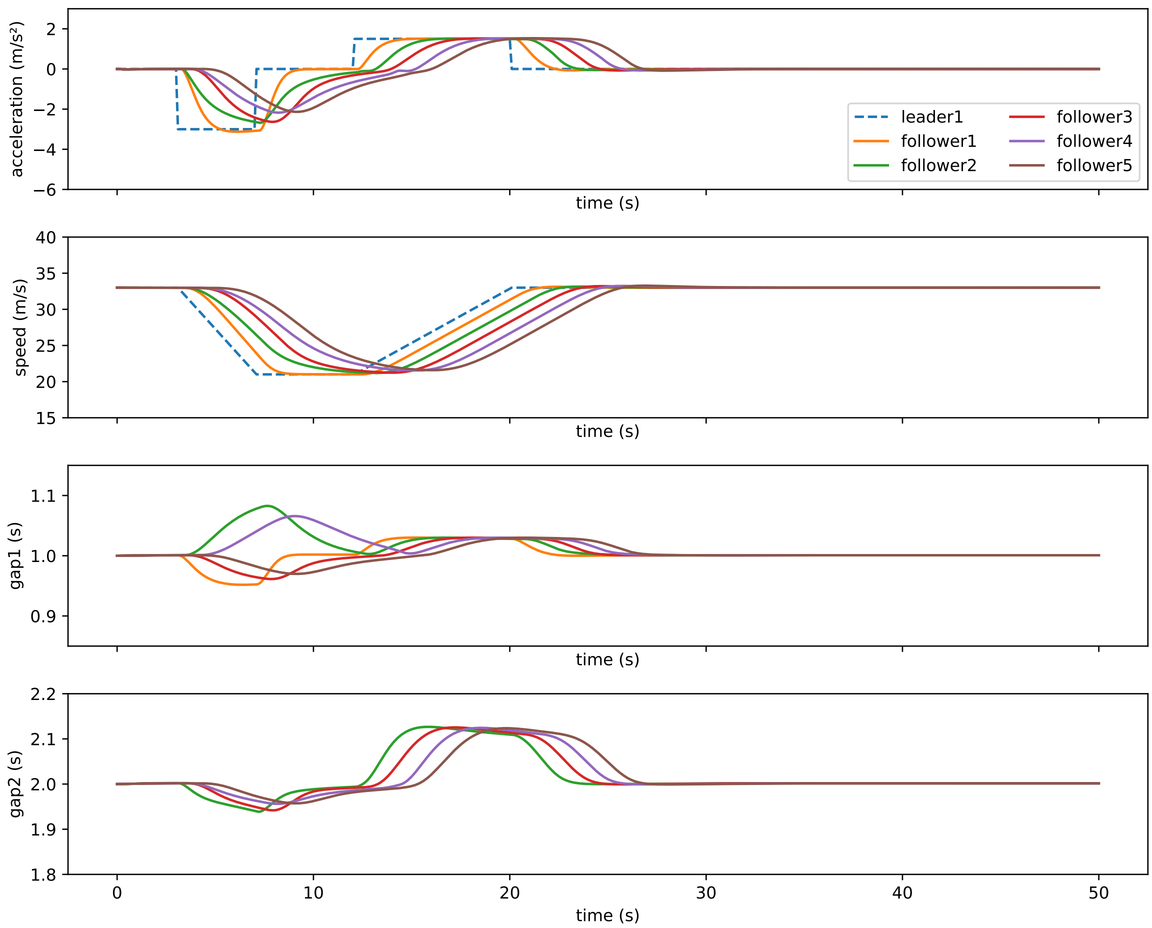


Figure 5.12: Acceleration, speed, and time gap profiles of the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

Figure 5.13 helps observe the benefit of the two-leader system. The minimum speed of the following vehicles becomes significantly higher (shorter red segments) while the disturbance propagates in the upstream direction. However, the length of yellow and green segments also becomes longer for followers at the upstream of the platoon, which implies that the two-leader ACC systems guide the vehicles to start slowing down earlier than the one-leader ACC systems do. This demonstrates the early slow down reaction resulted from the multi-anticipation ability. Vehicles tend to decelerate lightly for a longer time span instead of decelerating strongly for a shorter time period.

According to these two figures, the other aspect worth mentioning is that the overshooting behavior is not as significant as it is in the case for the one-leader system. This is believed to be the effect of looking at more than one leader downstream. In the acceleration and stabilization phases, although the behavior of the followers may be mostly influenced by the first leader, it may also be controlled by the second leader at some points in time. The aggressive driving maneuver can hence be suppressed, which mitigates the overshooting effect.

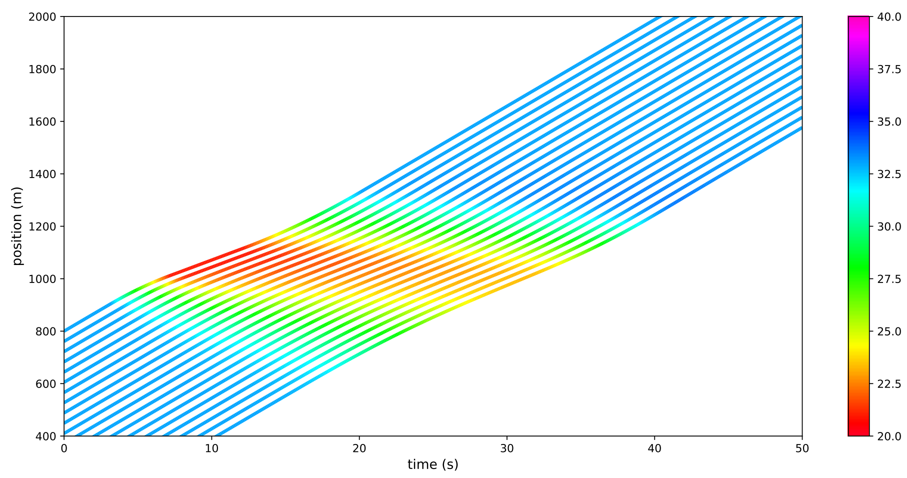


Figure 5.13: Vehicle trajectories and speed contours of the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

To explore the car-following dynamics of the two-leader system, the plots of time gaps 1 and 2 in Figure 5.12 first provide some insights. It is found that the time gaps of follower 2 and follower 4 with their first leader are larger than 1 s in the deceleration phase. On the other hand, their time gaps with the second leader (time gap 2) are below the desired value (2 s), which clearly implies that they are mostly reacting according to the movement of the second leader during deceleration. This is also the effect of the aforementioned early slow down behaviors. It is believed that this phenomenon can be significantly observed from the behaviors of followers 2, 4, and 6, etc., while the behavior of followers 3, 5, and 7, etc. are mainly influenced by their first leader since their first leaders (followers 2, 4, and 6, etc.) behave more conservative and slow down earlier than their second leaders (followers 1, 3, and 5, etc.). Therefore, the multi-anticipative car-following behaviors tend to occur intermittently within the platoon. Figure 5.14 plots the controlling leader at every time step of followers 2-5. As can be seen, followers 2 and 4 are controlled by the second leader more often during the deceleration phase, which demonstrates the mentioned phenomena. In addition, this figure also helps explain that the second leader does take over the control of the ACC task again between the acceleration phase and stabilization phase to suppress the overshooting behaviors of the followers.

Figures 5.15 and 5.16 can also be used to observe the propagation of this kind of car-following behavior. In Figure 5.15, it is found that the real trajectories of several followers, e.g., the third vehicle (follower 2), are slower than Newell's trajectories right before they reach the minimum speed. This early slow down behavior is similar to the conservative (or

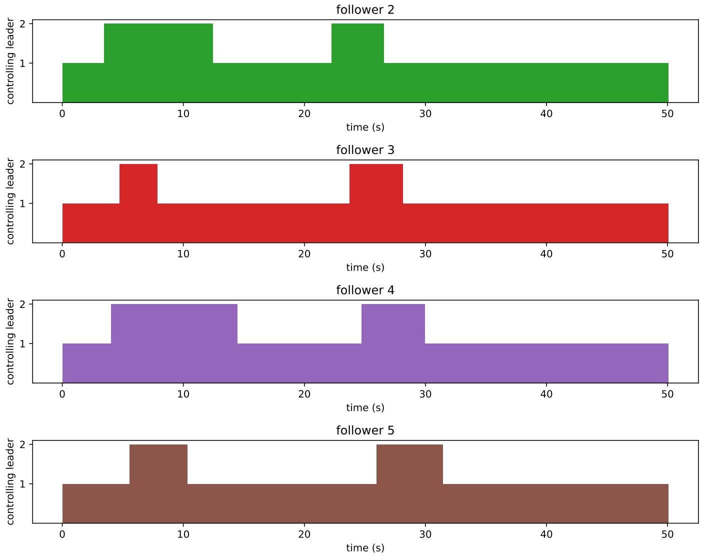


Figure 5.14: Controlling leader of followers 2-5 equipped with the two-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

timid) driving behavior of human drivers introduced in Laval and Leclercq (2010). Such phenomenon become less observable from the trajectories of vehicles upstream. Figure 5.16 clearly demonstrates such effect. The positions of followers 2 and 4 deviate from the Newell's trajectories at the beginning of the disturbance, while the deviation of follower 4 is also smaller than that of follower 2.

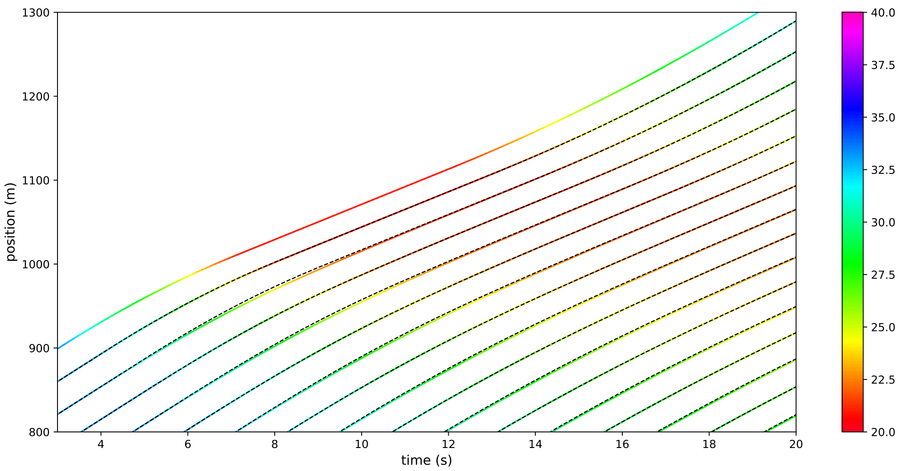


Figure 5.15: Real trajectories (colored lines) and Newell's trajectories (dotted lines) of the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

The performance of the two-leader ACC system designed with the second weighting combination is then analyzed. Figure 5.17 shows the acceleration/speed/gap profiles of the first five followers. Compared to the system designed with the first weighting combination, the behaviors of the followers are all quite similar. Followers 2 and 4 again exhibit the early

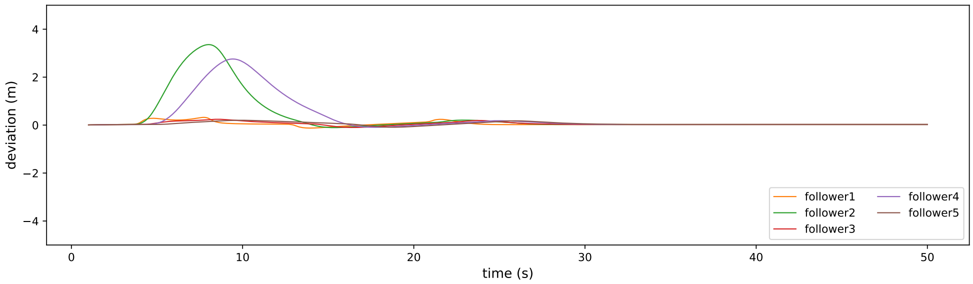


Figure 5.16: Position deviation with Newell's trajectories of the first five followers in the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with accurate measurements

slow down behavior as plotted in the profile of gap 1. The difference brought by the second weighting setup can be observed from the gap profiles that the gap errors become even smaller than before. This again demonstrates the effect of giving higher weights on the gap error term in the reward function.

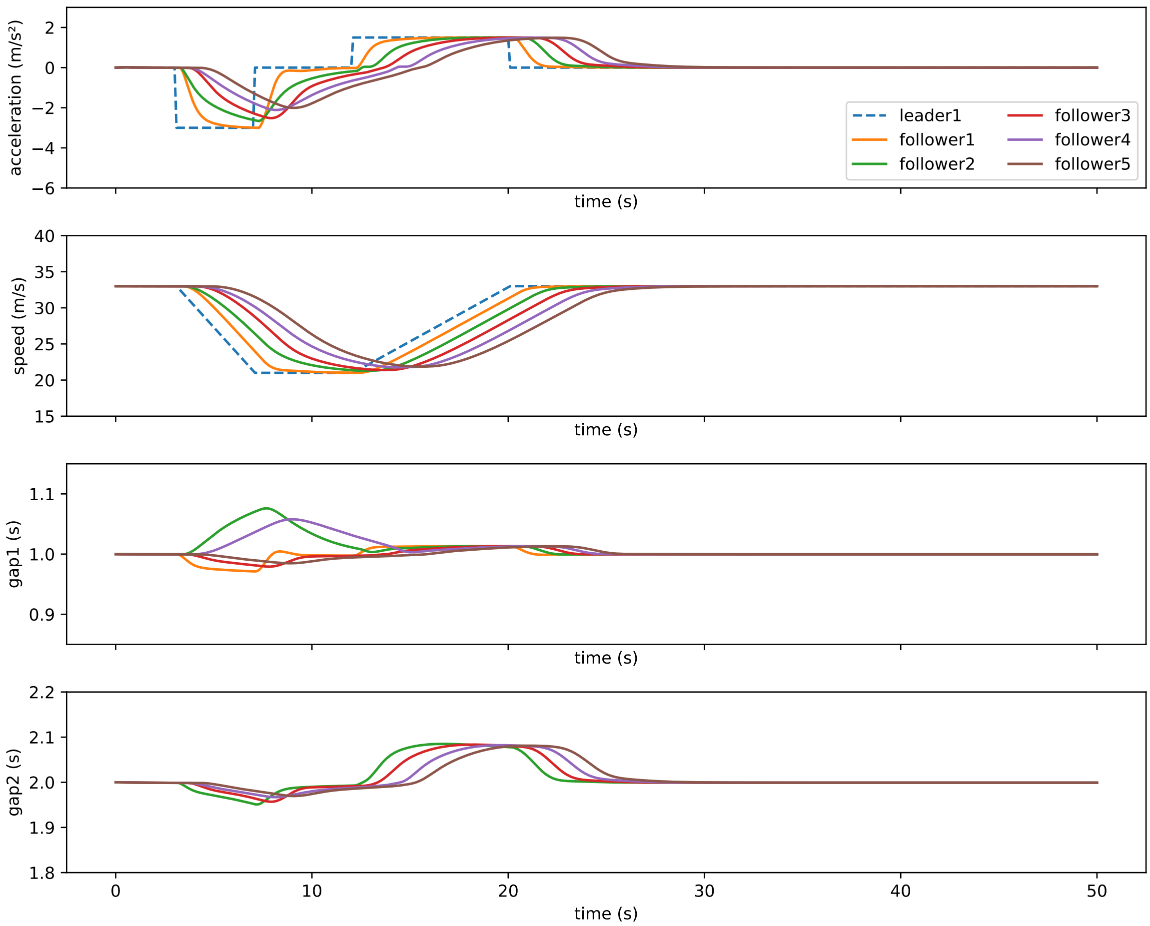


Figure 5.17: Acceleration, speed, and time gap profiles of the platoon using the two-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

More significant difference between the two weighting combinations can be found from the vehicle trajectories and speed contours in Figure 5.18. As can be observed from the propagation of the disturbance shockwave, the lowest speed of the very upstream vehicle stays above 24 m/s, which is higher than the platoon using the system designed with the first weighting combination. In addition, the overshooting behaviors between the acceleration and stabilization phase can barely be seen in the speed contours.

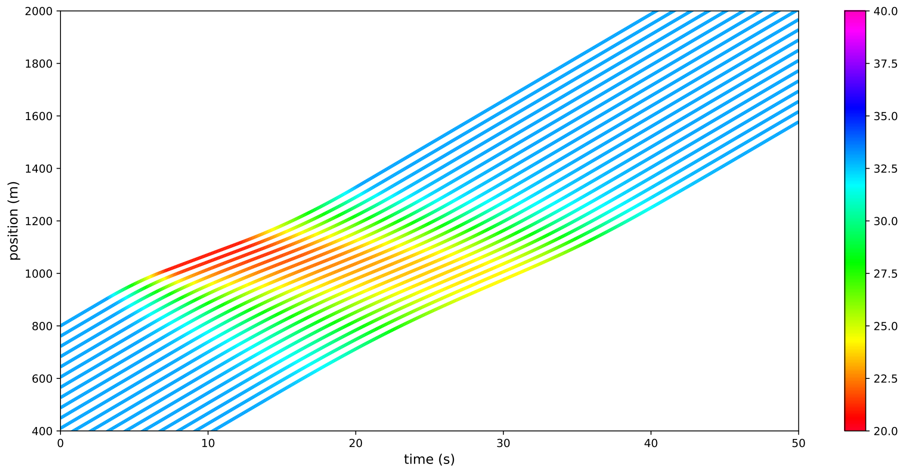


Figure 5.18: Vehicle trajectories and speed contours of the platoon using the two-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

Figure 5.19 also shows that not only followers 2 and 4 decelerate more than the theoretical behavior by looking further downstream, but other following vehicles also have relatively conservative car-following behavior. These behaviors collectively damp out the disturbance faster than the behaviors of the vehicles equipped with the previous system.

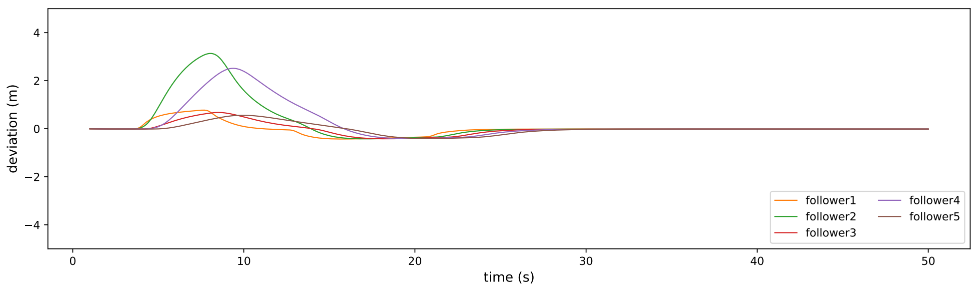


Figure 5.19: Position deviation with Newell's trajectories of the first five followers in the platoon using the two-leader ACC system designed with the second weighting combination in the scenario with accurate measurements

In conclusion for the two-leader ACC system, the disturbance can be effectively damped out when the following vehicles look further ahead. Hence, it has better string stability performance than the one-leader system. The overshooting phenomena between the acceleration and stabilization phases of every following vehicle are also mitigated. These are all benefits brought by the multi-anticipative car-following behavior, which can be found by looking at the acceleration/speed/gap profiles and position deviations between the real and theoretical trajectories.

5.2.3. Comparison of one-leader and two-leader ACC systems

To compare the one-leader and two-leader systems, the string stability indicators, including oscillation growth amplitudes *OG* and overshooting amplitudes *OS* of every vehicle in the platoon, and the ride comfort performance are analyzed from the simulation result.

The systems designed with the first weighting combination is first evaluated and compared. Figure 5.20 shows the two indicators, *OG* and *OS*, of each follower in the platoon when facing the disturbance. By looking at the oscillation growth curve of the one-leader system, it is found that the one-leader system can really achieve string stable performance since the *OG* value decreases as the disturbance propagates to upstream vehicles. The two-leader system, on the other hand, has a more significant reduction in the *OG* value than the one-leader system. The curve of the two-leader system also exhibits a step-like pattern. This feature is resulted from the dynamics of the multi-anticipative behavior. As discussed in subsection 5.2.2, the significant oscillation amplitude reductions and minimum deceleration increment would occur at followers 2, 4, and 6, etc.

The overshooting amplitude *OS* of each vehicle in the platoon using the two systems are shown on the right side in Figure 5.20. Both systems have increasing *OS* values along the platoon, which means the overshooting exists when using both systems. However, the slope of the *OS* curve is smaller for the two-leader system, indicating a mitigated overshooting effect. The result coincide with the finding from Figures 5.12 and 5.13 in subsection 5.2.2.

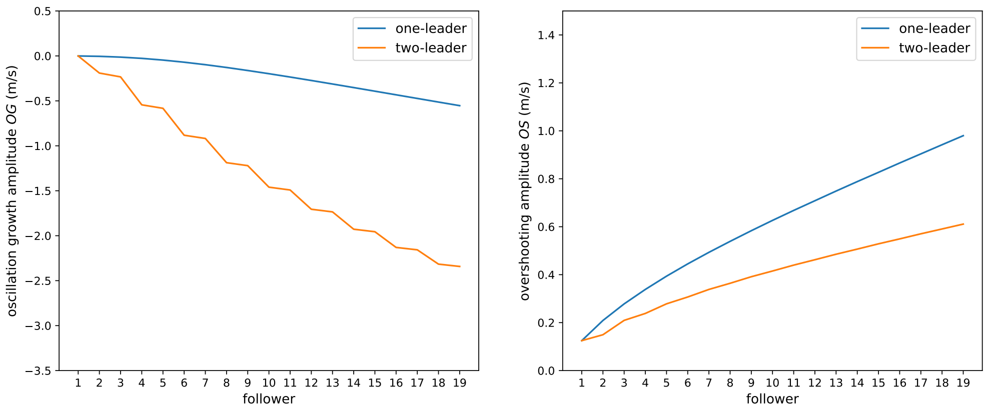


Figure 5.20: Comparison of the oscillation growth amplitude (left) and overshooting amplitude (right) between the one-leader and two-leader ACC systems designed with the first weighting combination in the scenario with accurate measurements

Figure 5.21 shows the two string stability indicators of the systems designed with the second weighting combination. As can be seen in the left figure, the oscillation growth amplitudes decrease faster in the curves of both the one-leader and two-leader systems.

This demonstrated that the proposed one-leader ACC system designed with the second weighting combination possess the ability to damp out the disturbance. The *OG* values of the vehicles equipped with the two-leader system also shows the step-like decreasing pattern while propagating upstream. Follower 19 in the platoon can reduce nearly 3.5 m/s of the oscillation.

On the other hand, the effect of the second weighting combination can also be discovered by looking at the overshooting amplitude, as shown in the right plot of Figure 5.21. The overshooting amplitudes are all smaller than or equal to 0.015 m/s for vehicles equipped with either the one-leader or two-leader systems, which are small enough to be negligible. This again indicates that the overshooting effect does not exist when using the systems designed with the second weighting combination.

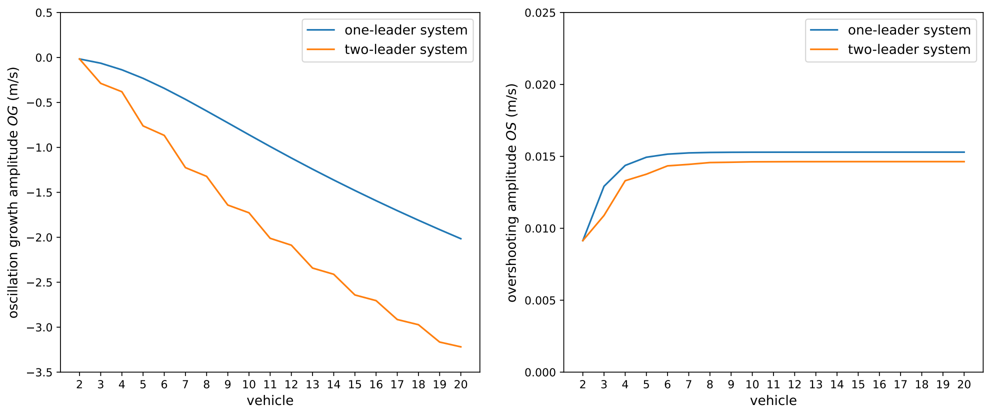


Figure 5.21: Comparison of the oscillation growth amplitude (left) and overshooting amplitude (right) between the one-leader and two-leader ACC systems designed with the second weighting combination in the scenario with accurate measurements

The ride comfort performance of the two systems designed with both weighting combinations can be compared by analyzing the jerk experienced by the following vehicles according to the method proposed in subsection 4.4.2. Figure 5.22 shows the distribution of jerk of all vehicles at every time step in the traffic disturbance. Although the difference between all the curves may not be significant, some features can still be observed. It is first found that the two-leader systems outperform the one-leader systems in both weighting combination setups. The jerk amplitudes distributions of the two-leader systems are more concentrated at around 0 m/s^3 than those of the one-leader systems, indicating that the early slow down behavior brought by the two-leader systems can slightly reduce the amount of large jerk amplitudes.

However, comparing the performance of two different weighting combination setups, it is found that the systems designed with the second weighting combination have better ride comfort performance than the systems designed with the first weighting combination as the green and red curves in Figure 5.22 are more concentrated than the blue and orange curves. This may be different from the initial hypothesis since the second weighting combination does not favor the limitation on jerk amplitudes. The possible reason could be that the upstream vehicles equipped with the systems designed with the second weighting combination in the platoon experience less disturbance since the string stability performance is also improved. The vehicles would not need to conduct relatively aggressive driving maneuvers any more when facing the disturbance. Hence, the aggregated ride comfort performance of the platoon is also improved.

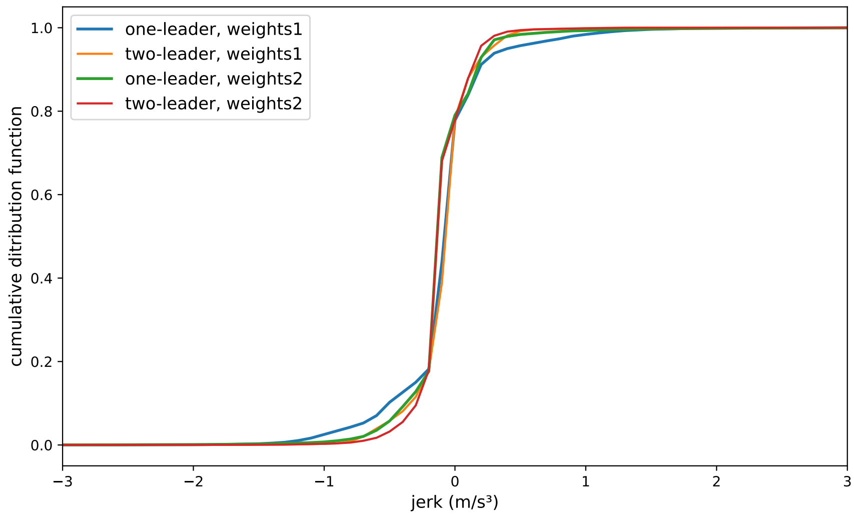


Figure 5.22: Cumulative distribution function of jerks experienced by the platoons using the one-leader and two-leader ACC systems in the scenario with accurate measurements

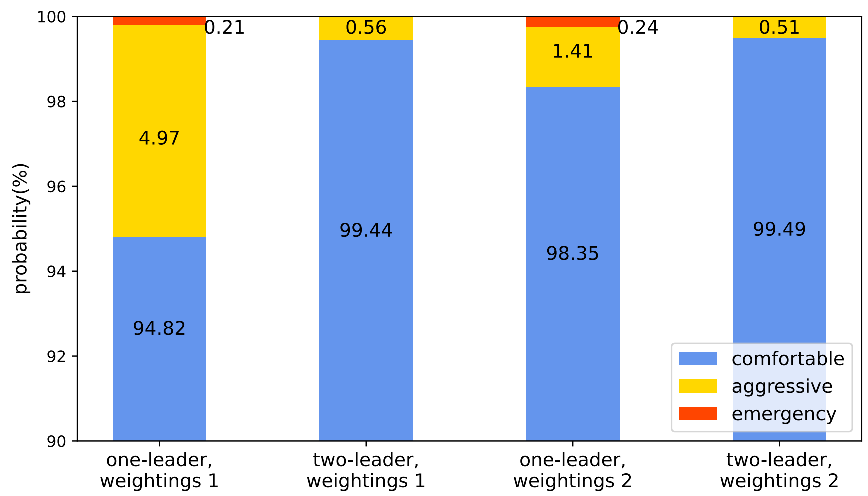


Figure 5.23: Probability distribution of jerks in each ride comfort level experienced by the platoons using the one-leader and two-leader ACC systems in the scenario with accurate measurements

Figure 5.23 shows the distribution of jerks in each level defined in subsection 4.4.2. For the systems designed with the first weighting combination, significant improvement on ride comfort can be found. When using the one-leader system, around 5% of the maneuvers are aggressive driving behaviors, while a small amount of emergency maneuvers can be found. When using the two-leader system, almost 100% of the jerks experienced by following vehicles stay in the range of comfortable driving maneuver. This implies that the multi-anticipation behavior helps the following vehicles to prevent aggressive accelerating and decelerating maneuver. When looking at the systems designed with the second weighting combination, the one-leader system can already reduce the amount of aggressive driving maneuvers significantly compared to the system designed with the first weighting combination. The two-leader system does not show much improvement when using these criteria to assess the ride comfort since they are already at a quite ideal level of performance. Still, the effect of multi-anticipation and different weighting setups on ride comfort has been found by investigating these four systems.

5.3. System performance under measurement uncertainty

In this section, both systems are simulated with different levels of sensor measurement uncertainties to explore the performance of the systems in scenarios which are closer to the real-world autonomous driving situations. It is also important to understand the influence of erroneous second leader information so that countermeasures can be designed and used to maintain the string stability.

5.3.1. Performance of one-leader ACC systems under measurement uncertainty

The performance of the one-leader ACC system when facing measurement noise is first discussed. As introduced in chapter 4, the first leader measurements have a small level of uncertainty with standard deviations $\sigma_{g_1} = 0.2$ m and $\sigma_{v_1} = 0.2$ m/s (measurement uncertainty level N0). 20 simulation runs were executed to account for the randomness.

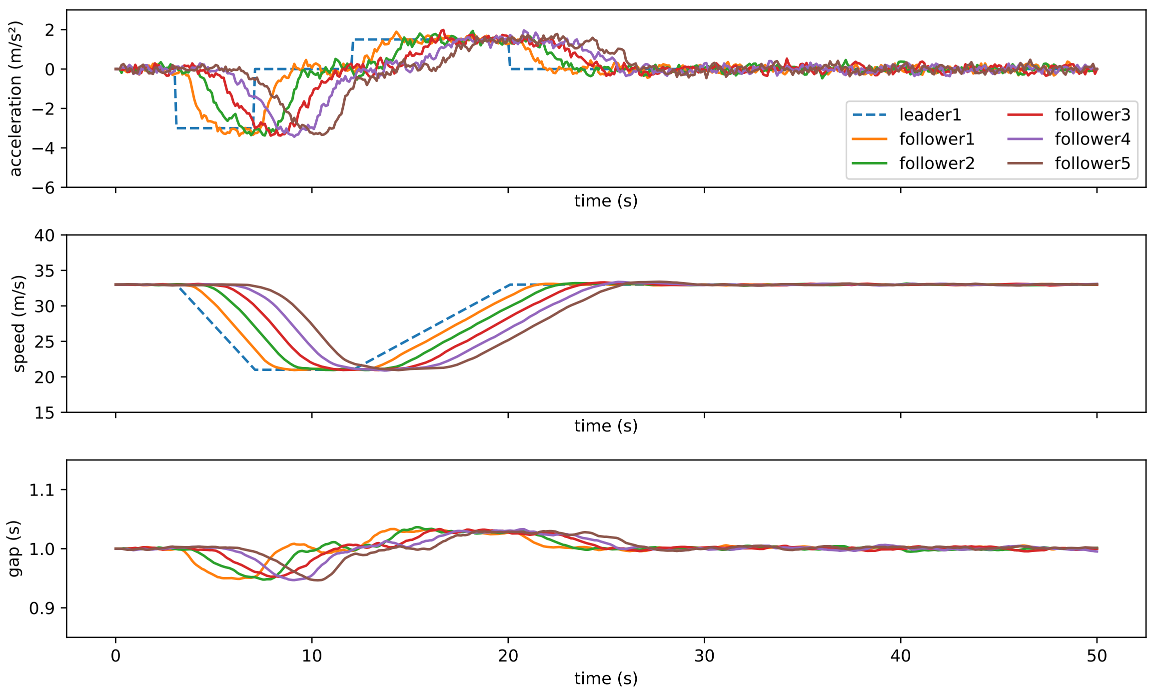


Figure 5.24: An example of acceleration, speed, and time gap profiles of the platoon using the one-leader ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

Figure 5.24 shows an example of the acceleration/speed/gap profiles of the one-leader ACC system with the specified level of sensor measurement uncertainty. It is found that the curves in the profiles fluctuate more often than they do in the case with perfect information. However, it is not observable that whether the string stability performance is deteriorated by simply looking at the profiles of the first five followers. The acceleration/speed/gap profiles of the one-leader ACC system designed with the second weighting combination

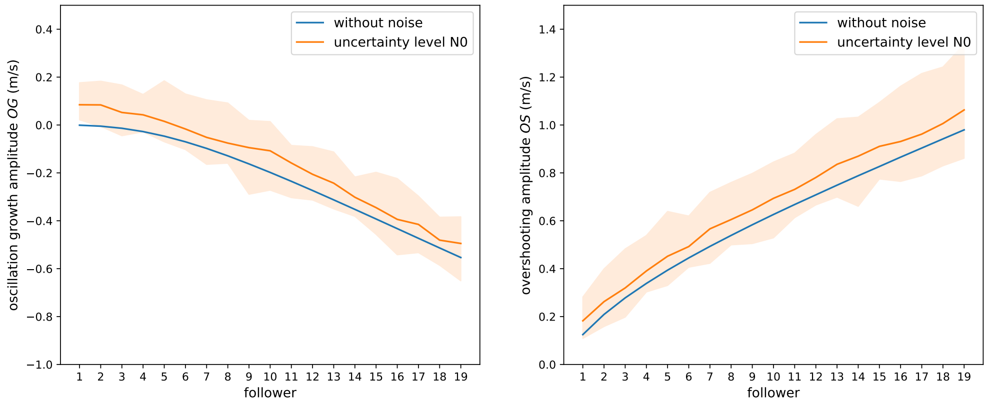


Figure 5.25: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader ACC system designed with the first weighting combination in scenarios with and without measurement noise

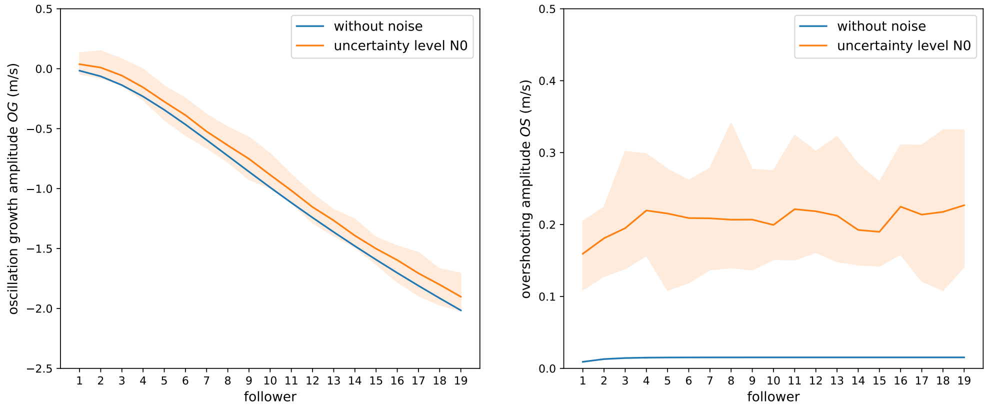


Figure 5.26: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader ACC system designed with the second weighting combination in scenarios with and without measurement noise

when facing the same level of measurement noise shows similar pattern and is therefore not shown in this subsection.

Figures 5.25 and 5.26 provide a quantitative way to look at the influence of the measurement noise on string stability indicators. As can be seen on the left side of the figures, the *OG* values of every vehicle in the platoon become higher when the measurements are noisy. In addition, the *OS* values of every vehicle in the platoon *OS* increases compared to the scenario with perfect information. Therefore, it is found that the noisy measurements bring negative effect on the string stability performance for the one-leader ACC systems, but the impact is not significant enough to cause string instability.

Figures 5.27 and 5.28 show the cumulative distribution of jerk amplitudes of the one-leader ACC systems before and after the measurement noise is considered in the simulation experiment. As can be observed in the distribution curves, there are more aggressive and emergency driving maneuvers present when the measurement noise is considered, indicating the influence of measurement noise on the ride comfort performance. By comparing the systems designed with different weighting setups, it can also be found that the system designed with the second weighting combination produce more driving maneuvers with large jerk amplitudes, which is because of the effect of the different weighting setups in the training process. The second weighting combination resulted in a trained policy which does not limit the jerk amplitudes as much as the one trained with the first weighting combination. Hence, when the measurements are noisy, its actions would also fluctuate more, which leads to larger jerks.

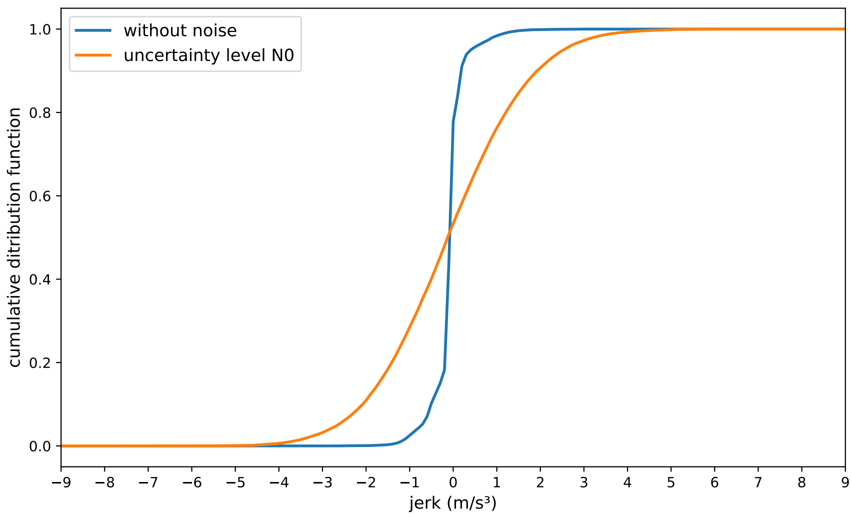


Figure 5.27: Cumulative distribution function of jerks experienced by the platoon using the one-leader ACC system designed with the first weighting combination in scenarios with and without measurement noise

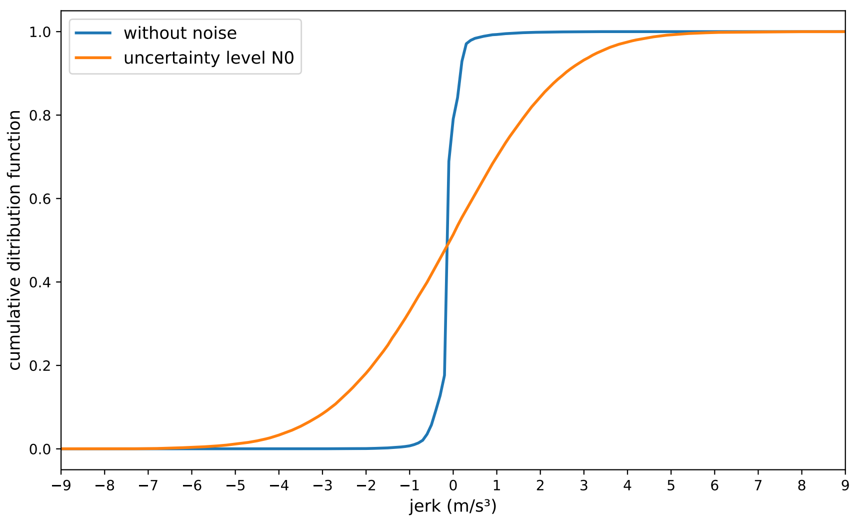


Figure 5.28: Cumulative distribution function of jerks experienced by the platoon using the one-leader ACC system designed with the second weighting combination in scenarios with and without measurement noise

The probability distribution of jerks in each ride comfort level can be found in Figure 5.29. As can be seen in the figure, the number of jerk amplitudes which are larger than 0.9 m/s^3 significantly increases when facing noisy measurements. In addition, a lot of emergency driving maneuvers with jerk amplitudes larger than 2 m/s^3 appear. This outcome can be caused by both the incorrect measurements and the abrupt actions required to reach the desired state. The ride comfort performance difference between the two weighting combination setups can also be found.

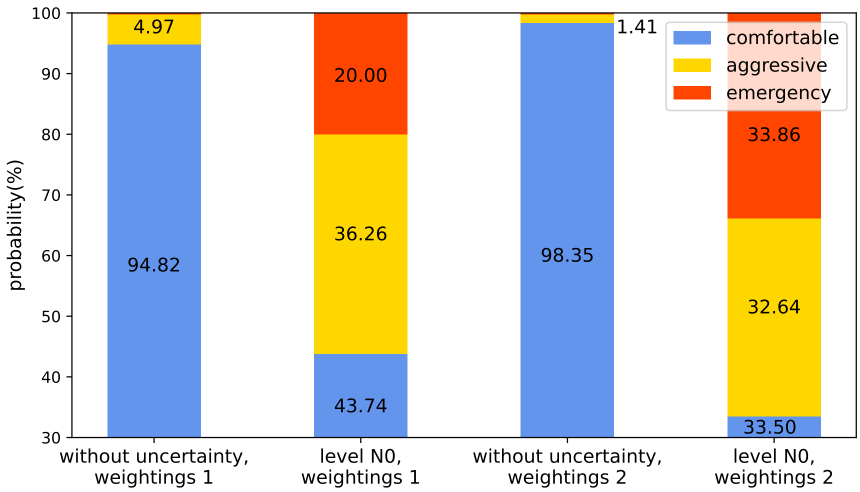


Figure 5.29: Probability distribution of jerks in each ride comfort level experienced by the platoons using the one-leader ACC systems in scenarios with and without measurement noise

5.3.2. Performance of two-leader ACC systems under measurement uncertainty

To evaluate the performance of the two-leader ACC system considering measurement uncertainties, it is simulated with four levels of measurement noise, as mentioned in subsection 4.2.2. Figures 5.30 and 5.31 show the acceleration, speed, and time gap profiles of the two-leader ACC system designed with the first weighting combination considering the measurement uncertainty levels N1 and N4 in the second leader measurements, respectively. In Figure 5.30, the pattern of every element is quite identical to the pattern in Figure 5.12 when there is no uncertainty. Although the acceleration decision (control action) made by the following vehicles is slightly fluctuating, the influence is not significant in the speed profile.

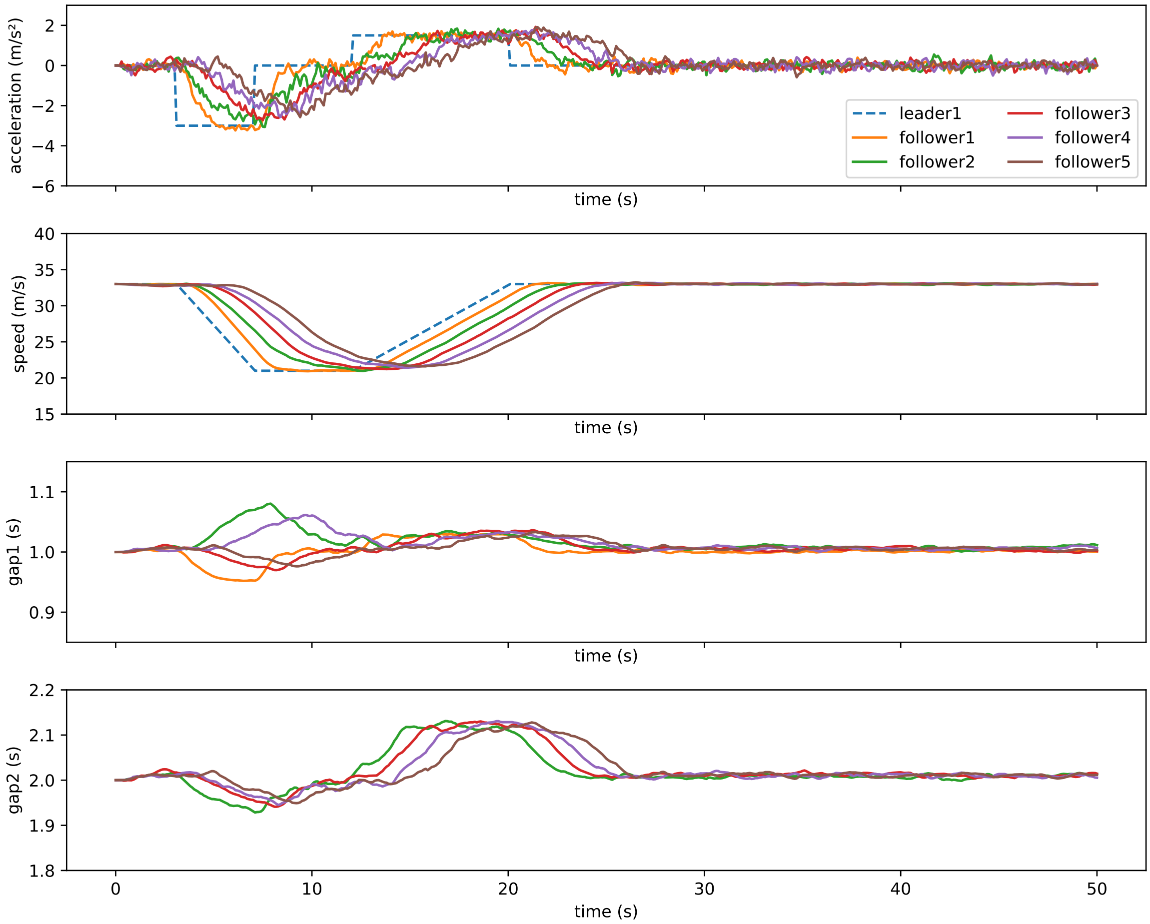


Figure 5.30: Acceleration, speed, and time gap profiles of the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N1

In Figure 5.31, however, the fluctuation of vehicle acceleration becomes more irregular than in the case of low measurement noise. This also makes the speed profile of each following vehicle more irregular. In addition, by looking at the time gap with the first leader (time gap 1), it is found that the early slow down behavior can hardly be observed. This means that the system does not utilize the full potential of the multi-anticipative car-following behavior because of the incorrect second leader measurements. The same consequence can also be found in the profile of time gap 2 with the second leader. As shown in Figure 5.30, in the scenario with relatively small measurement errors, time gaps 2 would not deviate too much away from the desired level (2 s) than in the case of perfect information. When the errors become larger, it is observed that time gap 2 deviates more than 0.1 s in both the deceleration and acceleration phases, which shows degraded string stability performance. Still, one can see that the oscillation is not amplified, the minimum deceleration rate does not decrease dramatically as well. It is therefore inferred that the string stability performance can still be ensured even if there is a certain level of measurement noise considered.

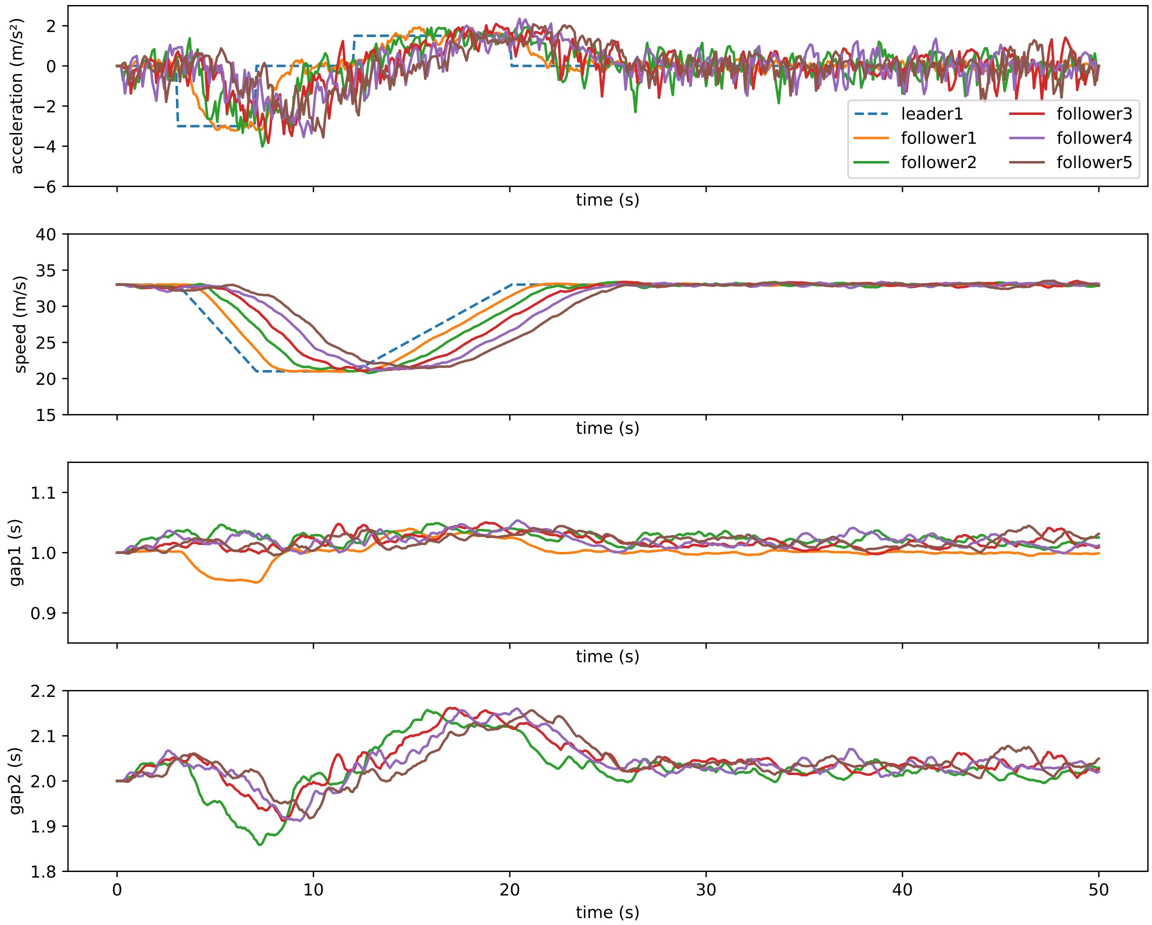


Figure 5.31: Acceleration, speed, and time gap profiles of the platoon using the two-leader ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N4

Once again, the acceleration/speed/gap profiles of the two-leader ACC system designed with the second weighting combination are skipped as they have similar patterns and would not provide added value for the discussion in this subsection. The quantitative analysis results of the system can still be found in the remainder of this subsection.

Figures 5.30 and 5.31 provide the first impression and comparison of the effect of different levels of measurement noise. The system performance is discussed by taking average over 20 simulation runs and analyzing the string stability indicators, as shown in Figures

5.32 and 5.33. For the two-leader ACC system, the indicator values would start from follower 2 since follower 1 only uses the one-leader system. It is shown that while the measurements become more noisy, the *OG* values becomes larger. However, the overall trend of the curves still remains, implying that the string stability can be maintained even if the ego-vehicle receives noisy measurements. The *OS* values basically show the same phenomena. The higher the uncertainty level is, the larger the overshooting amplitudes are. On the other hand, the *OS* curves of the system designed with the first weighting combination show slightly different outcomes. It can be seen on the right side of the Figure 5.32 that including certain levels of measurement noise helps mitigate the overshooting effect. It seems that the noisy measurements may lead to relatively conservative behavior between the acceleration and stabilization phase since the following vehicle would slow down more often due to the erroneous information, which eliminates the overshooting behaviors.

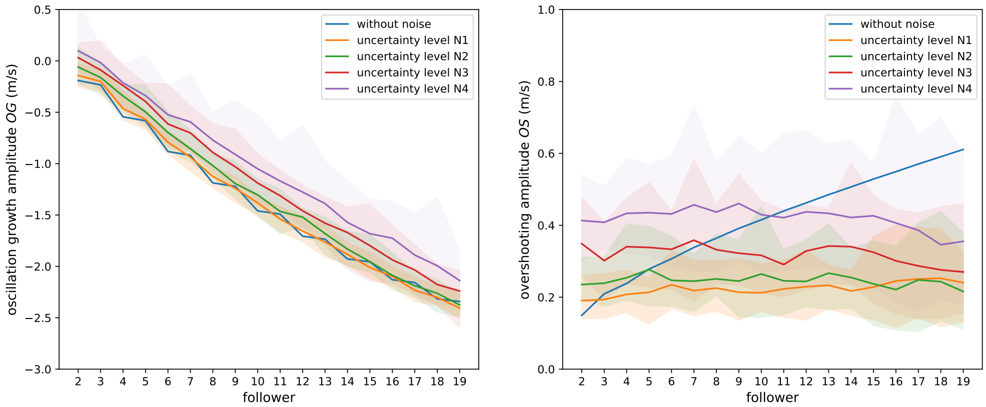


Figure 5.32: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader ACC system designed with the first weighting combination in scenarios with measurement uncertainty levels N1 - N4

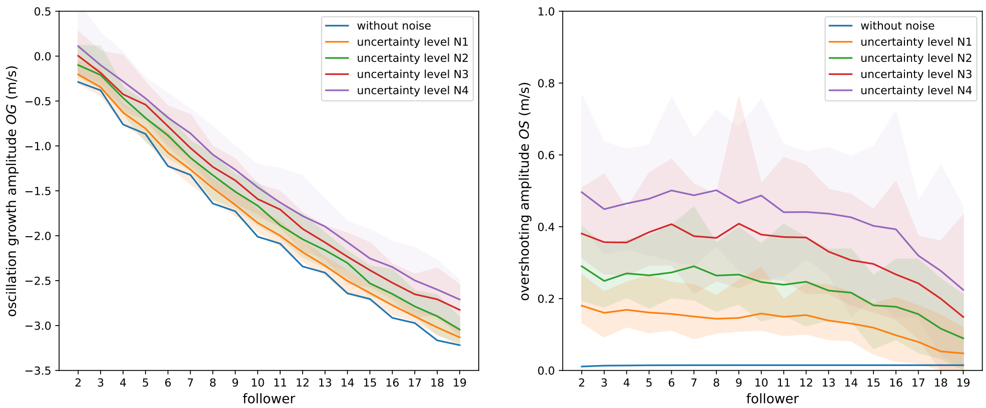


Figure 5.33: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader ACC system designed with the second weighting combination in scenarios with measurement uncertainty levels N1 - N4

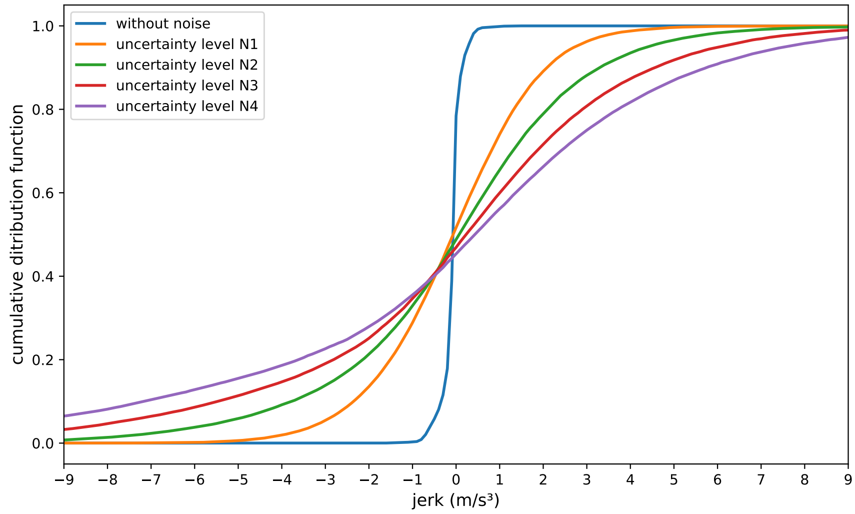


Figure 5.34: Cumulative distribution function of jerks experienced by the platoon using the two-leader ACC system designed with the first weighting combination in scenarios with measurement uncertainty levels N1 - N4

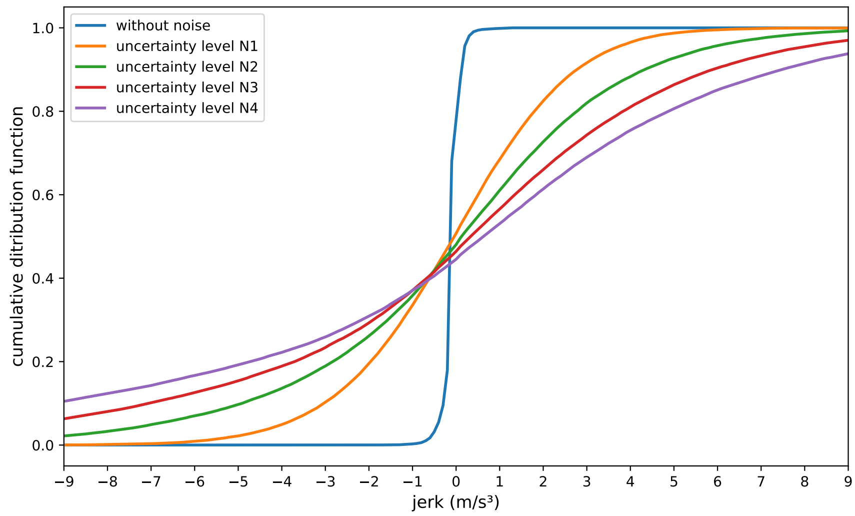


Figure 5.35: Cumulative distribution function of jerks experienced by the platoon using the two-leader ACC system designed with the second weighting combination in scenarios with measurement uncertainty levels N1 - N4

As already shown in Figures 5.30 and 5.31, the existence of measurement noise leads to more jerk throughout the whole simulation run. Figures 5.34 and 5.34 show the distribution of the jerk at every time step experienced by all following vehicles under each level of measurement uncertainty. Figures 5.36 and 5.37 plot the distributions in each ride comfort level. Note that follower 1 is not included in the analysis of the two-leader system since it only uses the one-leader ACC system. As shown by the curves and bar charts, more jerk amplitudes exceeding the range of the comfortable level appear while the measurement uncertainty level increases. At every level of measurement uncertainty, more than 50% of the jerk values exceed the range of comfortable driving maneuver. A large amount of jerk fall into the range of emergency driving behavior, while the situation gets even worse when the level of measurement uncertainty is higher. More than 50% of the jerk amplitudes are larger than 2 m/s^3 in measurement uncertainty level N3. The degraded ride comfort is even more severe for the system designed with the second weighting combination. The numbers of emergency driving maneuvers increase by around 10% in every measurement uncertainty level. This negative impact is expected to be removed after applying the properly designed state estimation methods.

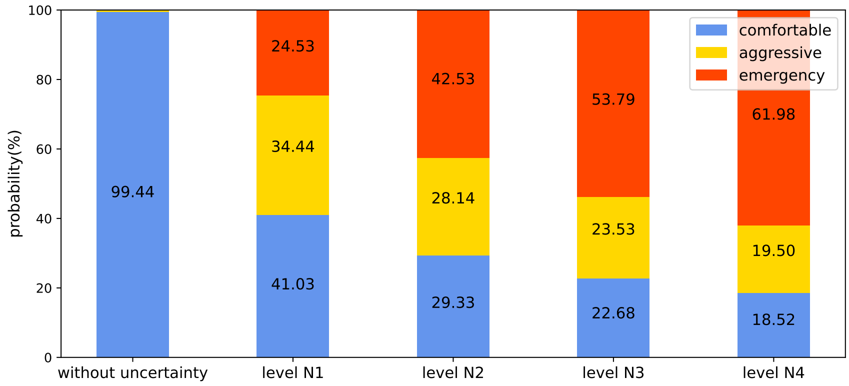


Figure 5.36: Probability distribution of jerks in each ride comfort level experienced by the platoons using the two-leader ACC systems designed with the first weighting combination in scenarios with measurement uncertainty levels N1 - N4

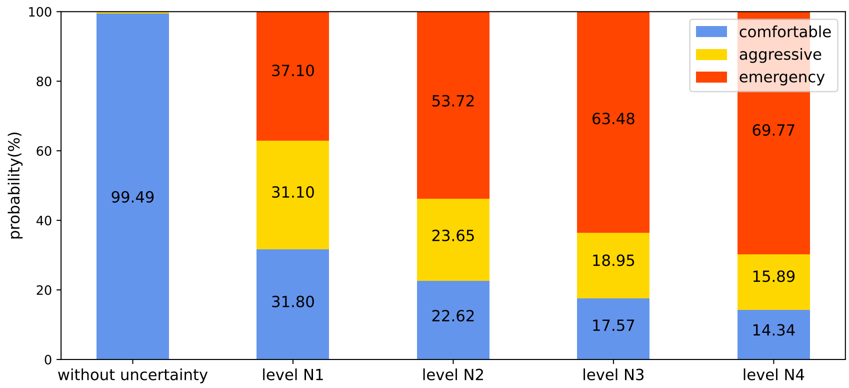


Figure 5.37: Probability distribution of jerks in each ride comfort level experienced by the platoons using the two-leader ACC systems designed with the second weighting combination in scenarios with measurement uncertainty levels N1 - N4

In general, the tested levels of measurement noise would not severely degrade the string instability performance in the deceleration phase given that the measurement of the first leader is still accurate enough, as demonstrated in the previous subsection. Possible reason for such outcome can also be that the random fluctuating effect (the switching between positive and negative measurement errors) can still averagely result in the desired car-following behavior. However, the impact of the noise can still be seen by looking at the ride comfort performance. The actions executed by the followers easily exceed the jerk threshold, indicating degraded and unreasonable driving maneuvers.

5.4. Performance of systems using Kalman filters

Many tracking methods can be adopted to estimate the true state of the leading vehicle and eliminate the influence of measurement noise for the ACC system. This section investigates the effect of Kalman filtering on the performance of the systems. It is applied based on the assumption made in this study that the measurement noise follows Gaussian distribution.

5.4.1. Performance of one-leader KF-ACC systems

The KF state estimator is first implemented to the one-leader system to create a one-leader KF-ACC system. Figure 5.38 presents an example of the filtered output of distance gap and relative speed measurements utilized by the ACC controller of follower 1. As can be seen, the tuned KF can adapt to the disturbance in the leader behavior and reduce the amplitudes of the measurement errors. Figure 5.39 then shows an example of the acceleration/speed/gap profiles of the one-leader KF-ACC system designed with the first weighting combination in the created scenario with measurement uncertainty level N0. By comparing it with Figure 5.24, one can clearly see that the vehicle acceleration does not fluctuate that often any more, demonstrating the positive effect of filtering for the range and range rate measurements.

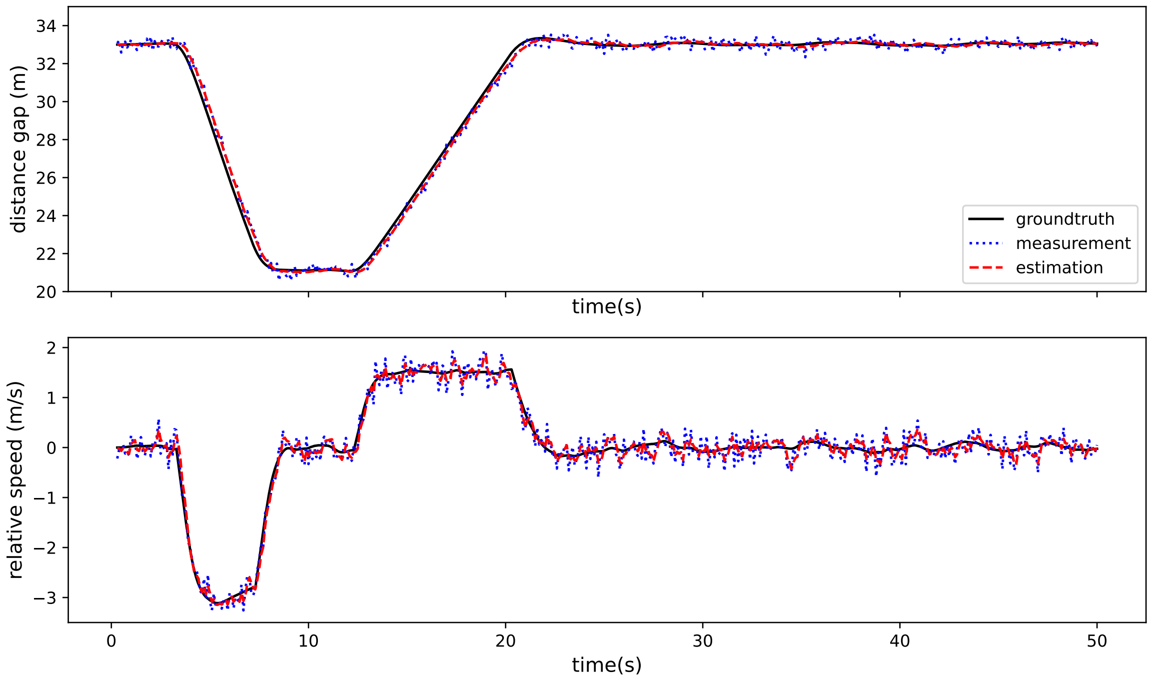


Figure 5.38: Filtered distance gap and relative speed measurements for the one-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

Figures 5.40 and 5.41 look at the string stability indicators in detail. As can be seen in Figure 5.40, the *OG* values after using KF can be reduced to almost the same level as the perfect information scenario although the improvement may be insignificant by looking at the profiles. The curve of *OS* values also shows similar improvement. On the other hand, the system designed with the second weighting combination does not show much improvement in both indicators after using a KF. Figure 5.41 shows that both the *OG* and *OS* curves overlaps with the curves without using any state estimator. This is probably due to the fact that the string stability of the system is already at the performance limit in the scenario with this level of measurement noise. More improvement is expected for ride comfort performance of the system after applying the KF.

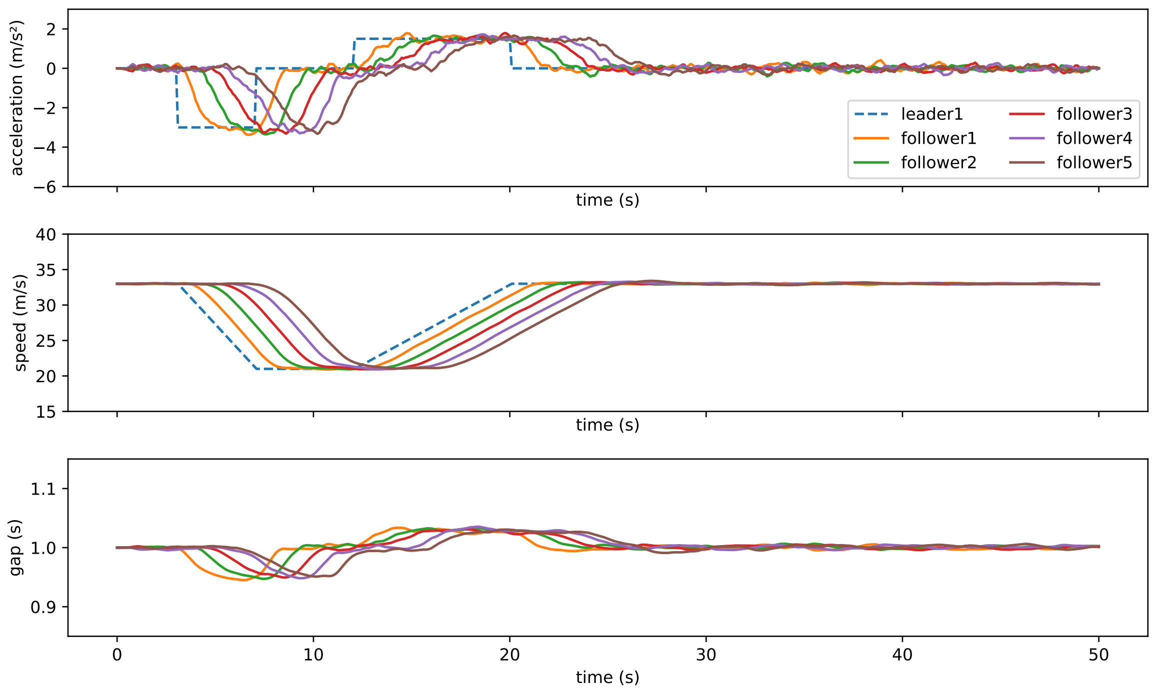


Figure 5.39: Acceleration, speed, and time gap profiles of the platoon using the one-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

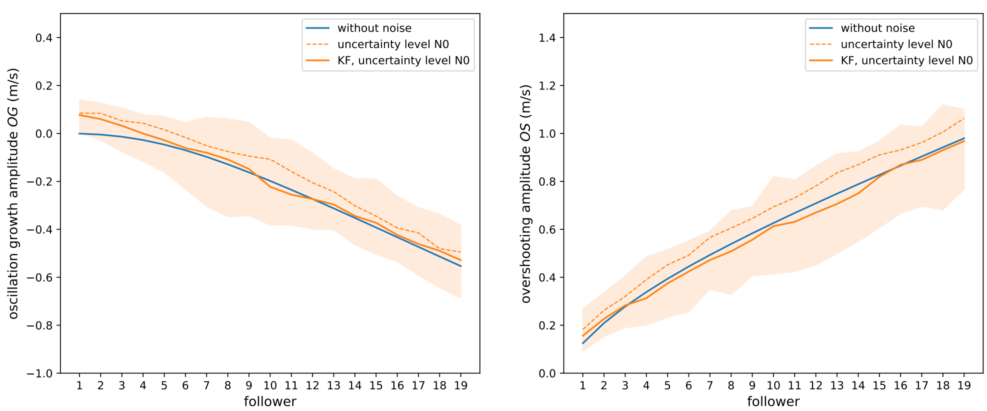


Figure 5.40: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader KF-ACC system designed with the first weighting combination with and without measurement noise

The ride comfort after using the one-leader KF-ACC system is evaluated. Figure 5.42

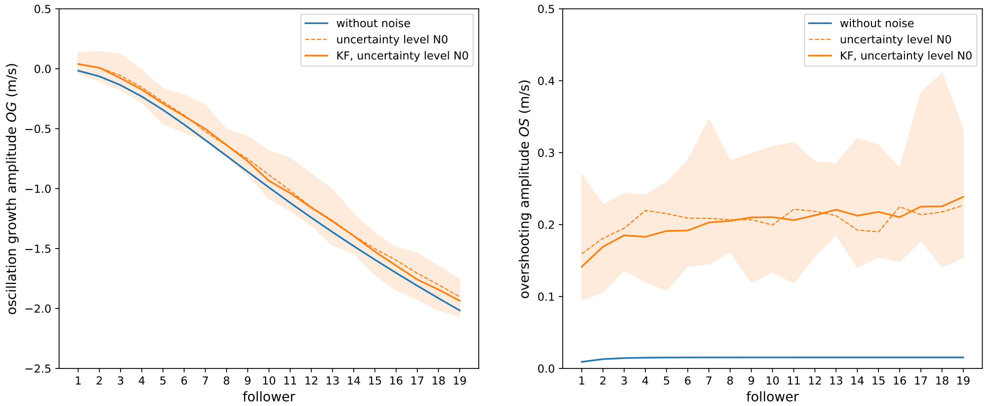


Figure 5.41: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader KF-ACC system designed with the second weighting combination with and without measurement noise

plots the distribution of jerks experienced by the followers equipped with the system designed with the first weighting combination. By comparing with the system without using any state estimator, it is found that the jerk amplitudes are significantly reduced. Figure 5.43 also plots the distribution of jerks experienced by the followers equipped with the system designed with the second weighting combination. The jerk amplitudes are also reduced after applying a KF in the system, but the effect is not as significant as it is in the previous system designed with the first weighting combination. There are still many jerk amplitudes which exceed the threshold of comfortable driving maneuvers.

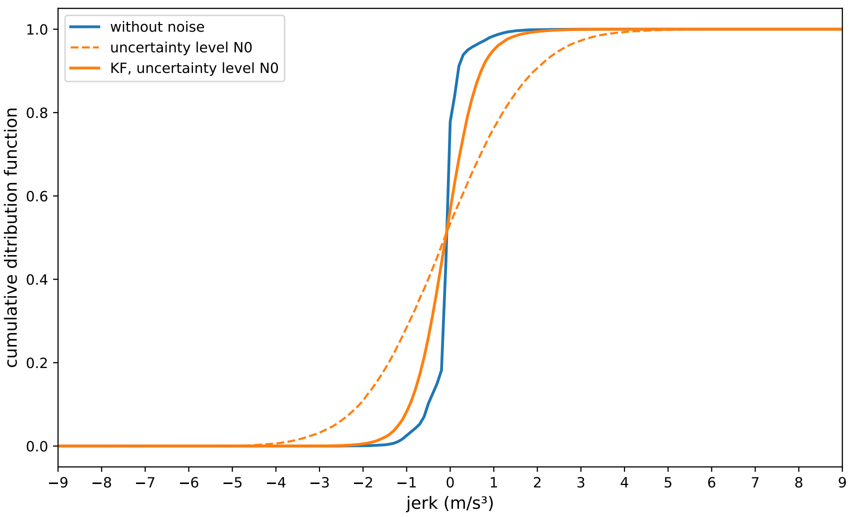


Figure 5.42: Cumulative distribution function of jerk experienced by the platoon using the one-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

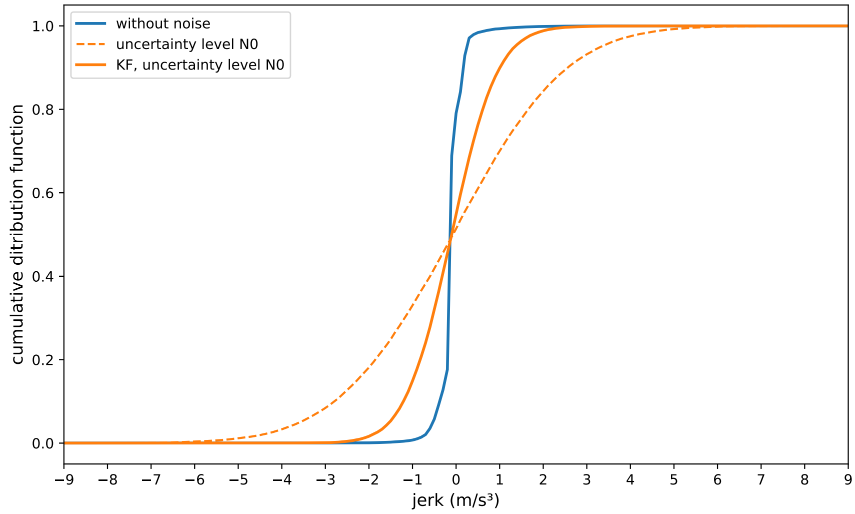


Figure 5.43: Cumulative distribution function of jerk experienced by the platoon using the one-leader KF-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N0

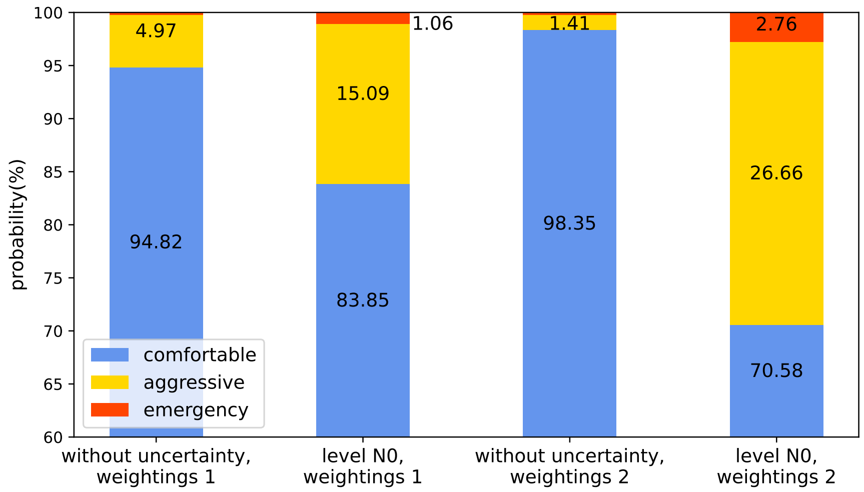


Figure 5.44: Probability distribution of jerks in each ride comfort level experienced by the platoons using the one-leader KF-ACC systems in scenarios with and without measurement noise

Figure 5.44 again provides the detailed probability distribution in each ride comfort level. By looking at the one-leader KF-ACC system designed with the first weighting combination, there are only 1% of jerk amplitudes in the range of emergency driving maneuvers, which is a lot closer to the performance in the scenario with accurate measurements than the ACC system without using any state estimator. The system designed with the second weighting combination also shows a lot of improvement after using a KF. These analysis results clearly implies the improved ride comfort brought by the filtering method adopted.

However, the number of aggressive driving maneuvers still increases, implying the influence of the weighting setup which does not favor jerk limitation.

5.4.2. Performance of two-leader KF-ACC systems

The performance of the two-leader KF-ACC system in multiple levels of measurement noise is then evaluated. An example of the filtered output of the second leader measurements with uncertainty level N4 is shown in Figure 5.45. The smoothing effect of the filter can be observed significantly. The filter can also follow the changes of leader behavior during the deceleration and acceleration phase.

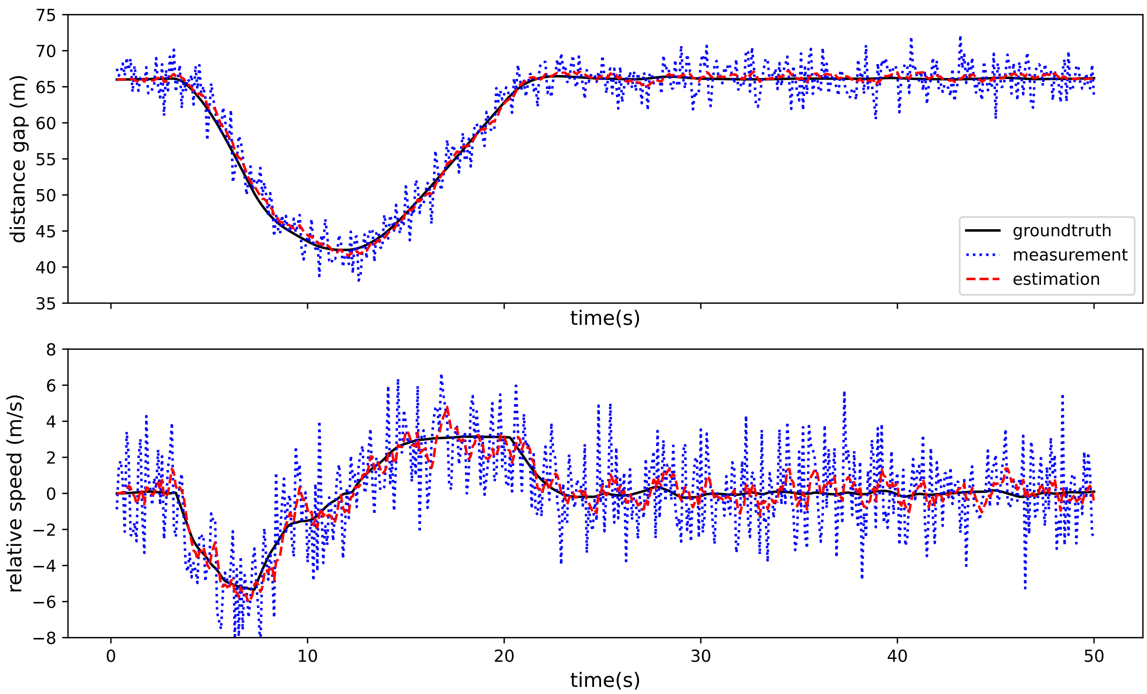


Figure 5.45: Filtered distance gap and relative speed measurements for the two-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N4

Figures 5.46 and 5.47 present examples of the profiles of the first six vehicles in the platoon which are equipped with the system designed with the first weighting combination when dealing with the uncertainty levels N1 and N4 in the second leader measurements, respectively. Compared to the system without using any state estimator, the acceleration profiles in both levels of measurement noise become smoother. On the other hand, the improvement can also be found by looking at the time gap profiles in Figure 5.47. The deviation of time gap 2 when facing the large measurement noise becomes smaller and closer to it is in the scenario with perfect information. The early slow down behaviors can also be observed again in the plot of time gap 1, which disappear originally when the system does not use any state estimator. However, there are still many large acceleration fluctuations in the acceleration profile in Figure 5.47. Such fluctuations lead to large jerks in the motions of the following vehicles, causing more irregular speed profile and even more severely degraded driving comfort than they are in the scenario without using any state estimator.

Figures 5.48 and 5.49 again plot the string stability indicators for the two-leader KF-ACC systems. Figure 5.48 plots the indicators for the system designed with the first weighting combination. As can be seen, the curves of the *OG* values largely overlap with the curve

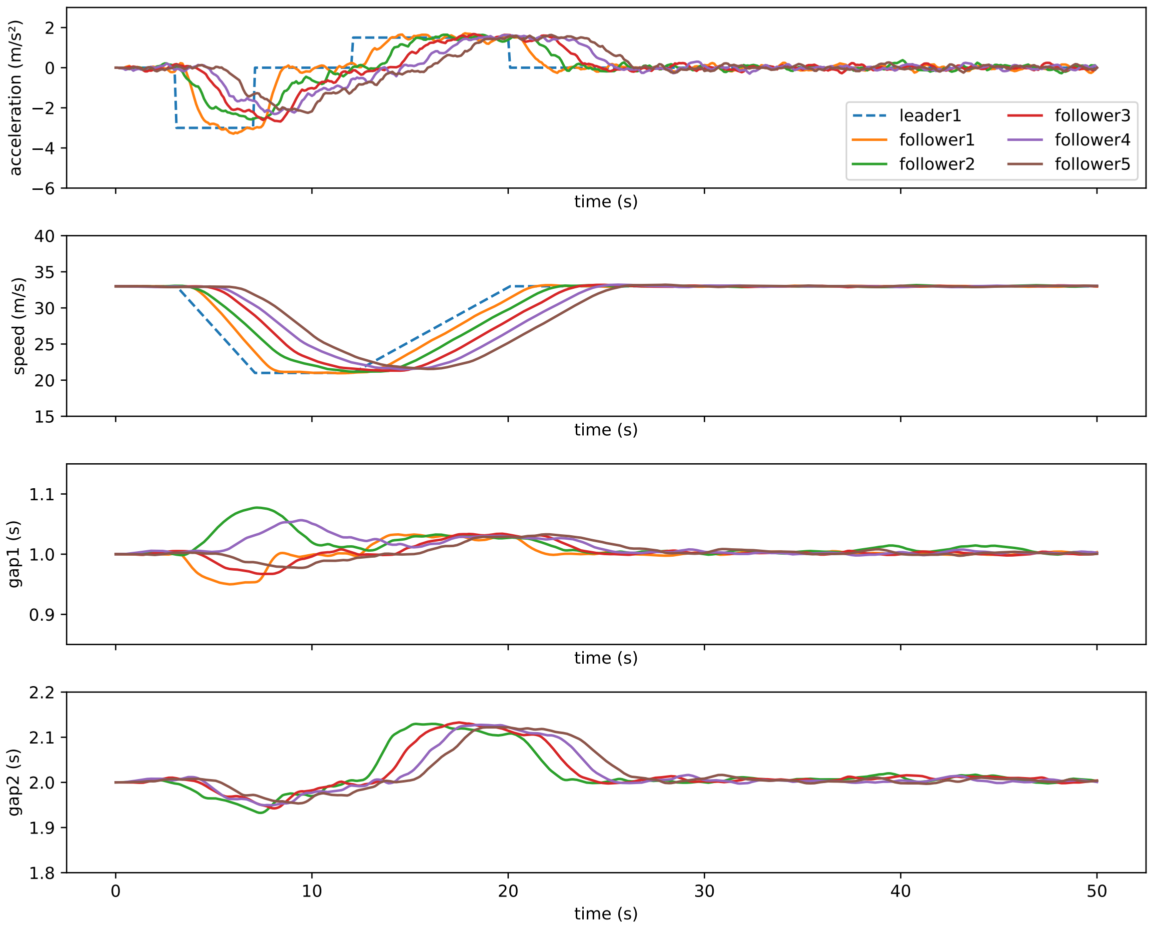


Figure 5.46: Acceleration, speed, and time gap profiles of the platoon using the two-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N1

in the scenario with perfect information, demonstrating the effectiveness of applying KFs. However, it is worth noting that KFs also bring negative effect to the *OS* values. As can be seen in the plot of *OS* curves, it is found that filtering out the measurement noise actually makes the overshooting effect more severe in the scenarios with measurement uncertainty levels N3 and N4. The *OS* values increase while propagating to the upstream of the platoon, which is similar to the effect in the scenario with perfect information. The system designed with the second weighting combination also shows improvement for the *OG* values. After using KFs, the curves of every measurement uncertainty level become closer to the curve of the scenario without any state estimator, indicating improved string stability. However, the curves of scenarios with measurement uncertainties still deviate more from the curve of the perfect information scenario than the situations in Figure 5.48. In addition, the *OS* curves do not show significant difference after applying KFs. This implies that the system designed with the second weighting combination benefit from the state estimation of KFs less than the system designed with the first weighting combination does, which is again the effect of different weighting setups in the policy training.

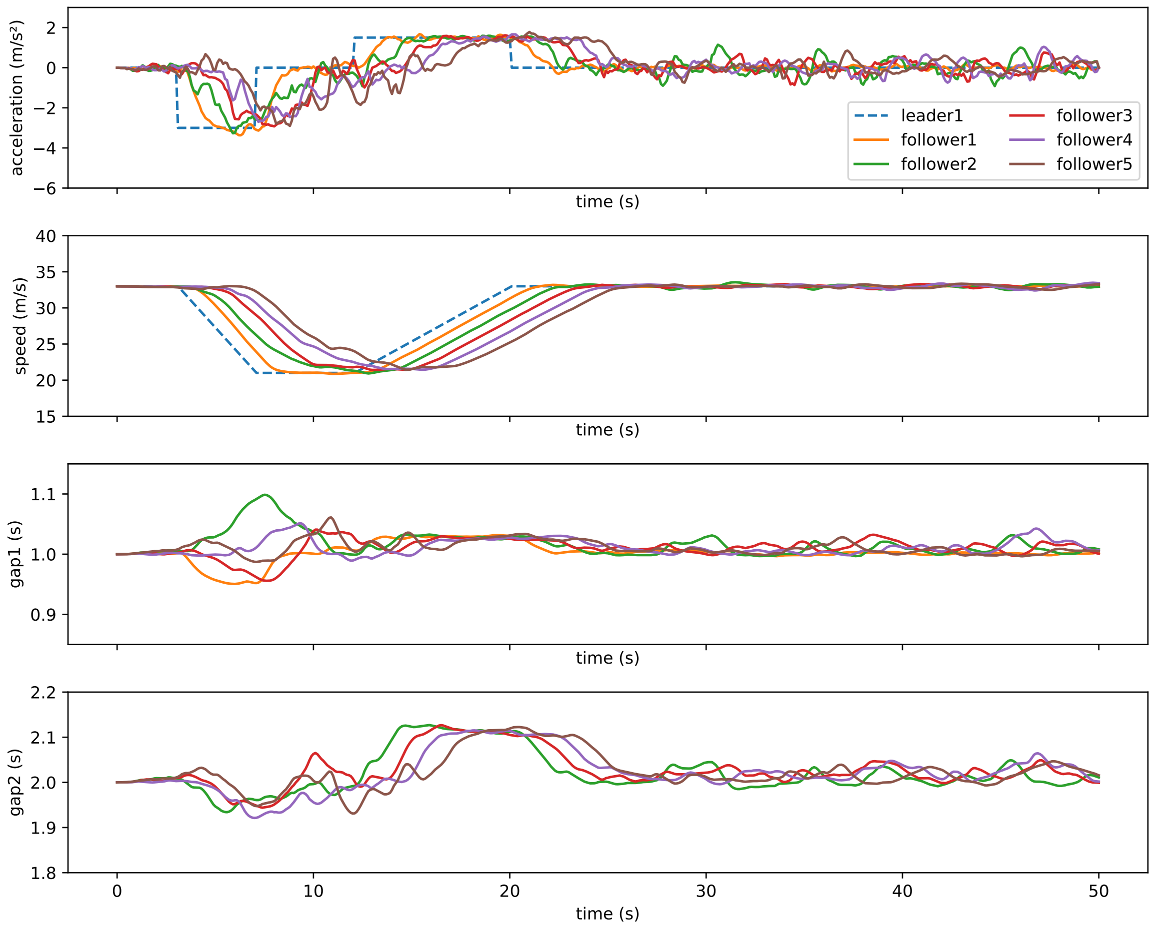


Figure 5.47: Acceleration, speed, and time gap profiles of the platoon using the two-leader KF-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N4

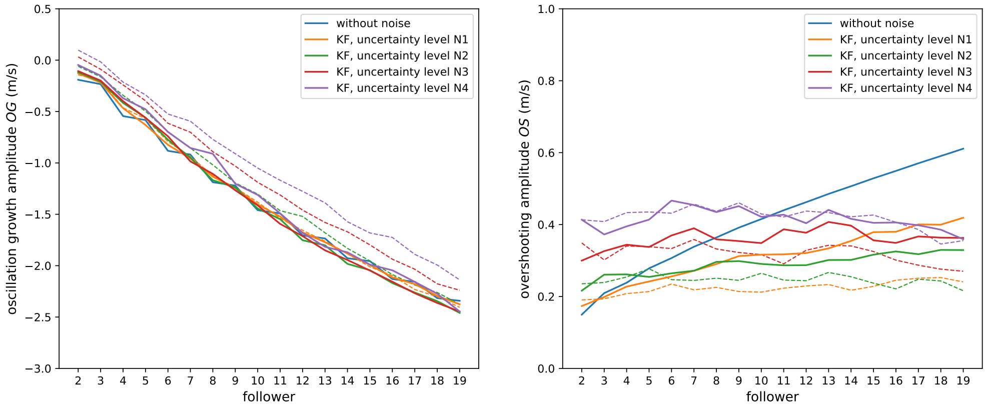


Figure 5.48: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader KF-ACC system designed with the first weighting combination in scenarios with and without measurement noise

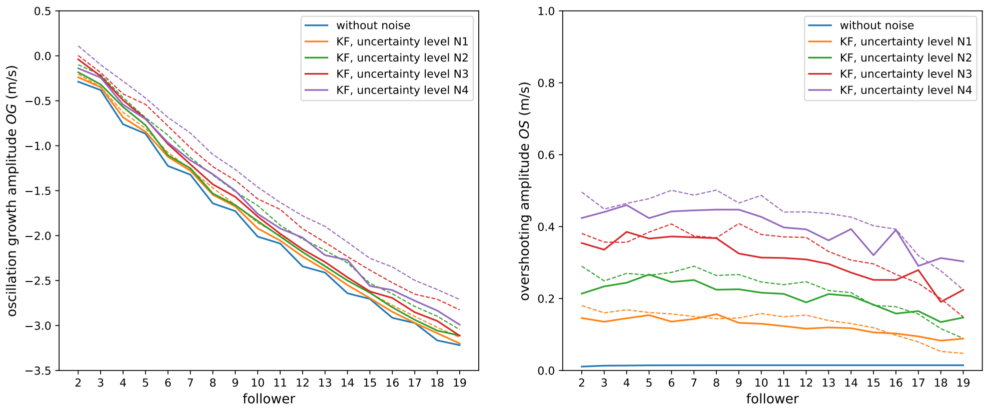


Figure 5.49: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader KF-ACC system designed with the second weighting combination in scenarios with and without measurement noise

Figures 5.50 and 5.51 present and compare the distribution of jerk and experienced by the followers before and after applying KFs as state estimators for the noisy sensor measurements. Similar to the improvement found in the one-leader KF-ACC system, the ride comfort of the two-leader KF-ACC system is also significantly enhanced and become closer to the level of the scenario with accurate measurements. As shown in the figure, a large portion of jerk amplitudes are decreased to the range of comfortable driving maneuver although there are still a certain number of aggressive or emergency driving maneuvers especially in relatively high measurement uncertainty levels. Comparing the systems designed with two different weighting combinations in the two figures, it is also found that the system designed with the first weighting combination still has better ride comfort performance after using KFs.

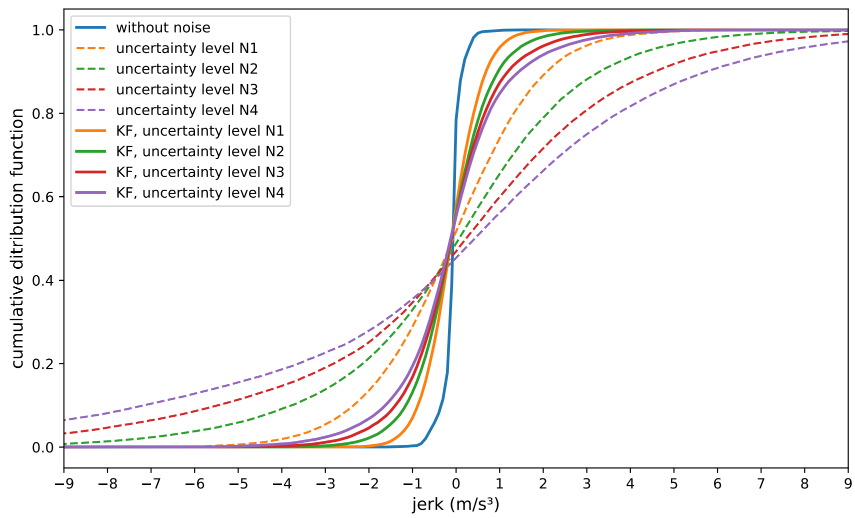


Figure 5.50: Cumulative distribution function of jerk experienced by the platoon using the two-leader KF-ACC system designed with the first weighting combination in scenarios with and without measurement noise

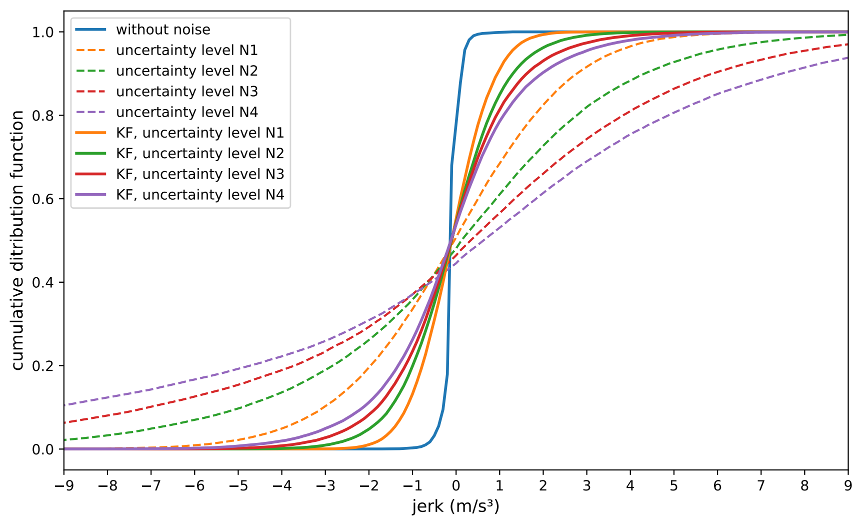


Figure 5.51: Cumulative distribution function of jerk experienced by the platoon using the two-leader KF-ACC system designed with the second weighting combination in scenarios with and without measurement noise

As shown in Figure 5.52, KFs in the system designed with the first weighting combination can successfully prevent at least 60-90% of the jerk values from exceeding the threshold of comfortable driving behavior depends on the measurement uncertainty level considered. However, the emergency driving maneuvers still exist. In uncertainty level N4, the number of emergency maneuver even exceeds 10%. For the performance of two-leader KF-ACC system designed with the second weighting combination shown in Figure 5.53, the ride comfort performance when facing measurement noise becomes even worse. Only 40-70% of the driving maneuvers can stay in the range of comfortable maneuvers. At measurement uncertainty level N4, more than 20% of the maneuvers exceed the 2 m/s^3 threshold.

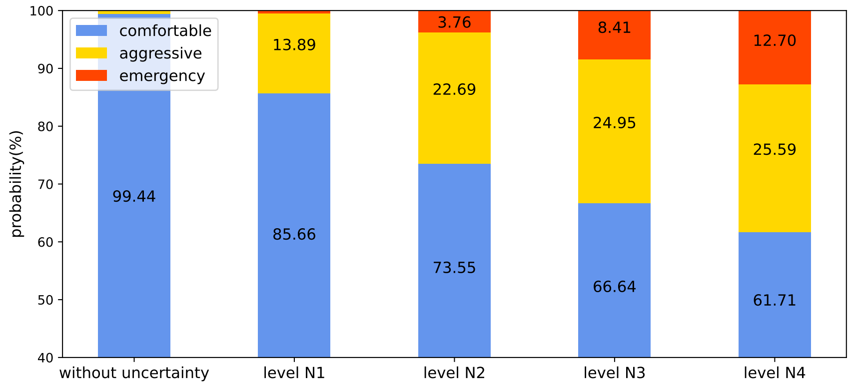


Figure 5.52: Probability distribution of jerks in each ride comfort level experienced by the platoons using the two-leader KF-ACC system designed with the first weighting combination in scenarios with and without measurement noise

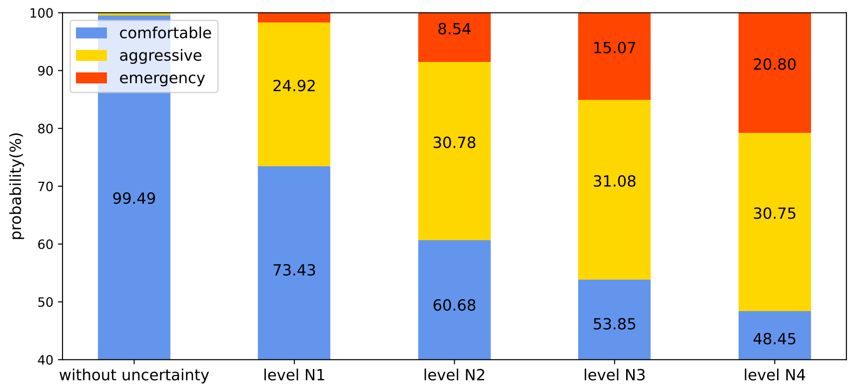


Figure 5.53: Probability distribution of jerks in each ride comfort level experienced by the platoons using the two-leader KF-ACC system designed with the second weighting combination in scenarios with and without measurement noise

By implementing KFs before the noisy measurements are fed back to the controller in the loop of the system architecture, the system aims to estimate the true state for the designed controllers. According to the results shown in this section, it is found that a tuned conventional filtering approach is able to reduce the impact of measurement noise and brings the string stability performance back to the level of scenarios with accurate measurements. In addition, the vehicle acceleration profile becomes smoother than in the scenarios without using any state estimator. However, the problem of fluctuating accelerations still remains especially when the measurement uncertainty level is high. This leads to severely degraded ride comfort performance of the platoon. Therefore, it may be concluded that there seems to be a performance limit in terms of ride comfort for the KF- ACC systems when encountering large measurement noise.

5.5. Performance of systems using recurrent policies

After exploring the performance of the KF-ACC systems and their potential strengths and shortcomings, this section analyzes the performance of the LSTM-ACC systems which use recurrent policies to control the car-following behavior of the following vehicles.

5.5.1. Performance of one-leader LSTM-ACC systems

This section presents the simulation and evaluation results of the one-leader LSTM-ACC system. Figure 5.54 first shows an example of the acceleration/speed/gap profiles of the first five followers in the platoon which are equipped with the system designed with the first weighting combination. Although it is still unclear whether the deceleration wave is amplified or not according to these profiles, it can be observed that the following vehicles exhibit abnormal behavior when speeding up again to the stabilization speed. They tend to accelerate more than their leaders at the beginning of the acceleration phase and adjust its acceleration afterward. This abnormal behavior, to a certain extent, indicates the imperfect performance of the trained policy in conducting the car-following behavior when facing the disturbance. In addition, the overshooting amplitudes are clearly increasing for the upstream vehicles.

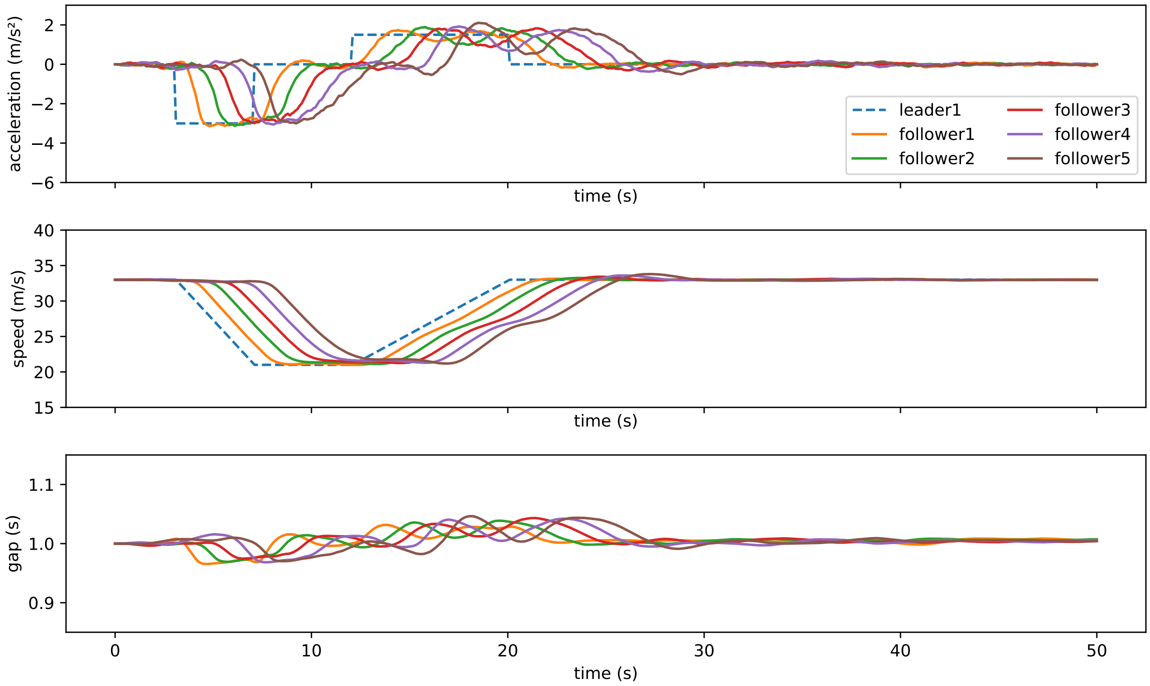


Figure 5.54: Acceleration, speed, and time gap profiles of the platoon using the one-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

From Figure 5.55, it can be observed that the oscillation is indeed slightly damped out for vehicles in the middle of the platoon, but the minimum speed seems to start decreasing again for vehicles further upstream. A second fluctuation wave emerges from the original fluctuation. In addition, the overshooting effect also significantly becomes more severe than any other systems have experienced before.

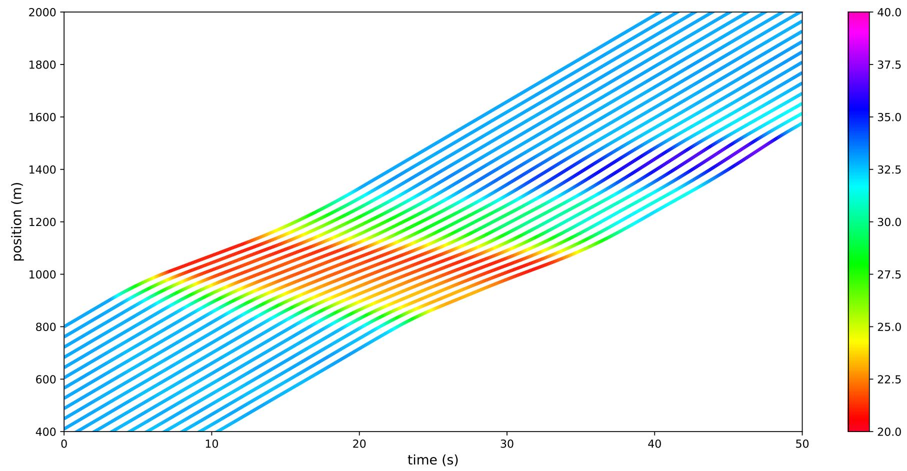


Figure 5.55: Vehicle trajectories and speed contours of the platoon using the one-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N0

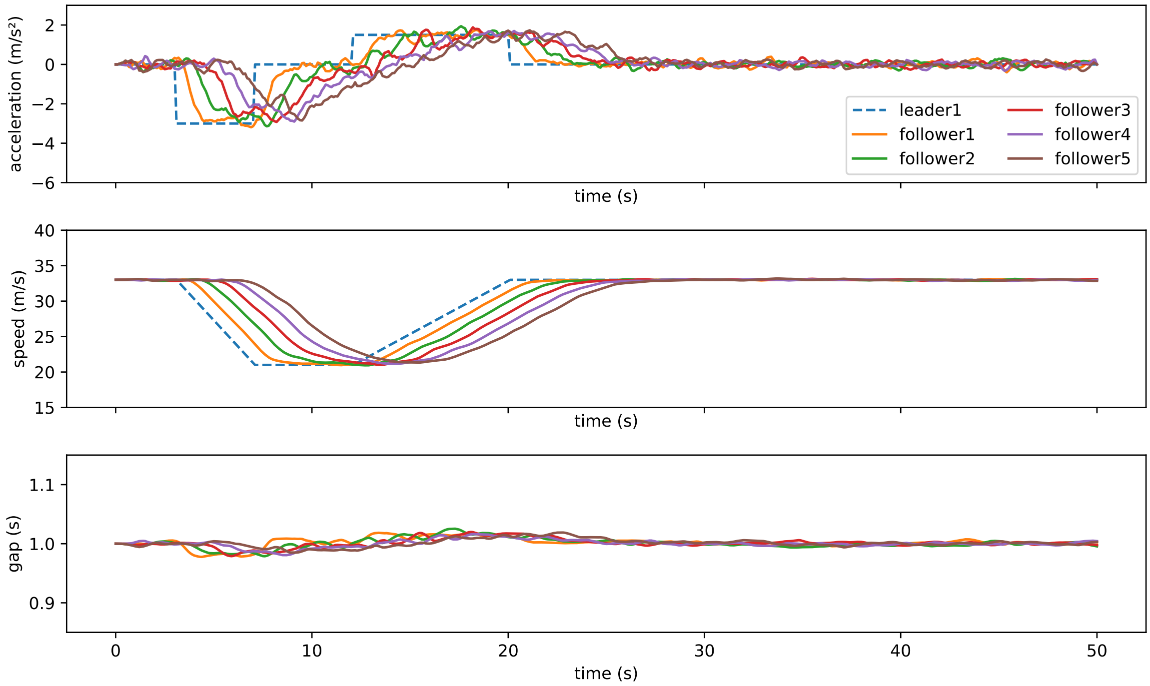


Figure 5.56: Acceleration, speed, and time gap profiles of the platoon using the one-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N0

The performance of the one-leader LSTM-ACC system designed with the second weighting combination is then discussed. Figure 5.56 shows the acceleration/speed/gap profiles of first five followers in the platoon equipped with the system. It is clearly seen that the system does not perform very well in smoothing out the acceleration fluctuations due to the measurement noise. However, the overall patterns of the speed and time gap profiles are more normal for the car-following behavior of vehicles when encountering disturbances than those vehicles equipped with the system designed with the first weighting combination. This shows that different weighting combination setups significantly affect the performance of the final policies. The second weighting setup generates better control policy in this case. As can be seen in the figure, the disturbance is not amplified, and the overshooting phenomena do not exist.

Figure 5.57 illustrates the speed contours of all the vehicles in the platoon equipped with the system designed with the second weighting combination. Different from the previous system designed with the first weighting combination, the disturbance can be damped out while propagating to the upstream. The conservative driving behaviors, as mentioned in subsection 5.2.1, can also be observed again. No overshooting behaviors can be found by looking at the speed contours. These all demonstrate the effectiveness and improvement brought by the one-leader LSTM-ACC system which was designed with the second weighting combination.

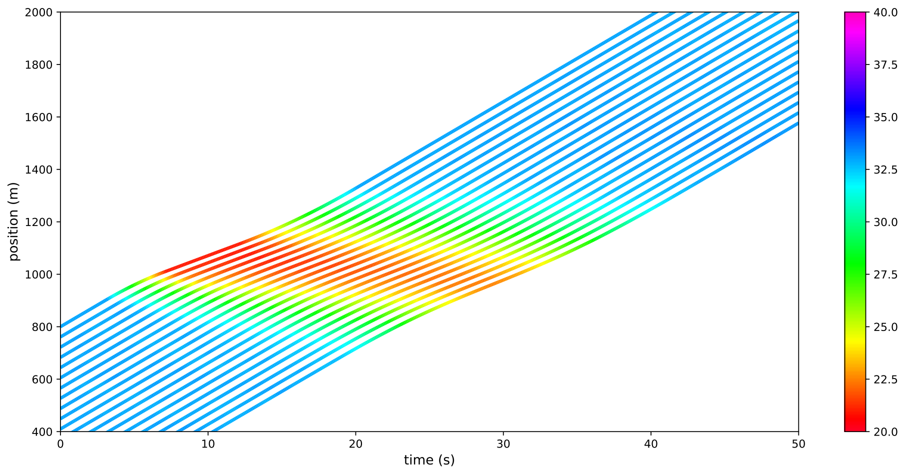


Figure 5.57: Vehicle trajectories and speed contours of the platoon using the one-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N0

The *OG* and *OS* curves of the vehicle platoon using the one-leader LSTM-ACC system designed with the first weighting combination are shown in Figure 5.58. The *OG* curve on the left side shows that the oscillation amplitude is decreasing even faster than the one-leader ACC system in the scenario with accurate measurements. However, the *OG* value starts increasing from follower 12. It is inferred that this is the influence of the other disturbance wave formed in the original disturbance. It is observed that vehicles tend to reach a constant speed which is slightly above the lowest speed of their leader first and decelerate again to reach the desired state, which later on results in the second disturbance which is amplified while propagating to the upstream. On the other hand, the *OS* curve provides a quantitative view of the severe overshooting effect. As shown by the curve, the *OS* value even exceeds 4 m/s for follower 19 at the end of the platoon. The string stability performance of the one-leader LSTM-ACC system designed with the first weighting combination is deteriorated to the level which is worse than the system without using any state estimator.

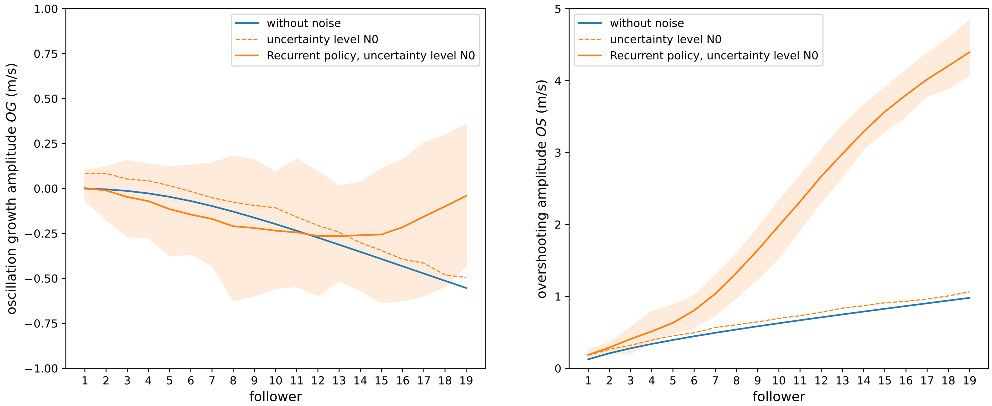


Figure 5.58: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader LSTM-ACC system designed with the first weighting combination in scenarios with and without measurement noise

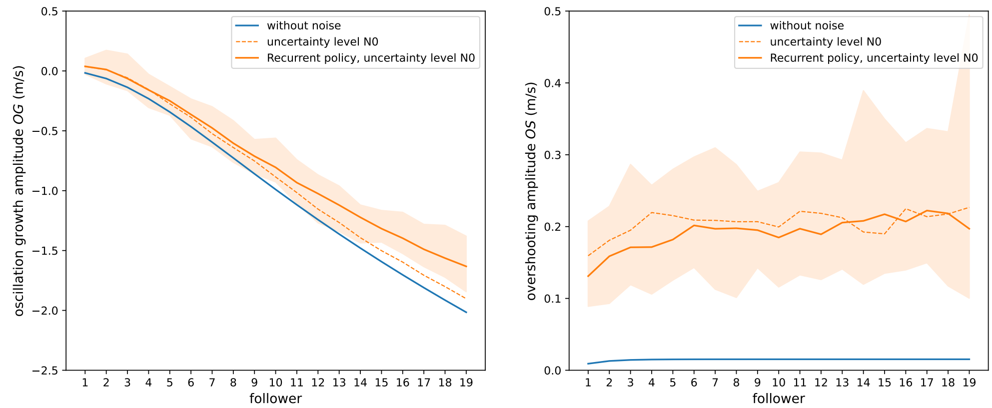


Figure 5.59: Oscillation growth amplitude (left) and overshooting amplitude (right) of the one-leader LSTM-ACC system designed with the second weighting combination in scenarios with and without measurement noise

The string stability performance of the one-leader LSTM-ACC system designed with the second weighting combination is quantitatively analyzed and shown in Figure 5.59. It is found that the *OG* curve is decreasing in the upstream direction, indicating that the system shows string stable performance. However, the curve of the one-leader LSTM-ACC system stays above the curve of the system without using any state estimator at the upstream of the platoon, meaning that the trained recurrent policy does not possess the same level of performance as the previous system when facing noisy measurements. To handle the measurement noise, the string stability performance is slightly degraded. By looking at the *OS* curves in the right plot of the figure, it is also found that the one-leader LSTM-ACC system does not improve the overshooting effect compared to the one-leader ACC system without using any state estimator. Therefore, it can be concluded that the one-leader LSTM-ACC systems do not show any improvement in terms of string stability.

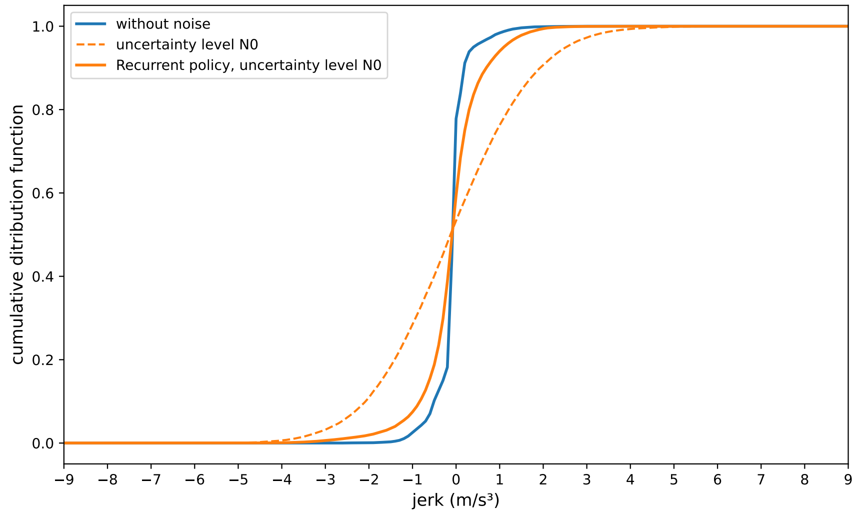


Figure 5.60: Cumulative distribution function of jerk experienced by the platoon using the one-leader LSTM-ACC system in the scenario with measurement uncertainty level N0

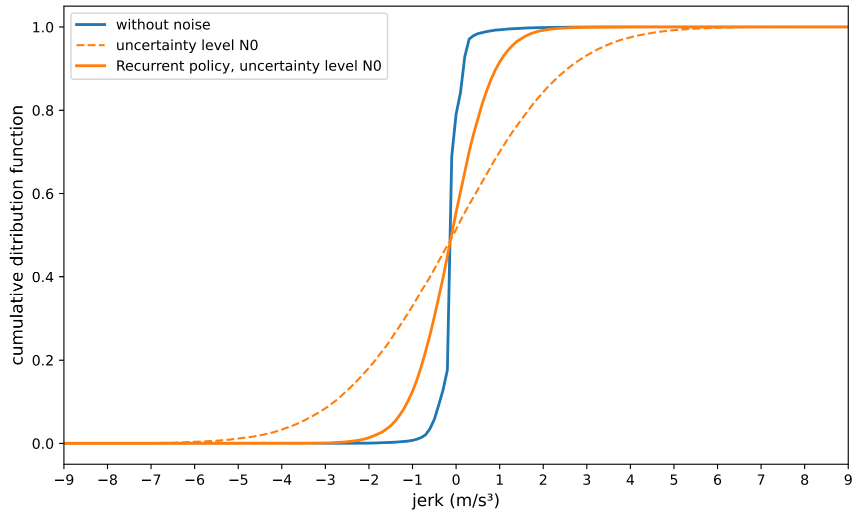


Figure 5.61: Cumulative distribution function of jerk experienced by the platoon using the one-leader LSTM-ACC system in the scenario with measurement uncertainty level N0

The ride comfort performance of the one-leader LSTM-ACC systems is also analyzed. Figure 5.60 and 5.61 show the distributions of the jerks experienced by the following vehicles equipped with the systems designed with the two weighting combination setups, respectively. As can be seen in the two figures, the cumulative distributions of jerks clearly get closer to the distributions in the scenario with accurate measurements. Comparing the two systems, it is also found that the system designed with the second weighting combination shows worse ride comfort performance than the system designed with the first weighting combination, which again demonstrates the influence of different weighting setups especially when measurement noise is considered.

Figure 5.62 summarizes the ride comfort performance of the one-leader LSTM-ACC systems by calculating the distribution of jerks in each ride comfort level. For the system designed with the first weighting combination, it is found that although it indeed significantly reduces the amount of jerks exceeding the threshold of comfortable maneuver compared to the scenario without using any state estimator (Figure 5.29), its performance is slightly worse than that of the system using a KF by looking at the amount of emergency maneuvers. The system designed with the second weighting combination, on the other hand, has a slightly higher number of jerks in the range of comfortable driving maneuver than the corresponding one-leader KF-ACC system, showing a better ride comfort performance.

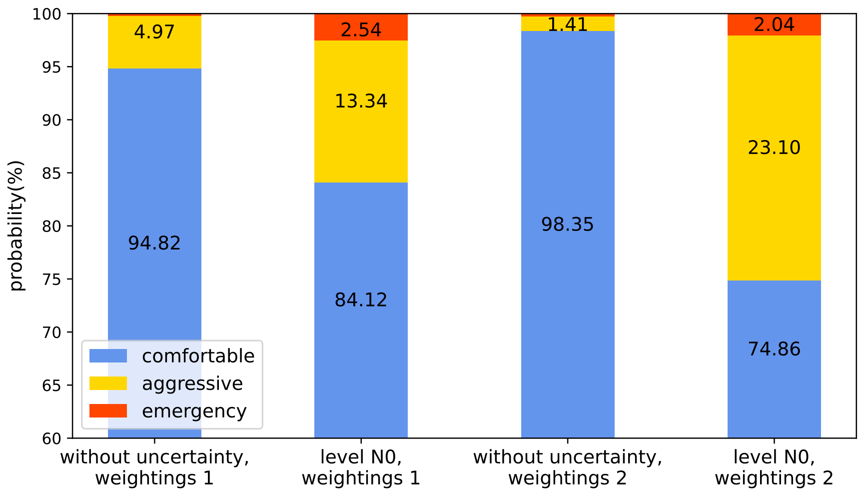


Figure 5.62: Probability distribution of jerks in each ride comfort level experienced by the platoons using the one-leader LSTM-ACC system in scenarios with and without measurement noise

To sum up for the one-leader LSTM-ACC system, the trained recurrent policy for the system designed with the first weighting combination shows string instability and worse ride comfort compared to traditional filtering approach. The system designed with the second weighting combination also does not show improved string stability performance compared to its counterpart without using any state estimator although it can still ensure string stability as the disturbance can be damped out. However, the system does have positive effect on the ride comfort performance compared to its counterpart using a KF. In general, it is still difficult to train a recurrent policy which can completely reduce the impact of measurement noise on the string stability and ride comfort at the same time.

5.5.2. Performance of two-leader LSTM-ACC systems

According to the results in the previous subsection, it is known that the one-leader LSTM-ACC systems do not perform better than the one-leader KF-ACC system or even the system without using any state estimator in terms of string stability. In particular, the system designed with the first weighting combination even leads to string instability at the upstream of the platoon. Therefore, for the first two-leader LSTM-ACC system in this subsection of evaluation, it will use the one-leader KF-ACC system for the following task with the first leader. The controllers using recurrent policies are only applied for the car-following task with the second leader. For the second system designed with the second weighting combination, both ACC controllers using recurrent policies will be used.

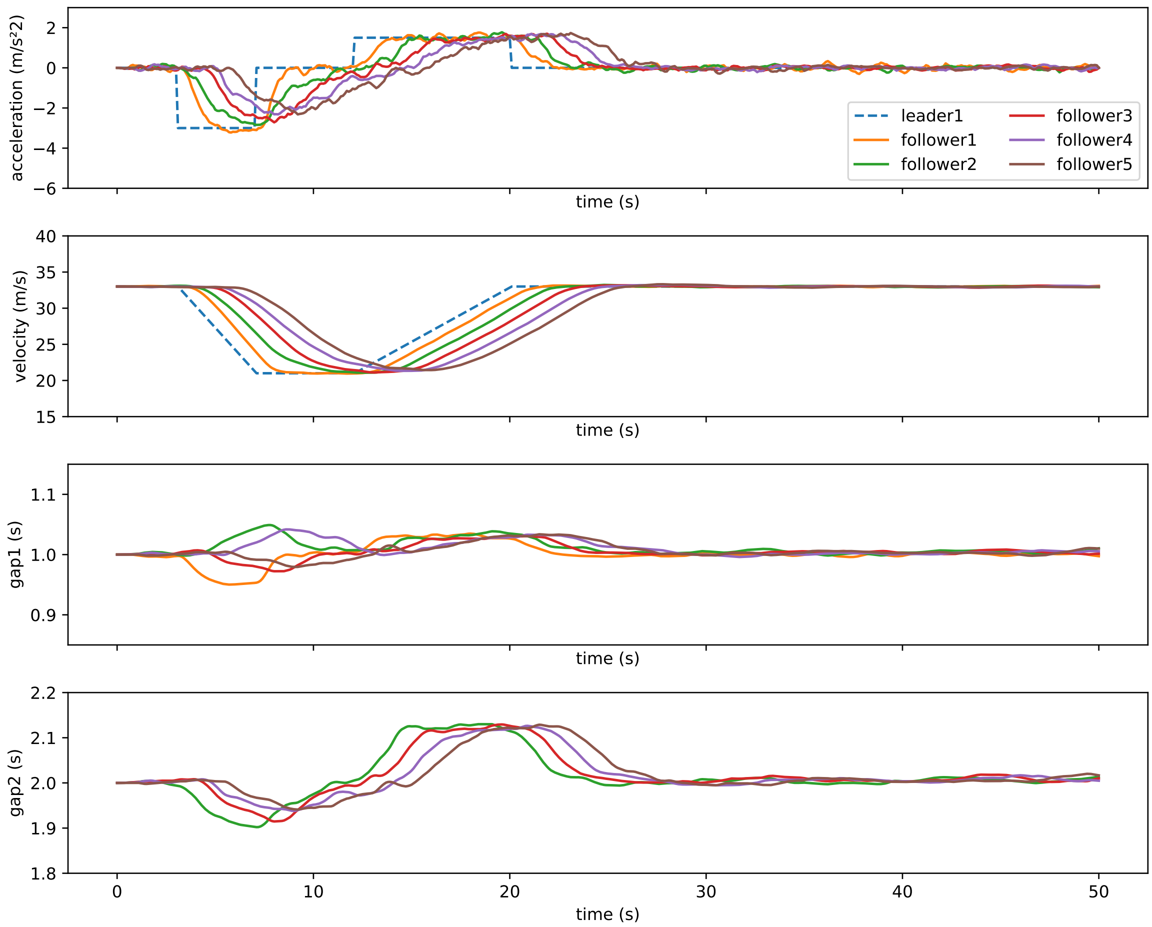


Figure 5.63: Acceleration, speed, and time gap profiles of the platoon using the two-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N1

The evaluation of two-leader LSTM-ACC systems starts from the system designed with the first weighting combination. Figure 5.63 shows the acceleration/speed/gap profiles of the system in the scenario with measurement uncertainty level N1. As can be seen in Figure 5.63, the behaviors of the following vehicles are pretty much similar to the behaviors shown in subsection 5.2.2 in which the two-leader ACC system is used in the scenario with accurate measurements. This implies that the ACC controllers using the recurrent policy can perfectly track the second leader and generate string stable performance even when the measurements are noisy and help the ego-vehicle leverage the benefit of multi-anticipation.

Figure 5.64 also shows that the disturbance is significantly damped out at the upstream of the platoon. In addition, the overshooting effect can hardly be observed between the acceleration and stabilization phase. From these two figures, it can already be found that the string stability performance of the two-leader LSTM-ACC system is better than the one-leader LSTM-ACC system designed with the first weighting combination.

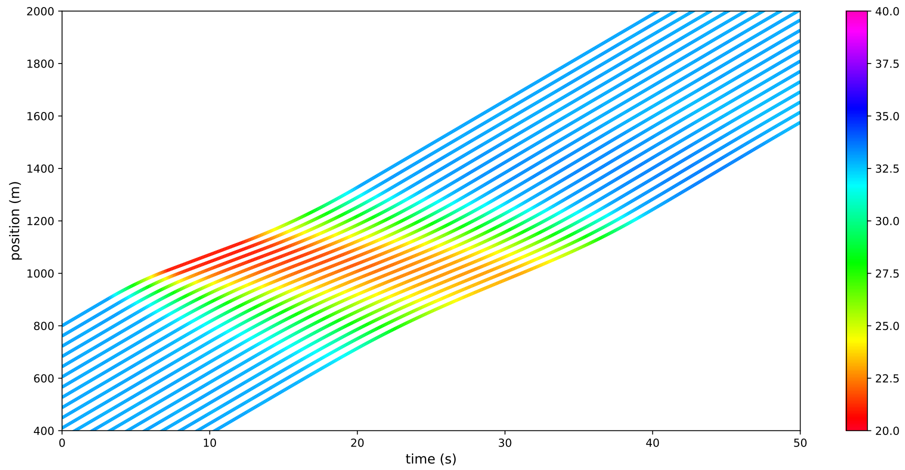


Figure 5.64: Vehicle trajectories and speed contours of the platoon using the two-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N1

The evaluation results of the two-leader LSTM-ACC system in the scenario with measurement uncertainty level N4 are then discussed by looking at Figures 5.65 and 5.66. As shown in the acceleration profile in Figure 5.65, the decreasing deceleration amplitude is still observable. The time gap profiles also exhibit the early slow down behavior mentioned before. However, the overall string stability performance is not as good as it is in the scenario with measurement uncertainty level N1. The time gap profiles show that the early slow down behaviors are not significantly exhibited anymore.

The same effect can also be found in Figure 5.66. It is shown that the vehicles at the upstream of the platoon still decelerate a lot. In addition, the overshooting behaviors appear again, as can be seen in stabilization phase of the vehicle trajectories in the figure. The probable reason of this outcome is the greater difficulty for the agent to estimate the state due to the increased uncertainty level. More measurements are required for the agent to perceive the occurrence of transitions between different phases in the traffic disturbance. The large measurement noise causes longer reaction delay than the scenarios with comparatively small noise, which deteriorates the string stability performance.

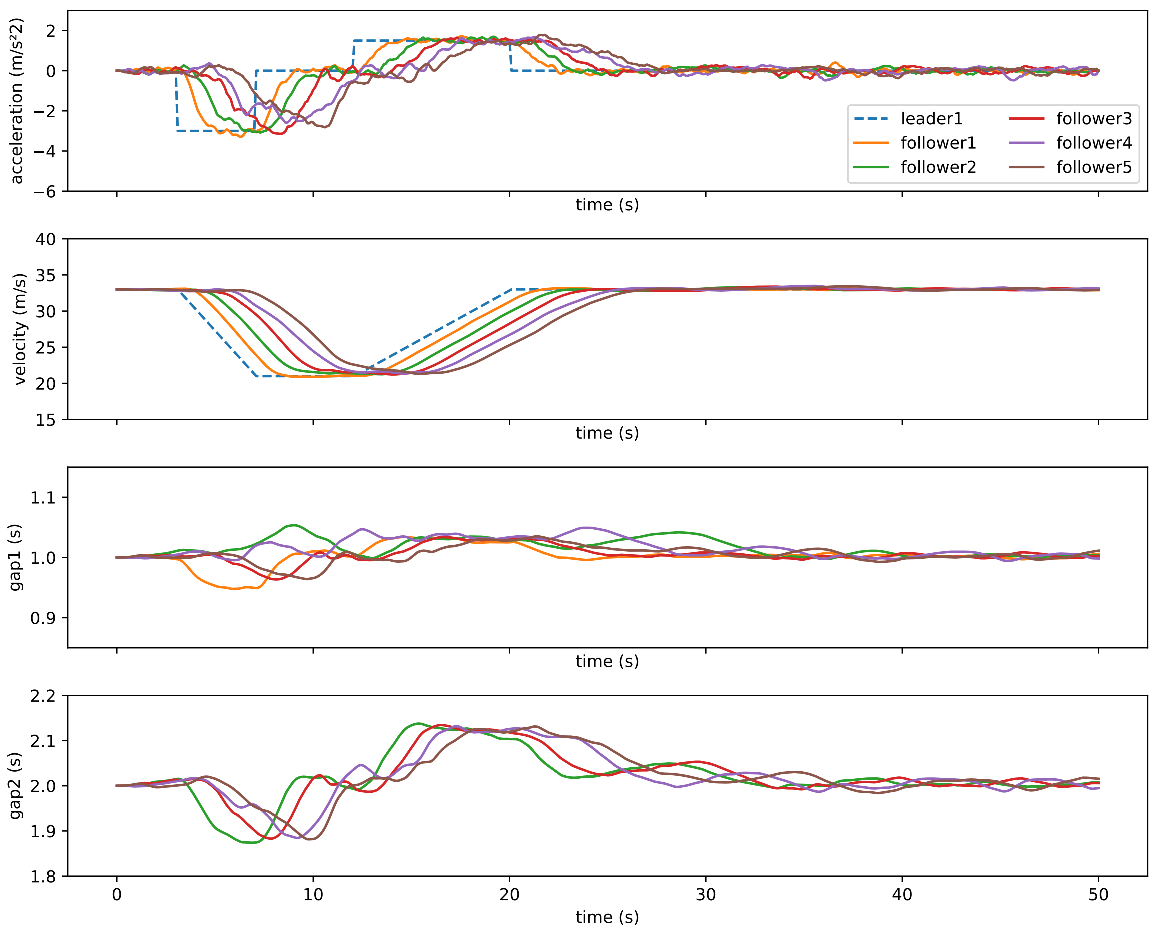


Figure 5.65: Acceleration, speed, and time gap profiles of the platoon using the two-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N4

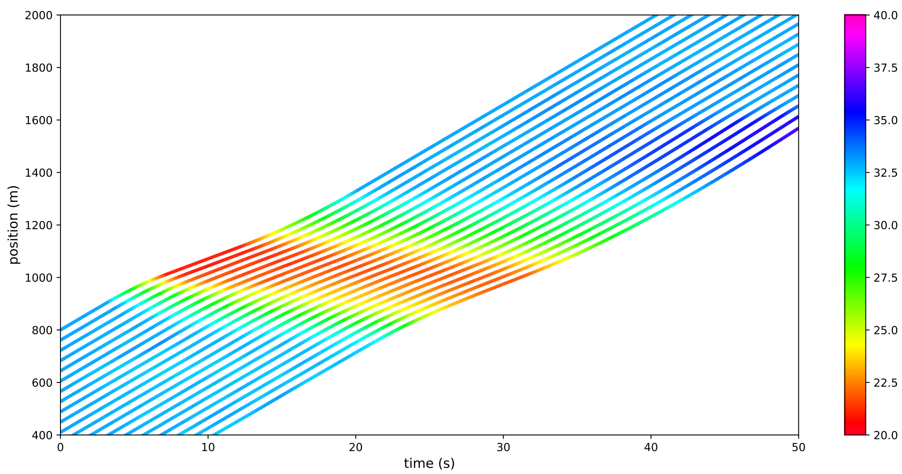


Figure 5.66: Vehicle trajectories and speed contours of the platoon using the two-leader LSTM-ACC system designed with the first weighting combination in the scenario with measurement uncertainty level N4

The performance of the two-leader LSTM-ACC systems designed with the second weighting combination are then presented. Figure 5.67 first shows the acceleration/speed/gap profiles of the first five followers equipped with the system in the scenario with measurement uncertainty level N1. It is found that minimum deceleration amplitude is decreasing while the disturbance is propagating to the upstream, indicating the string stability performance. The difference between the system considered here and the previous system designed with first weighting combination is the amplitude of gap error. As shown in the profile of gap 2, the maximum gap error can always be kept within 0.1 s within the second leader, which is better than the previous system in the scenario with the same level of measurement uncertainty. This also demonstrate the effect of the different weighting setup.

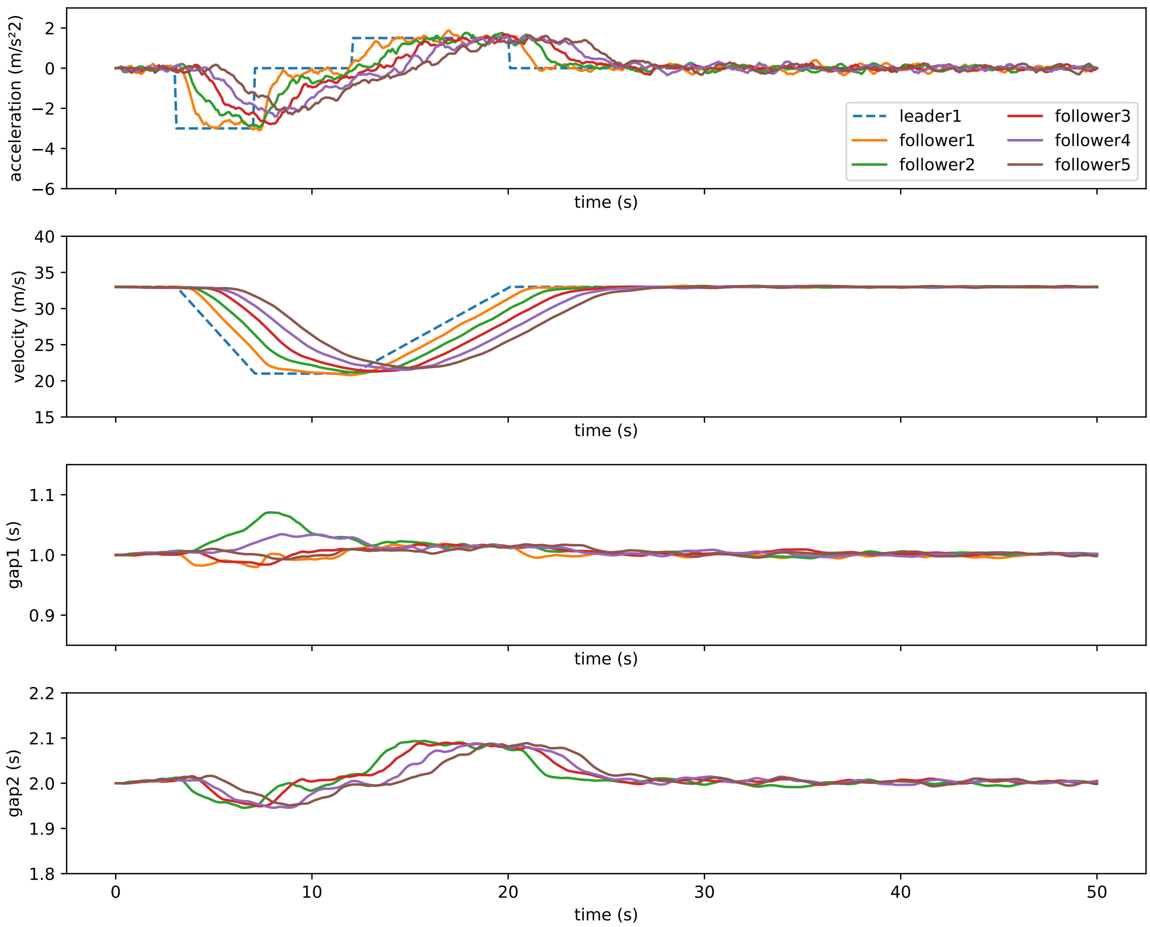


Figure 5.67: Acceleration, speed, and time gap profiles of the platoon using the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N1

The speed contours of the vehicles equipped with the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N1 are then shown in Figure 5.68. As can be seen from the speed contours, the minimum speed of the last follower in the platoon stays above 24 m/s, which is also better than the performance of the previous system designed with the first weighting combination. The disturbance is clearly damped out at the upstream of the platoon. Furthermore, no overshooting behavior can be observed.

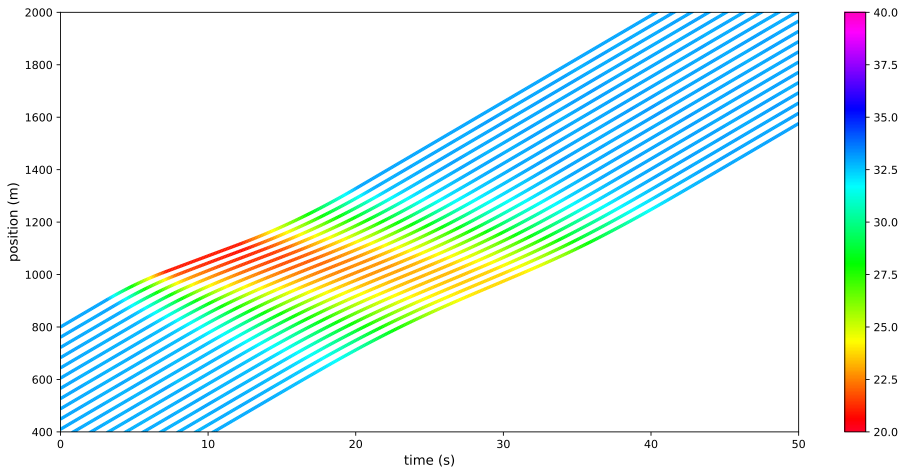


Figure 5.68: Vehicle trajectories and speed contours of the platoon using the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N1

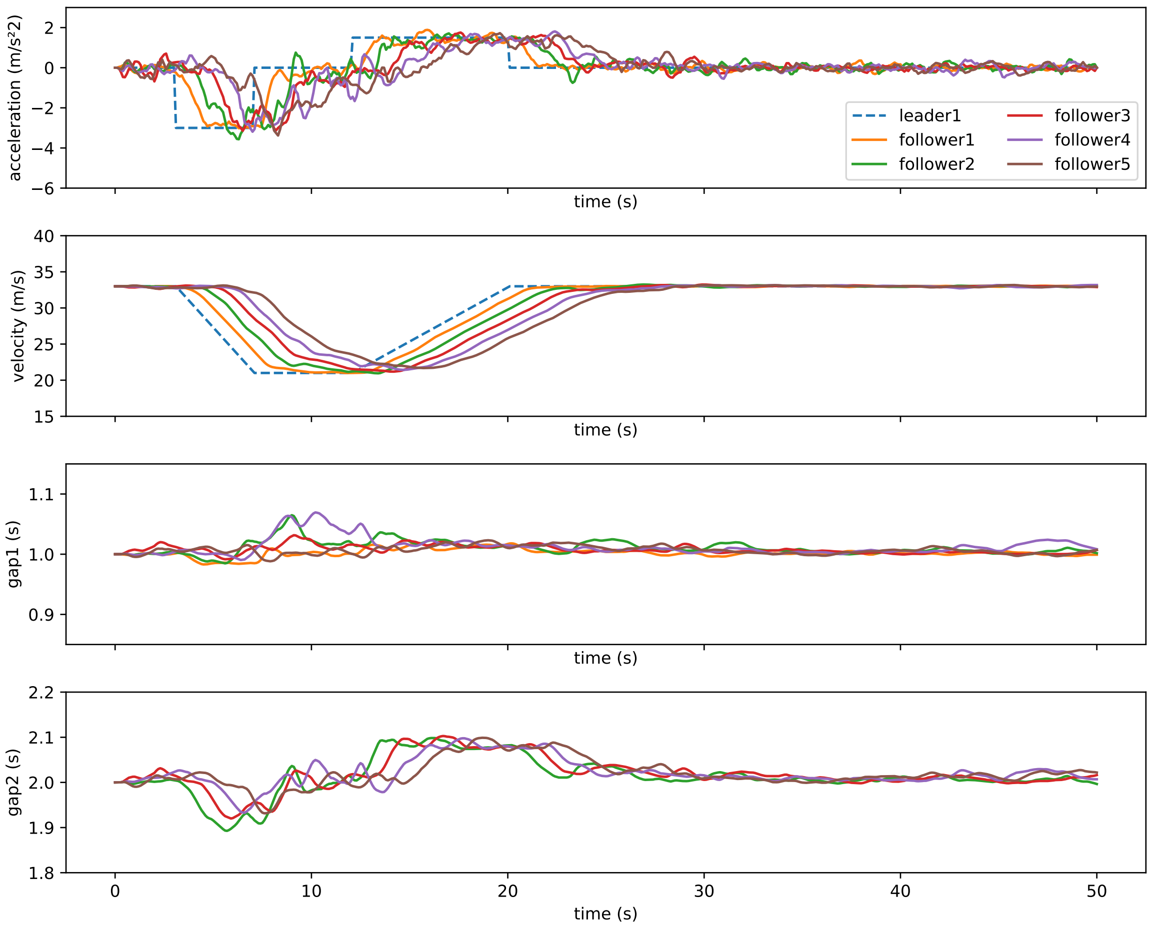


Figure 5.69: Acceleration, speed, and time gap profiles of the platoon using the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N4

Figure 5.69 then shows the acceleration/speed/gap profiles of the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N4. With such high level of measurement uncertainty, the string stability performance is degraded compare to the performance in the scenario with measurement uncertainty level N1, as shown in Figure 5.67. The deceleration amplitudes of the following vehicles become smaller, and the gap errors also become larger. Still, according to the profiles, the disturbance is not amplified, showing a certain level of string stability.

Compared to the speed profile of the previous system designed with the first weighting combination, the pattern of decelerating behaviors of the following vehicles become more uniform, indicating that the car-following behavior of the vehicles equipped with the system is not seriously influenced by the measurement noise. According to the gap profiles, the gap errors also become smaller, showing the improved string stability performance. However, it is also found that there are more fluctuations in the speed profile. These are all the influences caused by the different weighting setups.

The string stability performance of the entire platoon using the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N4 can be observed in Figure 5.70. Although not as significant as in the scenario with measurement uncertainty level N1, the disturbance is also damped out at the upstream of the platoon. In addition, no overshooting behaviors can be found even in such high level of measurement uncertainty. These imply that the string stability performance of the system considered is better than the system designed with the first weighting combination. However, as can be seen in the speed contours, the speed of the following vehicles fluctuates a lot during the entire simulation run. This can be the outcome of the relatively low weighting on the jerk term in the reward function. The drawback of this kind of setup on ride comfort performance will be discussed in detail later in this subsection.

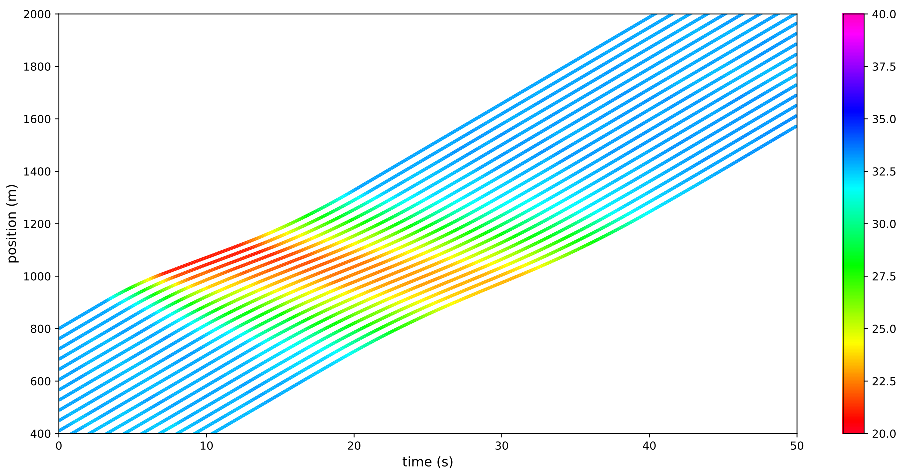


Figure 5.70: Vehicle trajectories and speed contours of the platoon using the two-leader LSTM-ACC system designed with the second weighting combination in the scenario with measurement uncertainty level N4

To have a quantitative understanding of the influence of measurement noise on these recurrent policies, the string stability indicators of the proposed two-leader LSTM-ACC systems in each measurement uncertainty level are shown in Figures 5.71 and 5.72. In Figure 5.71, it is found that the oscillation growth amplitudes are higher than the case without using any state estimator. The decreasing rate of the *OG* values within the platoon becomes slower while the measurement uncertainty level becomes higher. On the other hand, the overshooting effect can remain insignificant in scenarios with small measurement noise, but the *OS* value increases dramatically in uncertainty levels N3 and N4 while the disturbance propagates to the upstream of the platoon. This indicates the degraded string stability performance in scenarios with high level of measurement uncertainty.

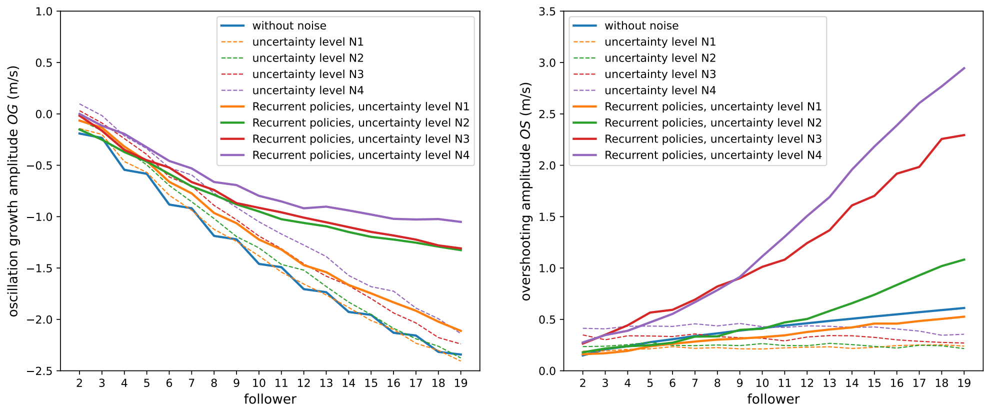


Figure 5.71: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader LSTM-ACC systems designed with the first weighting combination in scenarios with and without measurement noise

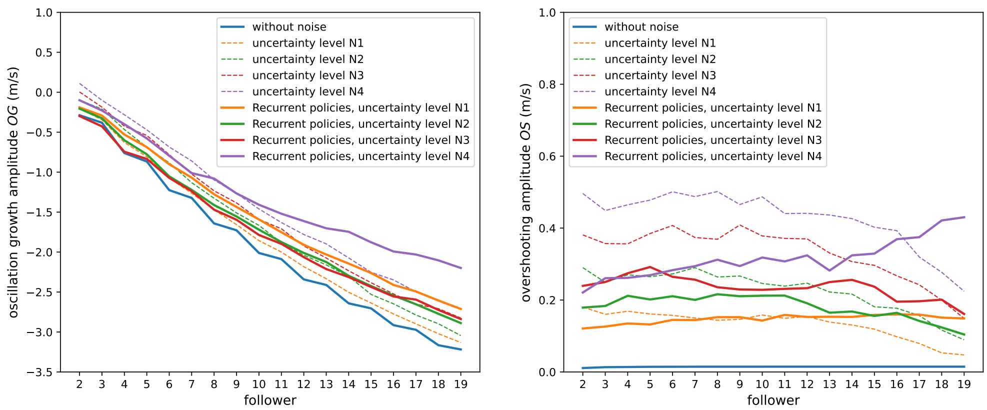


Figure 5.72: Oscillation growth amplitude (left) and overshooting amplitude (right) of the two-leader LSTM-ACC systems designed with the second weighting combination in scenarios with and without measurement noise

In general, Figure 5.72 shows similar trend in the *OG* curves. However, different from the situation in Figure 5.71, the *OG* values of the systems designed with the second weighting combination can be reduced significantly in every scenario with each level of measurement uncertainty. Although no significant improvement can be found from the *OG* curves, the overshooting effect is significantly mitigated when using the system designed with the second weighting combination compared to the scenarios without using any state estimator. This shows that the second weighting combination setup applied in the training of recurrent policies can still lead to a better string stability performance under every level of measurement uncertainty.

It is worth noting that the *OG* values of the last follower in measurement uncertainty level N4 in both Figures 5.71 and 5.72 are quite similar to those in both Figures 5.58 and 5.59, respectively. Therefore, it may be concluded that the performance of two-leader LSTM-ACC systems in scenarios with measurement uncertainty level N4 is at the same level with the one-leader ACC system in the scenario with accurate measurements. When the measurement uncertainty becomes too high, the two-leader LSTM-ACC systems may degrade to the one-leader ACC systems.

The ride comfort performance of the two-leader LSTM-ACC system designed with the first weighting combination is first presented by Figure 5.73. It is found that using the recurrent policies to follow the second leader can really reduce the number of large jerk amplitudes resulted from the noisy measurements. The cumulative distribution curves of the two-leader LSTM-ACC systems are pretty close to each other, indicating that there is no large difference between the ride comfort performance of the systems under every level of measurement uncertainty.

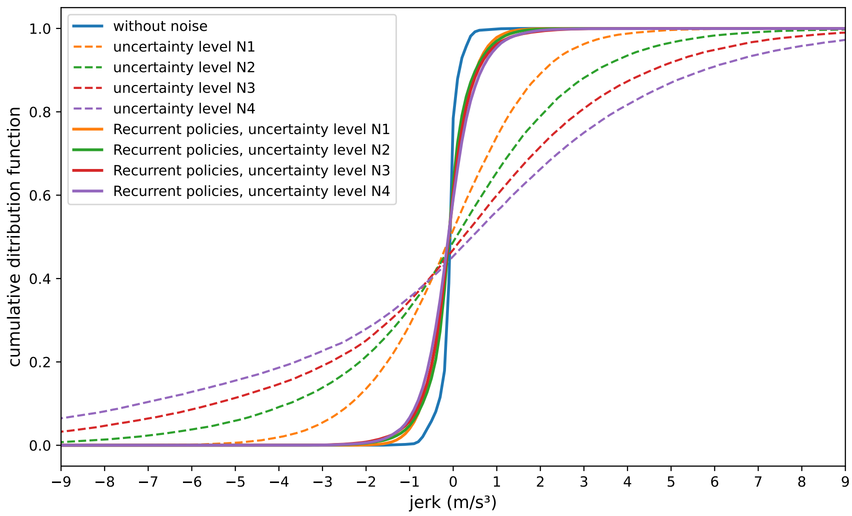


Figure 5.73: Cumulative distribution function of jerks experienced by platoons using the two-leader LSTM-ACC systems designed with the first weighting combination in scenarios with measurement uncertainty levels N1 - N4

Figure 5.74 then shows the cumulative distribution of jerks of the two-leader LSTM-ACC systems designed with the second weighting combination. Compared to the systems without using any state estimator, the amount of large jerk amplitudes can already be largely reduced by using these systems with recurrent policies. However, the curves deviate from the curve in the scenario with accurate measurements more than the situation in Figure 5.73, which is also the effect of different weighting setups.

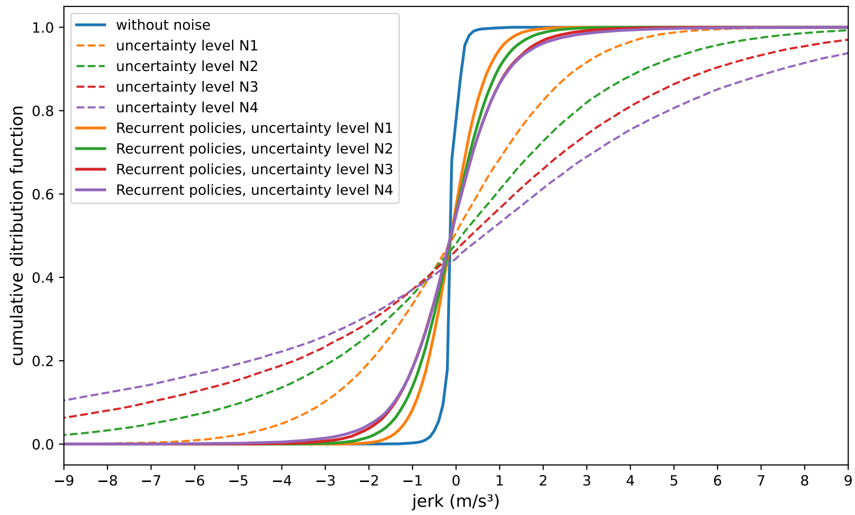


Figure 5.74: Cumulative distribution function of jerks experienced by platoons using the two-leader LSTM-ACC systems designed with the second weighting combination in scenarios with measurement uncertainty levels N1 - N4

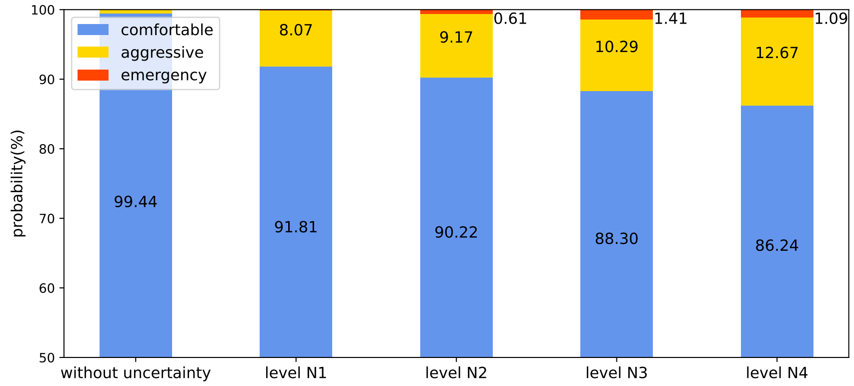


Figure 5.75: Probability distribution of jerks in each ride comfort level experienced by platoons using the two-leader LSTM-ACC systems designed with the first weighting combination in scenarios with and without measurement noise

Figures 5.75 and 5.76 again summarize the distributions of jerk amplitudes of the platoons using the two-leader LSTM-ACC systems in each ride comfort level defined. For those systems designed with the first weighting combination, nearly 90% of the driving maneuvers can stay within the comfortable level even when the measurement uncertainty level is high. Although there are still a certain percentage of aggressive driving maneuvers, only around 1% of the jerk amplitudes are in the range of emergency driving maneuvers. Compared to the KF-ACC systems shown in the previous section, the ride comfort performance of these systems is much closer to that of the level of the system in the scenario with accurate measurements. On the other hand, the systems designed with the second weighting combination show different results. At least 20% of the maneuvers are aggressive or emergency driving maneuvers under every level of measurement uncertainty. In addition, the ride comfort performance significantly degrades when the measurement uncertainty level increases. In the scenarios with measurement uncertainty levels N3 and N4, less than 70% of the jerk amplitudes can stay in the range of comfortable maneuvers.

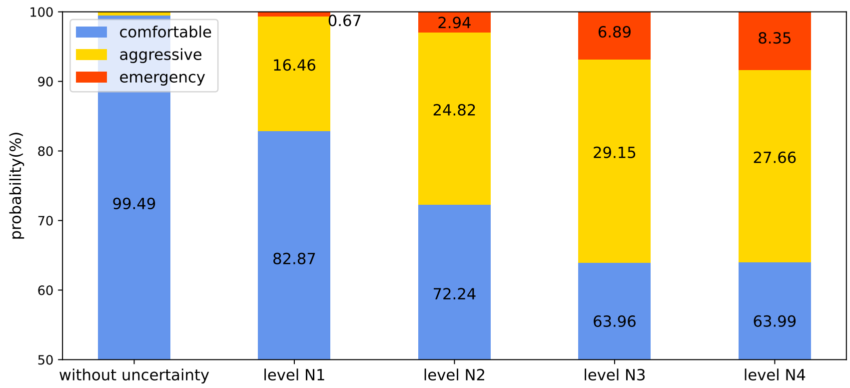


Figure 5.76: Probability distribution of jerks in each ride comfort level experienced by platoons using the two-leader LSTM-ACC systems designed with the second weighting combination in scenarios with and without measurement noise

The results in this section show that the two-leader LSTM-ACC systems can still produce better string stable performance than the one-leader LSTM-ACC systems through multi-anticipation. However, the string stability performance of these systems is not significantly improved compared to the systems without using any state estimator. In addition, the ability to damp out the disturbance would be clearly degraded if the measurement uncertainty level increases. The improvement of using the recurrent policies can, instead, be found by looking at the ride comfort performance. By using the recurrent policies, the systems produce better ride comfort performance than the systems using KFs. Different weighting combinations also show slightly different levels of improvement.

6

Discussions

This chapter first summarizes the evaluation results presented in chapter 5 and compares them with the hypothesis made in this study and the findings in previous studies. The limitations of both the findings and the research methodology of this study are also discussed.

6.1. Discussions on evaluation results

For the evaluation of the proposed multi-leader ACC systems, a 20-vehicle platoon is simulated. To analyze the string stability and ride comfort, which are the two main aspects of the system performance in this study, the performance indicators include oscillation growth amplitude, overshooting amplitude, and probability distribution of jerk amplitudes. Newell's car-following model is also applied as a comparison to help explain the longitudinal behavior of following vehicles equipped with the proposed systems. This chapter summarizes all the findings in chapter 5.

6.1.1. Effect of deep reinforcement learning and multi-anticipation

To examine the DRL performance in the training of ACC controller agents and understand the general effect of multi-anticipation, this subsection focuses on the performance evaluation results of ACC systems designed with accurate measurements.

From the results of ACC systems, it is found that the one-leader ACC system can already achieve a certain level of string stability performance when looking at the decreasing oscillation growth amplitude within the platoon despite that the string stability is not explicitly considered in the control system design. The outcome demonstrates the success of applying DRL to design ACC controllers which can preserve string stability unconditionally, which aligns with the results in Hart et al. (2021).

Compared to the one-leader systems, the two-leader ACC systems show even better string stability performance. The oscillation growth amplitude decreases faster when using the two-leader systems than it does when using the one-leader systems, indicating that the disturbance is damped out more significantly while propagating upstream. On the other hand, it can be observed from the *OS* curves that the overshooting effect between the acceleration and stabilization phases is also mitigated when the followers can look at two preceding vehicles.

In addition, the two-leader ACC systems also exhibit better ride comfort performance than the one-leader systems. The original aggressive driving maneuvers required by the vehicles equipped with one-leader systems to tackle the disturbance completely disappear when the following vehicles are equipped with the two-leader systems. The following vehicles in the platoon have smoother trajectories when they are equipped with the two-leader ACC systems. The improvement of ride comfort performance for the vehicle platoon brought by the ability of multi-anticipation was also pointed out by Wilmink et al. (2007) and Lee et al. (2021).

By inspecting the car-following behavior mechanism of those vehicles in many different ways, including looking at the acceleration/speed/gap profiles, position deviation with Newell's car-following model, and the plot showing the controlling leader at every time step, the multi-anticipation ability of vehicles equipped with the proposed two-leader ACC systems can be observed from their early slow-down behaviors when facing the traffic disturbance. Hence, the effect of looking further downstream for the following vehicles to improve the string stability and ride comfort of the car-following dynamics in the platoon is demonstrated. By comparing the evaluation results with other studies, similar characteristics regarding the multi-anticipation can be found. The early slow down behavior exhibited by the following vehicles equipped with the two-leader ACC systems was also pointed out in Wilmink et al. (2007). The improved string stability of the ACC system resulted from the multi-anticipation ability can also be found in the conclusion of Wang, Daamen, Hoogendoorn, et al. (2014a) and Donà et al. (2022). Therefore, it is found that the results of this study are indeed consistent with those in the previous studies which are also related to ACC/CACC systems with multi-anticipation.

6.1.2. Effect of measurement uncertainties and the handling methods

In the next part of the study, the effect of sensor measurement uncertainties is first investigated by including measurement noise into the simulation experiments. According to the evaluation results, it is found that measurement noise does degrade the string stability by slightly increasing the magnitude of the disturbance experienced by vehicles in the platoon, but the stability still remains as the oscillation growth decreases along the upstream of the platoon. On the other hand, although the overshooting amplitudes increase when the following vehicles are facing noisy measurements, the amplitudes are not amplified while the disturbance is propagating to the upstream. Therefore, it is found that both the oscillation and overshooting amplitudes increase when the measurement uncertainty level becomes higher, but the systems can still maintain a certain level of string stability performance when facing measurement noise without using any countermeasures. The result is slightly different from the original expectation made by the author at the beginning of this study. The random noise in the measurements does result in fluctuating behaviors. However, it turns out that these behaviors, on average, are still quite similar to the behaviors in the scenario with accurate measurements.

However, the random noise does severely increases the jerks experienced by the following vehicles, which leads to extremely uncomfortable and abnormal longitudinal driving maneuvers. This can be found by looking at the distributions of the jerks experienced by the following vehicles in the platoon.

To cope with the measurement uncertainties, this study first applies the practical and conventional Kalman filter (KF) as the state estimator to process the sensor measurements for the designed controllers. It is shown in the acceleration/speed/gap profiles that the KF successfully smooths out the influence of measurement noise. For string stability performance, it is shown in the *OG* curves that the oscillation growth amplitudes can be reduced and become closer to the level of scenarios with accurate measurements. On the other hand, the overshooting amplitudes of following vehicles in the platoon do not have such significant improvement. Regarding the ride comfort performance, the number of emergency actions are significantly reduced compared to the scenarios when facing measurement noise without using any state estimator. However, the jerk amplitudes cannot be completely reduced to the range of comfortable driving behavior. There is still a certain number of aggressive driving maneuvers in every measurement uncertainty level. The number of emergency behaviors which exceed the 2 m/s^3 threshold again increases in scenarios with relatively high levels of measurement uncertainties.

In addition to the conventional filtering approach, the LSTM-ACC systems which use recurrent policies as the controllers are proposed. These controllers are trained in a non-stationary partially observable environment with noisy measurements. They are expected to further reduce the jerk amplitudes resulted from noisy measurements by actively considering the jerk at every time step in the reward function. The simulation results show that several LSTM-ACC systems designed with the first weighting combination and relatively high measurement uncertainty levels still fail to achieve the desired string stability performance. Even though there are several LSTM-ACC systems which successfully preserve string stability for the platoon, the performance still deviates a lot from that of the scenarios with accurate measurements. It is also found that the string stability performance also degrades while the measurement uncertainty level increases. To comply with the spacing policy and reduce the jerk amplitudes experienced, those systems designed for scenarios with larger measurement noise tend to behave in a more conservative way when facing traffic disturbance. They decelerate more than the amount which is actually required and hence sacrifice the string stability performance. This kind of trade-off between noise robustness and string stability was mentioned as the future work by Donà et al. (2022). The results in this study further proves the existence of such trade-off.

When looking at the distribution of jerk amplitudes of the platoons using the LSTM-ACC systems, it is found that the number of aggressive and emergency driving maneuvers can be reduced significantly. This implies that the ride comfort performance becomes closer to level of the scenario with accurate measurements at every measurement uncertainty level. To sum up, although the proposed LSTM-ACC systems using recurrent policies cannot ensure ideal string stability in every setup and scenario, they manage to preserve ride comfort when facing noisy measurements.

The difference between KF-ACC systems and LSTM-ACC systems mainly lies in the way the erroneous information is handled. Although it is found that the LSTM-ACC systems are less subject to measurement noise compared to the KF-ACC systems, comparison between the performance of these two types of systems is not the key in this study. Different parameters used in the KF or even other filtering approach can also result in different string stability and ride comfort performance. This is something which is not explored in this study. The main finding after evaluating these systems is their potential and ability to preserve string stability and ride comfort even in scenarios with large measurement noise. With proper training and parameter setups, both kinds of system design can produce the desired performance.

6.1.3. Effect of different weighting combination setups

This study also aims to investigate the potential trade-off between string stability and ride comfort, which is a hypothesis made at the beginning of the study. However, the string stability is not explicitly consider in the control system design in this study. It is suspected that the gap-keeping performance also represents the string stability performance to a certain extent. Therefore, to help observe the trade-off, two different weighting combinations on the gap error and jerk terms are applied in the reward function of the DRL policy training for every ACC controller agent.

By looking at the performance of one-leader ACC systems, it is found that the system designed with the second weighting combination has better string stability performance than the system designed with the first weighting combination. The oscillation growth amplitude significantly decreases while the disturbance propagates to the upstream of the platoon. The two weighting combinations also lead to large difference between the overshooting effect of the two systems. The overshooting amplitude of the platoon using the system designed with the first weighting combination increases when the disturbance propagates to the upstream, while those of the platoon using the system designed with the second weighting combination are small enough to be ignored. This can already demonstrate that the gap-keeping performance can indeed represent string stability to a certain extent for the control system design.

For ride comfort performance, it is hence expected that the system designed with the first weighting combination may have better performance. In the end, the acceleration and speed profiles do present that the first following vehicle in the platoon seems to have larger and more aggressive deceleration reactions to the disturbance. However, according to the distribution of jerks experienced by the whole platoons, the system designed with the second weighting combination turns out to have slightly better ride comfort performance. It is speculated that the improved gap-keeping and stability performance damps out the disturbance faster than before, which also alleviates the required driving maneuvers for the following vehicles to handle the disturbance. Therefore, the system designed with the second weighting combination can still bring better ride comfort performance collectively. The same situation happens for the two-leader ACC systems as well. The trade-off cannot be shown when simply comparing the systems in the scenario with accurate measurements. Instead, the different weighting setups only influence the local stability of the first follower and its ride comfort performance.

When measurement noise is considered in simulation experiments, the effect of different weighting combinations can be observed. Although the systems designed with the second weighting combination still have better string stability performance compared to those designed with the first weighting combination, the ride comfort performance is significantly degraded. More driving maneuvers with large jerk amplitudes emerge, and the number of emergency maneuvers also significantly increases when the followers are equipped with the systems designed with the second weighting combination. This indicates that the policies trained with the second weighting combination tend to produce more jerks especially when the measurements are noisy, which cannot be shown in the scenario with accurate measurements. After implementing the KFs into the systems, the similar effect can still be found. The KF-ACC systems which use controllers designed with the second weighting combination have worse ride comfort performance.

The trade-off between string stability and ride comfort is even more significant when looking at the performance of the systems using recurrent policies. It is found that the one-leader LSTM-ACC system fails to preserve string stability in the platoon when it is designed with the first weighting combination as the oscillation amplitude increases at the end of the platoon due to an additionally induced disturbance, and the overshooting amplitude increases dramatically while the disturbance propagates. The problem is fixed when the system is instead designed with the second weighting combination. Although the overall string stability performance is not improved compared to the system without using any state estimators, the oscillation amplitude decreases, and the overshooting amplitude can be kept at a uniform level when the one-leader LSTM-ACC system is designed with the second weighting combination. Same situation exists for the two-leader LSTM-ACC systems. The systems designed with the first weighting combination do not show desired string stability, especially in scenarios with large measurement noise. When the systems are designed with the second weighting combination, the oscillation growth amplitude decreases, and the overshooting effect is not significant.

For the ride comfort performance of LSTM-ACC systems, it is found that systems designed with the first weighting combination have almost equally ideal level of performance. Around 99% of the jerk amplitudes stay within the 2 m/s^3 threshold in every measurement uncertainty level. When using the systems designed with the second weighting combination, the number of emergency driving maneuvers increases while the measurement uncertainty level becomes higher. This indicates that not only do the systems have worse ride comfort performance, but they are more sensitive to the influence of measurement noise.

To conclude, the scenario with accurate measurements can only present the contradiction between the local stability and ride comfort of the first following vehicle, which is complied with the findings in Shladover (1978) to a certain extent. When looking at the platoon as a whole, the trade-off between string stability and ride comfort by applying the different weighting combinations is shown when the measurement noise is considered in the simulation. Such phenomenon was indeed mentioned by Yamamura et al. (2008). The existence of measurement noise highlights the trade-off between string stability and ride comfort.

6.2. Research limitations

This section aims to discuss the applicability of the proposed ACC systems in other scenarios and details in the autonomous driving environment which are not considered in this study.

6.2.1. Applicability in other speed ranges

The controllers in the ACC systems are trained in a pre-determined speed range between 15 m/s and 35 m/s which is approximately the speed range of uncongested motorway traffic. This also defines the operational design domain of the proposed systems. However, it is believed by the author that the systems may still work in a slightly larger speed range since the trained control policies are all DNNs which basically approximate the parametric car-following models. Parametric car-following models and the control policies both seek to map the distance and speed information in the driving environment either perceived by a human driver or collected from a sensor to an output acceleration value. Such mathematical similarity was pointed out by H. Zhou et al. (2022). The defined speed range in this study is already large enough to create the variability to design controllers (DNNs) which possess sufficient generalizability. As long as the speed of either the leader or the follower would not reach the extreme situations, e.g., a complete standstill, the systems should be able to be applied. More testings are required to verify the above-mentioned hypothesis regarding the generalizability of the system.

6.2.2. Applicability in other traffic disturbances

The applicability of the proposed systems when facing other types of traffic disturbance can also be discussed. In real-world driving scenarios, a leader behavior which contains many disturbances with a higher oscillation frequency or many disturbances with more irregular fluctuation amplitudes may occur. When facing this kind of complicated leader behavior, the systems using controllers designed with accurate measurements in a stationary environment are believed to be more applicable than the systems using controllers designed with uncertain measurements. The policies trained with accurate measurements in a stationary environment are designed in a way that the controller agent make decisions based on the assumption that its leader maintains a constant speed at every time step. Although the assumption is far from reality, the policy is trained to be more deterministic for every possible system state and hence can be used in all kinds of conditions. On the other hand, the training of recurrent policies for LSTM-ACC systems is carried out in a non-stationary environment which uses an LSTM network as a future state estimator. The performance of such a controller heavily depends on the training task specified. In this study, the training tasks are manually designed and randomly generated during the training process instead of using real vehicle trajectory data. This method ensures the behavior of the trained policy follow the desired theoretical setting and seems to prevent the policy from generating any unexpected control actions. However, similar to the problem of data hungriness for supervised and unsupervised learning, if the training task setup fails to provide sufficient scenarios which the ACC controllers have to tackle, they may not be able to guarantee optimal performance in every car-following task encountered. The generalizability of the trained policies may be questioned if the training tasks are not well designed. This problem can be more serious when the learning is carried out in a non-stationary environment than it is in a stationary environment. Therefore, it is speculated that the recurrent policies in this study are more limited because they are only trained in scenarios with a single disturbance event in the time series. In the future work, the training of recurrent policies using leader behavior collected from real-world data should be considered.

6.2.3. Applicability in other types of measurement uncertainties

Another aspect which can be considered when discussing the applicability of the systems in this study is the types of measurement uncertainties. Measurement noise following Gaussian distributions is the only source of uncertainties considered in the system design in this study. However, other types of uncertainties, such as false positive alarms or false negative signals (losing detection) may also occur in the autonomous driving environment. In addition, the measurement noise can also contain a certain extent of time- or spatial-dependency, which is more difficult to be handled compared to the white noise. As mentioned previously, these characteristics are not included in this study due to the availability of empirical or field experimental data. The KF-ACC systems apply KFs for the state estimation task due to the assumption of Gaussian noise. The LSTM-ACC systems are also designed by training the controller agents in scenarios with Gaussian noise. Therefore, these two types of proposed system designs will certainly fail in scenarios with different types of uncertainties. In future work, this problem can be addressed by considering other filtering methods, such as particle filter, or including other sources of measurement uncertainties in the numerical simulation so that the trained controller agents with recurrent policies possess the ability to handle the other types of uncertainties. By doing so, the results can also become more persuasive for practitioners' interest.

6.2.4. Limitations in the numerical simulation environment

Apart from the applicability of the systems, there are several details regarding the modelling of the driving environment which are not considered in this study. This implies that the simulation approach adopted in this study does not precisely reproduce the vehicle dynamics in reality, which may affect the quality of the research findings. As discussed in section 1.4, the scope of the study does not consider the operation of lower-level ACC controller, internal driveline, and environmental factors. Most studies related to ACC systems in the traffic domain focused on the design of upper-level controller. The possible reason is that the discussion of lower-level controller involves the hardware design of commercial vehicles which is mainly manipulated by car manufacturers themselves (H. Zhou et al., 2021). The modelling and simulation of the lower-level controller is relatively difficult than simply simulating the upper-level planner which generates the acceleration command. For instance, factors which need to be considered in the simulation of lower-level ACC controller include the control gains in the control systems of these actuators to precisely describe how the vehicle speed and acceleration are updated to reach the desired motion state determined at the upper-level. On the other hand, there are also many environmental factors, such as aerodynamic drag forces, rolling resistance, and road slope, which can hardly be considered unless conducting field experiments. Given the complexity of the problem, these aspects are excluded in this study. Therefore, the simulation and analysis results in this study only provides a theoretical understanding of the performance of the proposed ACC system designs based on the scope defined and assumptions made at the beginning of the study.

7

Conclusions and Recommendations

Before concluding the study, this chapter again briefly summarizes the findings from the evaluation results and answers those research questions proposed in chapter 1. After the conclusions are made, section 7.4 seeks to reflect on the pros and cons of the selected control system design method from a scientific perspective and also provide several implications from a practical perspective. At the end, recommendations for future research regarding the design of multi-leader ACC systems are provided.

7.1. Research findings

According to the evaluation results, several important research findings are drawn in this section.

In the first part of the evaluation, it is discovered that DRL successfully trains control policies which can ensure string stability and maintain the ride comfort performance for the proposed ACC systems with proper training and parameter setups. It is worth noting that this is achieved without considering string stability in the system design explicitly. The multi-anticipation ability brought by the two-leader ACC systems benefits both string stability and ride comfort performance for the vehicle platoon. The improved performance is resulted from the early slow down behaviors which are exhibited when the following vehicles are equipped with the two-leader system. Such behavior prevents the vehicle from having to conduct a large deceleration or excessive acceleration.

The next part of the evaluation shows that both string stability and ride comfort performances of the ACC systems are degraded when the measurement noise is considered. The degradation condition becomes even worse when the uncertainty level increases. In addition, the trade-off between string stability and ride comfort can be clearly demonstrated from the results of systems designed with different weighting combinations.

To handle the measurement uncertainties, the effects of implementing KFs and using recurrent policies are presented in the third and fourth parts of the evaluation, respectively. Still, the higher measurement uncertainty level leads to a worse string stability and ride comfort performance of the systems. The trade-off between noise robustness and string stability is also demonstrated by looking at the performance of two-leader LSTM-ACC systems. However, both types of ACC system design show their abilities and potentials to handle noisy measurements. The performance limits of the systems highly depend on their internal parameter settings.

7.2. Discussions on research questions

The findings regarding the performance of the proposed ACC systems from the evaluation results are already summarized in the previous chapter and the previous section. This section then seeks to answer the research questions proposed in section 1.3.

- ***What kind of control method for ACC systems has the potential to outperform other types of controllers in terms of string stability and the handling of measurement uncertainties? How to design the ACC controllers using the selected control method? Which factors can and should be considered in the control system design?***

Although not being investigated in this study, a few studies using typical PD-like ACC controllers in the past have already shown a certain level of ideal performance in terms of string stability (Wang et al., 2018; Wang et al., 2017). This study applies DRL, an intelligent control approach, to design the ACC controllers instead of choosing other methods since it is believed that DRL possesses a great potential for non-linear control task using DNN and the handling of uncertainty than other control methods do. More relevant discussion regarding the choice of controller design method can be found in section 7.4. DRL learns a policy network by allowing the DNN agent to explore the environment through trial-and-errors. The state of the ACC controller agent consists of distance gap with the leader, speed, relative speed with the leader, and jerk amplitude. In the DRL framework, the time gap error and jerk amplitude at every time step are considered in the reward function design. The difficulty of including string stability in the controller design will also be elaborated in section 7.4.

In this study, the controllers without considering sensor measurement uncertainties are first designed by training with accurate measurements and randomly generated car-following tasks in which the leader keeps a constant speed. In the next part, the LSTM network is implemented into the agent to design controllers which have the state estimation ability on its own. The training of these LSTM-ACC controller agents with recurrent policies is carried out in non-stationary traffic disturbance cases which consider measurement noise.

The ability of DRL can first be discovered from the results of the one-leader ACC systems, which are developed through the training with accurate measurements. The decreasing oscillation growth amplitude and insignificant overshooting effect in the platoon demonstrate that the controller agents trained by the DRL setup can indeed achieve a certain extent of string stability.

On the other hand, the ability of DRL in the handling of measurement uncertainties can be discovered by looking at the performance of the LSTM-ACC systems which use recurrent policies. When inspecting the evaluation results of LSTM-ACC systems in chapter 5, one can see that the string stability can still be preserved albeit with degraded performance particularly in scenarios with a relatively high level of measurement uncertainties. The jerks produced by the following vehicles, however, become smaller because of the ability of the controller agents to react by actively considering the noise. Hence, in scenarios with a relatively small level of uncertainties, the controllers designed using the learning-based control method manage to preserve string stability and ensure a certain level of ride comfort in the mean time.

- ***What is the proper way to model and simulate the measurement uncertainties so that the real-world autonomous driving conditions can be reproduced? When it is known that the sensor measurements are erroneous, what kind of methods can be used by the ACC systems to cope with the uncertainties?***

In the defined problem context, two sources of measurement uncertainties, measurement noise and false negative detection, can be modelled. For simplicity in this study, only the measurement noise is considered.

In order to evaluate the proposed systems in an environment which is as close to the real-world autonomous driving conditions as possible, the noise error should ideally follow a pattern extracted from real data or possess a certain level of time-dependency to reflect the influence in certain environmental conditions. However, due to the data availability of the multi-leader detection task, which is a comparatively novel technique, and complexity of the problem, this study does not take any real-world pattern or time-dependency into account. The measurement noise follows zero-mean Gaussian distributions with specified standard deviations. This is also the most commonly adopted method to model measurement noise in the literature related to ACC system design and evaluation.

To handle measurement uncertainties for a dynamical control system, a filtering method is commonly applied to estimate the true state for the controller. In this study, a linear KF can be used since the measurement noise follows Gaussian distributions. The system state can be estimated by the prediction based on the known process dynamics and the collected real-time measurements. The filtered measurements would then be used by the controller as the estimated state information. This method is, hence, adopted by the KF-ACC systems in this study.

In this study, the ACC systems are first designed by DRL. Another method to handle the measurement uncertainties is to extend the original control problem based on MDP to a problem based on a POMDP. In this kind of problem, the agent can properly react only based on the erroneous or incomplete information in the environment. The problem can then be solved by DRL with RNNs as the state estimator in the policy. Therefore, this study proposes LSTM-ACC systems, which have an LSTM network layer before the original actor network in their controller agents. By doing so, the agent can utilize the history to estimate the true underlying state at every time instant and generate control actions accordingly.

- ***What is the criteria of string stability in this study? Which indicators can be used to appropriately evaluate the string stability performance of the proposed ACC systems?***

String stability is defined as the ability of the vehicle platoon to maintain or even dissipate the disturbance caused by downstream car-following dynamics. The propagation of disturbance can be investigated by looking at either the acceleration, speed, or gap error profiles of the following vehicles. To simplify the discussion in this study, the influence of the disturbance on vehicle speed is the major focus. Two indicators, oscillation growth amplitude and overshooting amplitude, are analyzed following the calculation method introduced in subsection 4.4.1. If the two indicators is decreasing or maintaining at the same level while propagating to the upstream of the platoon, the platoon is considered to be string stable.

- ***While aiming at preserving string stability, can the system still maintain a certain level of ride comfort for the platoon? How to quantitatively analyze the ride comfort performance of the proposed systems?***

In the literature, ride comfort is usually discussed by investigating the vehicle acceleration or jerk, which is the rate of acceleration changes. In this study, the jerk is selected since a large vehicle acceleration or deceleration is not completely avoidable when encountering a traffic disturbance.

Traffic stability (including local, string, and traffic flow stability) and the jerk of vehicle motion are known as two potentially contradicting factors in car-following dynamics. It is hence expected that the jerk amplitudes may become larger when the system seeks to ensure string stability for the platoon, and vice versa. Maintaining a certain level of ride comfort when seeking to preserve string stability for ACC systems is one of the main research objectives in this study. Therefore, two different weighting combination setups are applied in the reward function to investigate if the hypothetical trade-off can be observed and if the system can achieve a balance between these two factors.

To quantitatively analyze ride comfort, the cumulative probability distributions of jerk amplitudes experienced by every platoon using the proposed ACC systems are then generated. According to the distributions of jerk amplitudes, the influence of the measurement uncertainties and the measurement uncertainty handling methods on the ride comfort can be understood. In addition, the threshold values for jerk amplitudes are determined to define the range of comfortable, aggressive, and emergency driving maneuvers. By looking at the distribution of maneuvers in these ranges, the ride comfort performance of different systems in different scenarios can be compared.

Following this framework, the ride comfort performance of the proposed ACC systems, KF-ACC systems, and LSTM-ACC systems in different scenarios is analyzed. It is found that the ACC systems can ensure string stability and maintain an ideal level of ride comfort performance in the mean time when the measurements are accurate. Later on, the simulation results show that the measurement noise considered in several scenarios in the simulation experiment significantly affects the ride comfort performance of the proposed ACC systems. In addition, the trade-off between string stability and ride comfort is shown when the measurement noise is considered. The systems designed with the second weighting combination have better gap-keeping and string stability performance than the systems designed with the first weighting combination, but they also lead to degraded ride comfort.

For the handling of measurement noise, KFs can significantly reduce the number of large jerk amplitudes experienced by the following vehicles. However, there is still a certain number of aggressive and emergency maneuvers when measurement uncertainty level is high. It is believed that different parameter settings in the filter can produce different performance results. On the other hand, the LSTM-ACC systems have better ride comfort performance at every level of measurement uncertainties than the KF-ACC systems. However, it is indeed found that the string stability performance is slightly impacted when using this kind of system. The potential of the LSTM-ACC systems is still yet to be explored.

- ***What is the benefit of multi-anticipation for the ACC platoon? How to explore the positive effect of the proposed multi-leader ACC system compared to the one-leader system?***

The benefit of multi-anticipation can first be observed by directly looking at the acceleration, speed, and time gap profiles. If the amplitudes of maximum acceleration, minimum deceleration, speed deviations, and gap errors of the two-leader system become smaller than those of the one-leader system, it implies the positive string stability effect of enabling the followers to look at their second leader. More importantly, the string stability indicators, oscillation growth and overshooting amplitudes, help provide a quantitative way to evaluate the string stability. It is expected that these two indicators would decrease faster in the upstream direction of the platoon when using the proposed two-leader ACC systems. Furthermore, vehicle trajectories with speed contours can also provide a clear view of the propagation of disturbance. Newell's car-following model serves as a basis to help understand the mechanism behind the behavior of the two-leader systems.

First, compared to the one-leader ACC system, the acceleration/speed/gap profiles of the two-leader ACC system show higher minimum deceleration values, faster decrease of the oscillation growth amplitudes, mitigated overshooting amplitudes, and smaller time gap errors. These all demonstrate the positive effect of looking at more than one leader further downstream. The speed contours also show that the oscillation amplitude is significantly damped out, and the overshooting effect can hardly be observed. By comparing the vehicle trajectories and trajectories generated from Newell's car-following model, one can see the early slow-down behavior stemmed from the ability of multi-anticipation indicated by the positive position deviation. The early slow-down behavior allows vehicles to have smaller deceleration response to the preceding traffic disturbance than their first and direct leader.

The two-leader system also brings positive effect on the ride comfort performance. By comparing the probability distributions of jerk amplitudes, it is found that the two-leader system can avoid aggressive driving maneuvers even when facing the traffic oscillation event. Therefore, the multi-anticipation ability can also help preserve the ride comfort the following vehicles equipped with the proposed multi-leader ACC systems.

- ***What is the influence of the considered measurement uncertainties on the ACC system performance? What is the measurement uncertainty boundary for the proposed systems within which the desired performance in terms of string stability and ride comfort can still be preserved?***

This study tests one level of measurement noise on the proposed one-leader systems and four levels of noise on the proposed two-leader systems. To understand the impact of measurement noise, the proposed ACC systems without using any state estimator are first simulated. It is found that the random noise does not cause string instability although the oscillation growth and overshooting amplitudes slightly increase. However, the jerk amplitudes become significantly larger and exceed the threshold of comfortable driving maneuvers easily when considering measurement noise in the simulation experiment.

With the help of the filters, the KF-ACC systems can achieve the level of string stability performance in every measurement uncertainty level which is similar to that in the scenario with accurate measurements. The ride comfort is also improved since there are fewer jerk amplitudes exceeding the determined comfortable driving maneuver threshold. Still, when the uncertainty level becomes higher, the number of emergency driving maneuvers still increases significantly.

The LSTM-ACC systems actively consider the measurement noise while making control actions to take the ride comfort into account. With proper training setup, these systems using recurrent policies successfully produce string stable and comfortable platoon performance. The ride comfort can be better guaranteed at every measurement uncertainty level, but the overshooting effect may still become quite severe in uncertainty levels N3 and N4.

According to the evaluation results of KF-ACC systems and LSTM-ACC systems, it is found that, in general, both types of system design can preserve string stability and ensure ride comfort in scenarios in which the standard deviations of the measurement noise can be kept within 1 m and 1 m/s for the range (distance gap) and range rate (speed) measurements, respectively. Larger measurement noise makes it difficult for the systems to preserve string stability and maintain the desired ride comfort performance. However, such result may not be valid for other system designs with different training or parameter settings. When different types of measurement uncertainties are considered in the simulation experiment, the deduced theoretical boundary of measurement noise may even become more limited. This study explores the effect of a filtering method and a recurrent DRL policy design method for ACC systems to handle measurement uncertainties and provide recommendations for possible future extension of the system design, as will be discussed in section 7.5.

7.3. Conclusions

From a traffic engineering standpoint, string stability has long been a problem for commercial ACC systems. Small disturbance caused by a leading vehicle can propagate to the upstream of the traffic as a shockwave. To enhance the string stability of a vehicle platoon, this study seeks to propose several ACC system designs by leveraging the power of an advanced on-board RADAR instead of inter-vehicle communication technologies due to their potential challenges in practice. Multiple systems are proposed, while each of them has different setups regarding the multi-anticipation ability, the handling of sensor measurement uncertainties, and the weightings between the gap-keeping and ride comfort performance.

DRL is applied to design the controllers in the ACC system architecture.

Systems which are designed with accurate measurements can demonstrate the ability of DRL and the positive effect of multi-anticipation. Therefore, the performance of these systems is first evaluated and serves as a benchmark for the remaining scenarios with measurement uncertainties considered. Later on, the system performance are evaluated in scenarios with measurement noise considered to explore the influence of the erroneous information. Two methods are adopted to cope with noisy measurements. First, a tuned conventional KF, which works as an external state estimator to smooth out the measurement noise, is applied. In the second approach, the DRL-based ACC controllers are modified by adding an LSTM layer into the DNN agent so as to formulate recurrent policies. These controllers are trained in scenarios with different levels of measurement noise. The LSTM network is expected to serve as an internal state estimator and predict the future leader behavior in a traffic disturbance.

The evaluation of system performance focuses on two aspects, the string stability and ride comfort. A numerical simulation approach is adopted to test a platoon of vehicles equipped with the proposed ACC systems in a traffic disturbance. Other than evaluating the system performance by investigating the indicators developed in the quantitative analysis framework, the car-following mechanism is also understood by looking at the acceleration/speed/gap profiles, vehicle trajectories with speed contour, and comparison with trajectories generated from Newell's car-following model.

The evaluation results show the ideal ability of DRL in designing controllers for the proposed ACC systems and the benefit of multi-anticipation ability for the overall platoon performance. Hence, the importance of developing an ACC system which can detect and react to the preceding car-following dynamics of multiple downstream leaders is revealed.

To reflect the real-world autonomous driving conditions, uncertainties lying in the sensor measurements should be included in the simulation. Therefore, measurement noise is considered in this study. After exploring the influence of measurement noise on the system performance, the measurement noise is also considered in the control system design to propose systems which can possess the ability to handle measurement noise. First, the KF-ACC systems which apply KFs as the state estimator to filter the noisy measurements are proposed. LSTM-ACC systems which use a recurrent policy containing an LSTM network as the controller are the second measurement handling approach in this study. With appropriate setups, it is found that both systems can preserve string stability and ride comfort to a certain extent in several scenarios. In general, it is observed that an ideal platoon performance can be guaranteed if, for more than 68% of the time, the noise of the second leader range and range rate measurements collected from the RADAR can be kept within [-1 m, 1 m] and [-1 m/s, 1 m/s], respectively. Still, different settings regarding the DRL training and parameter tuning should be conducted to further optimize the system performance and carefully explore its limit.

This study provides a theoretical understanding of the car-following behavior and performance of the proposed ACC systems in terms of string stability and ride comfort through experimental simulations. The performance limit of these system designs with regard to the handling of measurement uncertainties is also explored. It is one of the earliest attempts to discuss the string stability performance of learning-based ACC systems and the benefit of implementing multi-anticipation ability on ACC systems. To the best of the author's knowledge, this is the second research work which considers sensor measurement uncertainties in both the design and evaluation of multi-leader ACC systems, as Donà et al. (2022) being the first attempt. The results of this study are expected to motivate the development of advanced sensor technologies and ACC systems with the ability of multi-anticipation and measurement uncertainty handling in the future.

7.4. Implications

This section discusses how the research findings are important to the researchers in this field of study and practitioners in the automotive industry.

7.4.1. Scientific implications

In this subsection, the strength and potential of the methodology adopted to design the proposed ACC systems in this study are reflected by comparing it with two other alternative methods. How they can contribute to the future research is particularly important for the conclusion in this subsection.

In this study, string stability and measurement uncertainty handling are the two focal points. In a control theoretical approach, a transfer function can be used to describe string stability in the frequency domain, as mentioned in section 2.2. However, when a scenario with complicated measurement uncertainties are considered, the analytical method may not be able to easily describe the string stability. Instead, a simulation-based or optimization-based approach should be applied to design ACC controllers in this situation.

To design ACC controllers considering measurement uncertainties, there are three possible control methods which can be adopted, including the optimization of a parametric model, MPC, and the DRL-based method adopted in this study. The first method requires a hypothetical parametric car-following model being proposed beforehand. A multi-objective optimization approach can then be applied to determine the parameters which can generate optimal performance in terms of string stability and ride comfort. String stability and ride comfort can be expressed by using representative performance indicators in the objective function. Genetic algorithm is one of the most commonly-applied heuristic optimization methods for this kind of simulation-based optimization problem. However, the parametric model which needs to be formulated based on expert knowledge may not be able to capture all the non-linearity in the control task. The performance would be limited due to the formulated parametric model. Therefore, an MPC or a DRL-based method which can consider the non-linearity in the ACC control task is preferred.

Decision-making under uncertainty is a popular topic in the research community nowadays. A few studies have already proposed MPC-based ACC controllers to deal with uncertainties resulted from erroneous sensor measurements, sensor or actuator delay, and actuator behaviors, as mentioned in subsection 2.1.2. They tried to utilize the prediction and optimization models in the MPC framework to limit the influence of the uncertainties. Different from these approaches, the LSTM-ACC controllers designed in this study apply a learning-based method for both the state estimation/prediction and control task. Instead of designing or formulating a complex model for the system dynamics beforehand like an MPC approach, this kind of method only relies on a careful training setup. Although it has been mostly applied in the area of gaming and robotics until now, several studies have started to investigate the possibility of using DRL based on POMDPs for the control of autonomous driving tasks, as mentioned in subsection 2.1.3.

The other commonly stated advantage of DRL over MPC is its computation requirement. It is known that the direct mapping of a offline-trained DRL controller is significantly faster than the online optimization process in an MPC-based controller. This was one of the main reasons why DRL was adopted in those ACC-related studies. However, it is also known that an MPC can be replaced by a DNN function approximator through the so-called imitation learning approach. By transforming an MPC into a supervised learning problem to approximate the behavior of the controller, the execution of an MPC controller can become significantly faster. Piecewise affine system is also a commonly-used method to approximate the nonlinear optimal controller. Besides, the computation power is growing rapidly nowadays. Therefore, with these alternative ways to reduce the computation requirement, the argument that DRL is more computationally efficient than MPC may not entirely hold. Still, it is worth noted that the efficacy of replacing an MPC controller by these methods may be questioned since its performance may not be as good as using the original online optimization approach.

However, either MPC or DRL has a drawback in this study. Although string stability is the major focus in this study, it is not explicitly considered in the reward function design although the goal of minimizing gap error in the reward function does, to a certain extent, ask the agent to prevent the propagation of disturbance. The difficulty of considering string stability explicitly lies in the sequential decision-making nature of RL problems. The string stability cannot be easily described in the time domain (every single time step in the training episode), as pointed out by Mirwald et al. (2021). There are studies using DRL to design CACC controllers which considers string stability in the reward function design since the ego-vehicle receives the acceleration of the leader via vehicle communication technologies. With the known immediate acceleration command of the leader, the ego-vehicle can learn to limit its acceleration response to the preceding car-following dynamics when there is a disturbance. There are also some studies which aim to achieve cooperative behaviors through centralized control method. In this kind of problem context, the string stability can be ideally considered in the RL reward function. However, this is not the case in this study since the inter-vehicle communication does not exist in the defined problem context.

Even though there is such a drawback in the control system design, the evaluation results in this study show that DRL can still be applied to design ACC controllers which can ensure string stability. The different weighting combinations in the policy training setup also highlight the trade-off between string stability and ride comfort and show flexibility in the control system design. For the handling of measurement uncertainties, an LSTM network is directly connected with the actor-critic network structure of the LSTM-ACC systems. Although it is difficult to particularly examine the state estimation performance of the LSTM network, it is believed to be able to surpass KF especially when the assumption of Gaussian noise distribution is violated. However, it is also worth mentioning that there are other filtering approaches, such as the particle filter, which get rid of the linearity and Gaussian limitation. Using these types of filters as the state estimators can also be considered when the measurement errors are modelled differently. Still, as discussed in the previous section, if the measurement uncertainties follow a certain extent of time- or spatial-dependency, the LSTM network may be able to perform better than filtering approaches by directly learning the pattern of the uncertainty development throughout the time series.

According to the findings in this study, the ability and suitability of applying DRL to design ACC controllers which can ensure string stability and handle noisy measurements are demonstrated. The influence of measurement noise on the system performance is also shown by the simulation experiment. This study aims to highlight the importance of considering more types of measurement uncertainties so that the real-world autonomous driving condition can be reproduced. In addition, the method still contains a lot of potential in terms of uncertainty handling. Future studies can continue to propose advanced control system design to address those problems which have not been considered in this study, as will be discussed in section 7.5.

7.4.2. Practical implications

This subsection first focuses on the development of multi-leader detection functionality using various on-board sensors. The applicability of the proposed DRL-based ACC systems in practice is then discussed in the second part of this subsection.

From the results of the performance of the two-leader ACC systems in this study, it is found that a large level of measurement uncertainties may induce negative impacts on the system performance in terms of both string stability and ride comfort, which makes it even worse than the one-leader system. Therefore, it is suggested that the second leader measurements should only be used when its accuracy can be guaranteed at a certain level, or other possible methods to utilize the second leader measurements should be applied to prevent the dramatic influence of measurement uncertainties, which will be further discussed in the next section.

The detection of the second leader has not yet been widely adopted by commercial vehicles nowadays. It is known that the advanced RADAR sensor implemented on Tesla Autopilot v8.0 possesses the ability to see two vehicles ahead (Donà et al., 2022). However, the discussion regarding this functionality mainly lies in its positive improvement on safety. It is only known that the vehicle equipped with the Autopilot can detect heavy braking of the leaders further ahead and conduct emergency response even before the action of the direct leader. To which extent the functionality contributes to the traffic stability performance of the ACC system has not been emphasized.

While trying to implement this kind of multi-leader detection functionality for ACC systems to enhance string stability, a consideration can be raised. When the second leader is quite far away from the ego-vehicle in a high-speed driving scenario on a multi-lane motorway, it may be difficult for the RADAR sensor to distinguish the second leader from vehicles driving on other lanes. The same problem may also occur for the multi-leader detection task at curves or other road sections with complicated geometry, which is also a difficulty for one-leader ACC systems to operate nowadays. Therefore, it is suggested that the point cloud data collected from the on-board RADAR sensor should also be fused with vision-based sensors, such as a camera or LiDAR, to verify the presence and measurements of the second leader by looking at the lane markings and filtering out echo signals from other vehicles. By doing so, the feasibility and operation of the multi-leader ACC system can be better ensured.

There are also several manufacturers using camera or LiDAR to measure the first leader measurements for the ACC system, as introduced in Lee et al. (2021). It is also believed that when the second leader is only partially occluded, the cameras should be able to see and perceive the movement of the second leader. The other possibility for this kind of vision-based sensor setup to conduct multi-leader detection task is using the cameras to detect the second leader by looking through the window of the first leader. By applying the advanced image recognition techniques, the cameras can possibly observe whether the second leader is getting closer or farther away from the first leader, which can also help accomplish a certain level of multi-anticipative car-following behavior. Still, this type of sensor setup suffers from the problem of low reliability and unknown ability to collect measurements in a long range. How to ensure a reliable and accurate detection of the second leader through proper sensor fusion techniques can also be the major development direction for the industry in favor of the multi-leader ACC systems.

Regardless of the sensor technology for the multi-leader ACC systems, this study applies DRL to design the controllers. DRL can be regarded as an intelligent control method which has not yet been widely applied to solve any real-world problems despite its popularity in the research community. One of the major concern lies in the explainability of such a black box approach.

Despite this great challenge, a lot of studies and research endeavors have already been investigating the possibility of applying DRL on various self-driving challenges, including longitudinal speed control, lateral lane-changing maneuver, and the overall vehicle motion and trajectory planning. The ACC controller design and control logic of commercial vehicles nowadays are classified information for automakers. It is still unclear when DRL-based ACC systems will enter the market.

Every detail of the control system design in this study, including the DNN structures of the controller agents, state and observation space, reward function design, and training task are carefully considered to prevent unexplainable outcome. Even if the controller design method may not be of practical interest in the near future, the findings and conclusion regarding the effect of multi-anticipation ability and measurement uncertainty boundary drawn from this study can still be applied to the design of multi-leader ACC systems using other control strategies or uncertainty handling methods.

7.5. Recommendations for future work

This study proposes several multi-leader ACC system designs. The limitations of the proposed systems discussed in section 6.2, including the applicability to other types of disturbance events and measurement uncertainties, are all possible and interesting aspects to tackle for future research. For instance, it is mentioned that real vehicle trajectory data should be considered for the training of LSTM-ACC controller agents with recurrent policies. In addition, when false negative signals of the second leader detection are considered, there are different strategies for the ego-vehicle to cope with this kind of situation. The difference between the complete fall-back to a one-leader system and a vehicle motion tracking method can be explored. This section highlights other issues related to the defined problem in this study and further extends the problem for future studies.

For the controller design setups, the ACC controllers in this study are DNN agents with two hidden fully connected layers between the input and output, while the LSTM-ACC controllers contain an LSTM network layer as the state estimator before the two fully connected layers. To enhance the performance of the controller, the DNN can be redesigned to see if a deeper or wider network architecture or a different activation function can better describe the non-linearity in the control task and generate better car-following performance. Another issue which may degrade the power of DRL is related to the algorithm selected to train the policy network. In this study, the controller agents of the ACC systems are designed by PPO, an on-policy DRL algorithm. Although PPO is believed to be the most powerful on-policy DRL algorithm for the time being, a lot of attention nowadays has started to develop new DRL algorithms which leverage the merits of both the on-policy and off-policy algorithms, such as DDPG and TD3. However, PPO is applied in this study due to the current limitation regarding the applicability of recurrent policies in the selected DRL library. Therefore, it is worth investigating that whether more powerful DRL algorithms can result in an even better ACC controller performance.

Another possible change which can be made in the control system design is related to the utilization of second leader measurements. In the proposed multi-leader ACC system architecture, the minimal acceleration command generated by the two controllers would be selected to ensure safe driving maneuver and follow the determined constant time gap spacing policy. However, this control strategy may lead to large jerk amplitudes when the controlling leader switches. In addition, the presence of noise in the second leader measurements can cause unnecessary slow down behaviors to the ego-vehicle when it falsely perceives a slower speed or decreased distance gap with the second leader. This kind of behavior creates more uncomfortable driving maneuvers and can potentially cause additional traffic disturbance.

To prevent these two possible problems, a different control strategy can be applied to determine the acceleration command in the next time step. In Donà et al. (2022), the measurements of two leaders were incorporated by using a weighting parameter to describe how much influence the second leader measurements have on the linear controller. Another simple logic which can be applied is pre-defining a threshold for the changes in second leader measurements between two consecutive time steps based on the level of measurement noise. The acceleration command generated from the controller following the second leader would only be activated if the change of second leader measurements between two time steps exceeds the threshold. By doing so, the jerk experienced by the vehicle is expected to be reduced. However, whether the string stability performance would be affected by the delayed action caused by the threshold setting when encountering a disturbance should be investigated. These are two possible methods to utilize the second leader measurements collected. An even more advanced approach can be the training of a single

ACC controller agent which takes the measurements of both the first and second leaders as state input at the same time. However, such controller design may not be suitable when platoon heterogeneity is considered.

Platoon heterogeneity can also be the other focus of the future work. In practice, the penetration rate of vehicles equipped with a multi-leader ACC system may not be sufficiently high at the beginning when it hits the market. Whether the string stability can be ensured when a platoon consists of many different types of vehicles, including human-driven vehicle, vehicles equipped with a one-leader ACC system, and those equipped with a two-leader system is an important question to answer. How much improvement a multi-leader ACC system can provide when there are only 10%, 30%, or 50% of vehicles in the platoon possessing the multi-anticipation ability can be quantitatively analyzed through a simulation approach.

At the next level, the platoon heterogeneity may even be considered in the ACC control system design. A single controller which knows how to make decisions no matter what kind of ACC systems is used by the downstream leader can be proposed. A problem can arise for this kind of control system design. What if the direct leader is a human-driven vehicle so that it does not obey the same driving behavior and spacing policy as the ego-vehicle does? How the controller in the multi-leader ACC system should be designed to follow the two leaders ahead which use unknown ACC systems and have different setups in the system is a problem which needs to be addressed. The proposed multi-leader ACC systems in this study are not able to tackle these complicated situations. In this case, the formulation of the DRL problem may have to be entirely modified to train the controller how to react in such a driving environment which has unknown and heterogeneous leader behaviors.

Bibliography

Abdu, F. J., Zhang, Y., Fu, M., Li, Y., & Deng, Z. (2021). Application of Deep Learning on Millimeter-Wave Radar Signals: A Review. *Sensors*, *21*(6), 1951. https://doi.org/10.3390/s21061951

Akita, T., & Mita, S. (2019). Object Tracking and Classification Using Millimeter-Wave Radar Based on LSTM. *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 1110–1115. https://doi.org/10.1109/ITSC.2019.8917144

Albeaik, S., Wu, T., Vurimi, G., Chou, F.-C., Lu, X.-Y., & Bayen, A. M. (2022). Deep truck cruise control: Field experiments and validation of heavy duty truck cruise control using deep reinforcement learning. *Control Engineering Practice*, *121*, 105026.

Aldrich, R., & Wickramarathne, T. (2018). Low-Cost Radar for Object Tracking in Autonomous Driving: A Data-Fusion Approach. *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 1–5. https://doi.org/10.1109/VTCSpring.2018.8417751

Alegre, L. N., Bazzan, A. L., & da Silva, B. C. (2021). Quantifying the impact of non-stationarity in reinforcement learning-based traffic signal control. *PeerJ Computer Science*, *7*, e575. https://doi.org/10.7717/peerj-cs.575

Bae, I., Moon, J., & Seo, J. (2019). Toward a Comfortable Driving Experience for a Self-Driving Shuttle Bus. *Electronics*, *8*(9), 943. https://doi.org/10.3390/electronics8090943

Bakker, B. (2001). Reinforcement Learning with Long Short-Term Memory. *Advances in Neural Information Processing Systems 14 (NIPS 2001)*.

Bareket, Z., Fancher, P. S., Peng, H., Lee, K., & Assaf, C. A. (2003). Methodology for assessing adaptive cruise control behavior. *IEEE Transactions on intelligent transportation systems*, *4*(3), 123–131.

Berthold, P., Michaelis, M., Luettel, T., Meissner, D., & Wuensche, H.-J. (2017). Radar reflection characteristics of vehicles for contour and feature estimation. *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 1–6.

Bifulco, G. N., Simonelli, F., & Di Pace, R. (2008). Experiments toward an human-like Adaptive Cruise Control. *2008 IEEE Intelligent Vehicles Symposium*, 919–924. https://doi.org/10.1109/IVS.2008.4621213

Blauw, M. (2019). *Representing the Car-Following Behaviour of Adaptive Cruise Control (ACC) Systems Using Parametric Car-Following Models*. Master thesis. Delft University of Technology.

Bock, F., Siegl, S., Bazan, P., Buchholz, P., & German, R. (2018). Reliability and test effort analysis of multi-sensor driver assistance systems. *Journal of Systems Architecture*, *85*, 1–13.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). OpenAI Gym.

Calvert, S. C., Schakel, W. J., & van Lint, J. W. C. (2017). Will Automated Vehicles Negatively Impact Traffic Flow? *Journal of Advanced Transportation*, *2017*. https://doi.org/10.1155/2017/3082781

Chandler, R. E., Herman, R., & Montroll, E. W. (1958). Traffic Dynamics: Studies in Car Following. *Operations Research*, *6*(2), 165–184. https://doi.org/10.1287/opre.6.2.165

Chandramouli, A. (2021). *Prediction of Vehicles' Movement based on Kalman-filter and LSTM Algorithms*. Master thesis. Chalmers University of Technology.

Chenna, S. K., Jain, Y. K., Kapoor, H., Bapi, R. S., Yadaiah, N., Negi, A., Rao, V. S., & Deekshatulu, B. (2004). State Estimation and Tracking Problems: A Comparison Between Kalman Filter and Recurrent Neural Networks. *International Conference on Neural Information Processing*, 275–281.

Chong, L., Abbas, M. M., Medina Flintsch, A., & Higgs, B. (2013). A rule-based neural network approach to model driver naturalistic behavior in traffic. *Transportation Research Part C: Emerging Technologies*, *32*, 207–223. https://doi.org/10.1016/j.trc.2012.09.011

Ciuffo, B., Mattas, K., Makridis, M., Albano, G., Anesiadou, A., He, Y., Josvai, S., Komnos, D., Pataki, M., Vass, S., & Szalay, Z. (2021). Requiem on the positive effects of commercial adaptive cruise control on motorway traffic and recommendations for future automated driving systems. *Transportation Research Part C: Emerging Technologies*, *130*, 103305. https://doi.org/10.1016/j.trc.2021.103305

Corona, D., & De Schutter, B. (2008). Adaptive cruise control for a SMART car: A comparison benchmark for MPC-PWA control methods. *IEEE Transactions on Control Systems Technology*, *16*(2), 365–372.

de Bruin, D., Kroon, J., van Klaveren, R., & Nelisse, M. (2004). Design and test of a cooperative adaptive cruise control system. *IEEE Intelligent Vehicles Symposium, 2004*, 392–396. https://doi.org/10.1109/IVS.2004.1336415

Deo, N., & Trivedi, M. M. (2018). Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs. *2018 IEEE Intelligent Vehicles Symposium*, 1179–1184. https://doi.org/10.1109/IVS.2018.8500493

Desjardins, C., & Chaib-Draa, B. (2011). Cooperative adaptive cruise control: A reinforcement learning approach. *IEEE Transactions on intelligent transportation systems*, *12*(4), 1248–1260.

Dijker, T., Bovy, P. H., & Vermijs, R. G. (1997). *Car-following under Non-congested and Congested Conditions* (tech. rep.). LVV report, VK 2206.301.

Dollar, R. A., Molnár, T. G., Vahidi, A., & Orosz, G. (2021). MPC-based connected cruise control with multiple human predecessors. *2021 American Control Conference (ACC)*, 405–411.

Donà, R., Mattas, K., He, Y., Albano, G., & Ciuffo, B. (2022). Multianticipation for string stable Adaptive Cruise Control and increased motorway capacity without vehicle-to-vehicle communication. *Transportation Research Part C: Emerging Technologies*, *140*, 103687. https://doi.org/10.1016/j.trc.2022.103687

Elmquist, A., & Negrut, D. (2020). Methods and Models for Simulating Autonomous Vehicle Sensors. *IEEE Transactions on Intelligent Vehicles*, *5*(4), 684–692. https://doi.org/10.1109/TIV.2020.3003524

Erez, T., & Smart, W. D. (2010). A scalable method for solving high-dimensional continuous POMDPs using local approximation. *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, 160–167.

European Commission. (2018). *On the road to automated mobility: An EU strategy for mobility of the future* (tech. rep.).

Floudas, N., Polychronopoulos, A., & Amditis, A. (2005). A survey of filtering techniques for vehicle tracking by radar equipped automotive platforms. *2005 7th International Conference on Information Fusion*, 8 pp. https://doi.org/10.1109/ICIF.2005.1592024

Forbes, J. R. N. (2002). *Reinforcement learning for autonomous vehicles*. Doctoral dissertation. University of California, Berkeley.

Gamba, J. (2020). *Radar Signal Processing for Autonomous Driving*. Springer Singapore. https://doi.org/10.1007/978-981-13-9193-4

Godjevac, J. (1995). Comparative Study of Fuzzy Control, Neural Network Control and Neuro-Fuzzy Control. *Fuzzy set theory and advanced mathematical applications* (pp. 291–322). Springer.

Gorter, C. M. (2015). *Adaptive Cruise Control in Practice: A Field Study and Questionnaire into its influence on Driver, Traffic Flows and Safety*. Master thesis. Delft University of Technology.

Gunter, G., Gloudemans, D., Stern, R. E., McQuade, S., Bhadani, R., Bunting, M., Delle Monache, M. L., Lysecky, R., Seibold, B., Sprinkle, J., Piccoli, B., & Work, D. B. (2021). Are Commercially Implemented Adaptive Cruise Control Systems String Stable? *IEEE Transactions on Intelligent Transportation Systems*, *22*(11), 6992–7003. https://doi.org/10.1109/TITS.2020.3000682

Gunter, G., Janssen, C., Barbour, W., Stern, R. E., & Work, D. B. (2019). Model-based string stability of adaptive cruise control systems using field data. *IEEE Transactions on Intelligent Vehicles*, *5*(1), 90–99.

Hallouzi, R., Verdult, V., Hellendoorn, H., & Ploeg, J. (2004). Experimental evaluation of a co-operative driving setup based on inter-vehicle communication. *IFAC Proceedings Volumes*, *37*(8), 126–131. https://doi.org/10.1016/S1474-6670(17)31962-6

Han, Y., Wang, M., Li, L., Roncoli, C., Gao, J., & Liu, P. (2022). A physics-informed reinforcement learning-based strategy for local and coordinated ramp metering. *Transportation Research Part C: Emerging Technologies*, *137*, 103584. https://doi.org/10.1016/j.trc.2022.103584

Hanke, T., Hirsenkorn, N., Dehlink, B., Rauch, A., Rasshofer, R., & Biebl, E. (2016). Classification of sensor errors for the statistical simulation of environmental perception in automated driving systems. *2016 IEEE 19th International Conference on Intelligent Transportation Systems*, 643–648.

Hart, F., Okhrin, O., & Treiber, M. (2021). Formulation and validation of a car-following model based on deep reinforcement learning. *arXiv preprint arXiv:2109.14268*.

Hasch, J., Topak, E., Schnabel, R., Zwick, T., Weigel, R., & Waldschmidt, C. (2012). Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. *IEEE Transactions on Microwave Theory and Techniques*, *60*(3), 845–860.

Hasebe, K., Nakayama, A., & Sugiyama, Y. (2003). Dynamical model of a cooperative driving system for freeway traffic. *Physical review E*, *68*(2), 26102.

Hausknecht, M., & Stone, P. (2015). Deep Recurrent Q-Learning for Partially Observable MDPs. *2015 AAAI Fall Symposium Series*.

Hayashi, S., Saho, K., Isobe, D., & Masugi, M. (2021). Pedestrian detection in blind area and motion classification based on rush-out risk using micro-doppler radar. *Sensors*, *21*(10), 3388.

He, Y., Montanino, M., Mattas, K., Punzo, V., & Ciuffo, B. (2022). Physics-augmented models to simulate commercial adaptive cruise control (ACC) systems. *Transportation Re-*

search Part C: Emerging Technologies, *139*, 103692. https://doi.org/10.1016/j.trc.2022.103692

Heess, N., Hunt, J. J., Lillicrap, T. P., & Silver, D. (2015). Memory-based control with recurrent neural networks. *arXiv:1512.04455*.

Herman, R., Montroll, E. W., Potts, R. B., & Rothery, R. W. (1959). Traffic dynamics: analysis of stability in car following. *Operations research*, *7*(1), 86–106.

Hill, A., Raffin, A., Ernestus, M., Gleave, A., Kanervisto, A., Traore, R., Dhariwal, P., Hesse, C., Klimov, O., Nichol, A., Plappert, M., Radford, A., Schulman, J., Sidor, S., & Wu, Y. (2018). Stable Baselines.

Hoogendoorn, S. P., Ossen, S., & Schreuder, M. (2006). Empirics of multianticipative car-following behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *1965*(1), 112–120.

James, R. M., Melson, C., Hu, J., & Bared, J. (2019). Characterizing the impact of production adaptive cruise control on traffic flow: an investigation. *Transportmetrica B: Transport Dynamics*, *7*(1), 992–1012.

Kesting, A., & Treiber, M. (2008). How reaction time, update time, and adaptation time influence the stability of traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, *23*(2), 125–137.

Kesting, A., Treiber, M., & Helbing, D. (2010). Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4585–4605. https://doi.org/10.1098/rsta.2010.0084

Kesting, A., Treiber, M., Schönhof, M., & Helbing, D. (2008). Adaptive cruise control design for active congestion avoidance. *Transportation Research Part C: Emerging Technologies*, *16*(6), 668–683.

Kim, T., & Park, T.-H. (2020). Extended Kalman filter (EKF) design for vehicle position tracking using reliability function of radar and lidar. *Sensors*, *20*(15), 4126.

Knoop, V. L., Wang, M., Wilmink, I., Hoedemaeker, D. M., Maaskant, M., & Van der Meer, E.-J. (2019). Platoon of SAE Level-2 Automated Vehicles on Public Roads: Setup, Traffic Interactions, and Stability. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(9), 311–322. https://doi.org/10.1177/0361198119845885

Ko, S.-J., & Lee, J.-J. (2007). Fuzzy logic based adaptive cruise control with guaranteed string stability. *2007 International Conference on Control, Automation and Systems*, 15–20. https://doi.org/10.1109/ICCAS.2007.4406871

Kreuzen, C. (2012). *Cooperative adaptive cruise control: using information from multiple predecessors in combination with MPC*. Master thesis. Delft University of Technology.

Laval, J. A. (2011). Hysteresis in traffic flow revisited: An improved measurement method. *Transportation Research Part B: Methodological*, *45*(2), 385–391. https://doi.org/10.1016/j.trb.2010.07.006

Laval, J. A., & Leclercq, L. (2010). A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4519–4541. https://doi.org/10.1098/rsta.2010.0138

Lee, D., Lee, S., Chen, Z., Park, B. B., & Shim, D. H. (2021). Design and field evaluation of cooperative adaptive cruise control with unconnected vehicle in the loop. *Transportation Research Part C: Emerging Technologies*, *132*, 103364.

Li, M., Li, Z., Xu, C., & Liu, T. (2020). Deep Reinforcement Learning-Based Vehicle Driving Strategy to Reduce Crash Risks in Traffic Oscillations. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(10), 42–54. https://doi.org/10.1177/0361198120937976

Li, S. E., Jia, Z., Li, K., & Cheng, B. (2014). Fast online computation of a model predictive controller and its application to fuel economy–oriented adaptive cruise control. *IEEE Transactions on intelligent transportation systems*, *16*(3), 1199–1209.

Li, T., Chen, D., Zhou, H., Laval, J., & Xie, Y. (2021). Car-following behavior characteristics of adaptive cruise control vehicles based on empirical experiments. *Transportation Research Part B: Methodological*, *147*, 67–91. https://doi.org/10.1016/j.trb.2021.03.003

Li, T., Chen, D., Zhou, H., Xie, Y., & Laval, J. (2022). Fundamental diagrams of commercial adaptive cruise control: Worldwide experimental evidence. *Transportation Research Part C: Emerging Technologies*, *134*, 103458. https://doi.org/10.1016/j.trc.2021.103458

Liang, C.-Y., & Peng, H. (1999). Optimal adaptive cruise control with guaranteed string stability. *Vehicle system dynamics*, *32*(4-5), 313–330.

Lin, Y., McPhee, J., & Azad, N. L. (2020). Comparison of deep reinforcement learning and model predictive control for adaptive cruise control. *IEEE Transactions on Intelligent Vehicles*, *6*(2), 221–231.

Littman, M. L. (1996). *Algorithms for sequential decision-making*. Doctoral dissertation. Brown University.

Liu, Z., Cai, Y., Wang, H., Chen, L., Gao, H., Jia, Y., & Li, Y. (2021). Robust Target Recognition and Tracking of Self-Driving Cars With Radar and Camera Information Fusion Under Severe Weather Conditions. *IEEE Transactions on Intelligent Transportation Systems*, 1–14. https://doi.org/10.1109/TITS.2021.3059674

Lu, B., Ni, S., & Washburn, S. S. (2015). A Support Vector Regression Approach for Investigating Multianticipative Driving Behavior. *Mathematical Problems in Engineering*, *2015*, 1–10. https://doi.org/10.1155/2015/701926

Ma, X., Karkus, P., Hsu, D., Lee, W. S., & Ye, N. (2020). Discriminative Particle Filter Reinforcement Learning for Complex Partial observations. *International Conference on Learning Representations*.

Makridis, M., Mattas, K., Ciuffo, B., Re, F., Kriston, A., Minarini, F., & Rognelund, G. (2020). Empirical study on the properties of adaptive cruise control systems and their impact on traffic flow and string stability. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(4), 471–484.

Mani, K., Kaushik, M., Singhania, N., & Krishna, K. M. (2019). Learning Adaptive Driving Behavior Using Recurrent Deterministic Policy Gradients. *2019 IEEE International Conference on Robotics and Biomimetics*, 2092–2098. https://doi.org/10.1109/ROBIO49542.2019.8961480

Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource Management with Deep Reinforcement Learning. *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, 50–56. https://doi.org/10.1145/3005745.3005750

Marsden, G., McDonald, M., & Brackstone, M. (2001). Towards an understanding of adaptive cruise control. *Transportation Research Part C: Emerging Technologies*, *9*(1), 33–51.

McAllister, R. T., & Rasmussen, C. E. (2017). Data-Efficient Reinforcement Learning in Continuous State-Action Gaussian-POMDPs. *The 31st Conference on Neural Information Processing Systems*.

Mehra, R. (1972). Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, *17*(5), 693–698. https://doi.org/10.1109/TAC.1972.1100100

Meng, L., Gorbet, R., & Kulic, D. (2021). Memory-based Deep Reinforcement Learning for POMDPs. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5619–5626. https://doi.org/10.1109/IROS51168.2021.9636140

Milakis, D., Snelder, M., van Arem, B., van Wee, B., & de Almeida Correia, G. H. (2017). Development and transport implications of automated vehicles in the Netherlands: scenarios for 2030 and 2050. *European Journal of Transport and Infrastructure Research*, *17*(1).

Milanes, V., Villagra, J., Godoy, J., & Gonzalez, C. (2012). Comparing Fuzzy and Intelligent PI Controllers in Stop-and-Go Manoeuvres. *IEEE Transactions on Control Systems Technology*, *20*(3), 770–778. https://doi.org/10.1109/TCST.2011.2135859

Milanés, V., & Shladover, S. E. (2014). Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. *Transportation Research Part C: Emerging Technologies*, *48*, 285–300. https://doi.org/10.1016/j.trc.2014.09.001

Mirwald, J., Ultsch, J., de Castro, R., & Brembeck, J. (2021). Learning-Based Cooperative Adaptive Cruise Control. *Actuators*, *10*(11), 286. https://doi.org/10.3390/act10110286

Mitra, P., Choudhury, A., Aparow, V. R., Kulandaivelu, G., & Dauwels, J. (2018). Towards Modeling of Perception Errors in Autonomous Vehicles. *2018 21st International Conference on Intelligent Transportation Systems*, 3024–3029. https://doi.org/10.1109/ITSC.2018.8570015

Montanino, M., & Punzo, V. (2021). On string stability of a mixed and heterogeneous traffic flow: A unifying modelling framework. *Transportation Research Part B: Methodological*, *144*, 133–154. https://doi.org/10.1016/j.trb.2020.11.009

Moser, D., Schmied, R., Waschl, H., & del Re, L. (2017). Flexible spacing adaptive cruise control using stochastic model predictive control. *IEEE Transactions on Control Systems Technology*, *26*(1), 114–127.

Naranjo, J., Gonzalez, C., Garcia, R., & dePedro, T. (2006). ACC+Stop&Go Maneuvers With Throttle and Brake Fuzzy Control. *IEEE Transactions on Intelligent Transportation Systems*, *7*(2), 213–225. https://doi.org/10.1109/TITS.2006.874723

Naranjo, J., Gonzalez, C., Reviejo, J., Garcia, R., & de Pedro, T. (2003). Adaptive fuzzy control for inter-vehicle gap keeping. *IEEE Transactions on Intelligent Transportation Systems*, *4*(3), 132–142. https://doi.org/10.1109/TITS.2003.821294

Naus, G., Ploeg, J., Van De Molengraft, R., & Steinbuch, M. (2008). Explicit MPC design and performance-based tuning of an adaptive cruise control Stop-&-Go. *2008 IEEE Intelligent Vehicles Symposium*, 434–439.

Newell, G. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, *36*(3), 195–205. https://doi.org/10.1016/S0191-2615(00)00044-8

Ng, L., Clark, C. M., & Huissoon, J. P. (2008). Reinforcement learning of adaptive longitudinal vehicle control for dynamic collaborative driving. *2008 IEEE Intelligent Vehicles Symposium*, 907–912.

Ohno, H. (2001). Analysis and modeling of human driving behaviors using adaptive cruise control. *Applied Soft Computing*, *1*(3), 237–243. https://doi.org/10.1016/S1568-4946(01)00022-9

Olah, C. (2015). Understanding LSTM Networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Ossen, S. J. L. (2008). *Longitudinal driving behavior: theory and empirics*. Doctoral dissertation. Delft University of Technology.

Palffy, A., Pool, E., Baratam, S., Kooij, J. F. P., & Gavrila, D. M. (2022). Multi-Class Road User Detection With 3+1D Radar in the View-of-Delft Dataset. *IEEE Robotics and Automation Letters*, *7*(2), 4961–4968. https://doi.org/10.1109/LRA.2022.3147324

Papacharalampous, A. E., Wang, M., Knoop, V. L., Ros, B. G., Takahashi, T., Sakata, I., van Arem, B., & Hoogendoorn, S. P. (2015). Mitigating congestion at sags with adaptive cruise control systems. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2451–2457.

Piazzoni, A., Cherian, J., Slavik, M., & Dauwels, J. (2021). Modeling perception errors towards robust decision making in autonomous vehicles. *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 3494–3500.

Ploeg, J., van de Wouw, N., & Nijmeijer, H. (2014). Lp String Stability of Cascaded Systems: Application to Vehicle Platooning. *IEEE Transactions on Control Systems Technology*, *22*(2), 786–793. https://doi.org/10.1109/TCST.2013.2258346

Pueboobpaphan, R., & van Arem, B. (2010). Driver and vehicle characteristics and platoon and traffic flow stability: Understanding the relationship for design and assessment of cooperative adaptive cruise control. *Transportation Research Record: Journal of the Transportation Research Board*, *2189*(1), 89–97.

Qiao, Z., Muelling, K., Dolan, J., Palanisamy, P., & Mudalige, P. (2018). POMDP and Hierarchical Options MDP with Continuous Actions for Autonomous Driving at Intersections. *2018 21st International Conference on Intelligent Transportation Systems*, 2377–2382. https://doi.org/10.1109/ITSC.2018.8569400

Qu, X., Yu, Y., Zhou, M., Lin, C.-T., & Wang, X. (2020). Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach. *Applied Energy*, *257*, 114030.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, *22*, 1–8.

Revach, G., Shlezinger, N., Ni, X., Escoriza, A. L., van Sloun, R. J. G., & Eldar, Y. C. (2022). KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics. *IEEE Transactions on Signal Processing*, *70*, 1532–1547. https://doi.org/10.1109/TSP.2022.3158588

Rosique, F., Navarro, P. J., Fernández, C., & Padilla, A. (2019). A systematic review of perception system and simulators for autonomous vehicles research. *Sensors*, *19*(3), 648.

Sakhdari, B., Shahrivar, E. M., & Azad, N. L. (2017). Robust tube-based mpc for automotive adaptive cruise control design. *2017 IEEE 20th International Conference on Intelligent Transportation Systems*, 1–6.

Saxena, S., Isukapati, I. K., Smith, S. F., & Dolan, J. M. (2019). Multiagent Sensor Fusion for Connected & Autonomous Vehicles to Enhance Navigation Safety. *2019 IEEE Intelligent Transportation Systems Conference*, 2490–2495. https://doi.org/10.1109/ITSC.2019.8917298

Schakel, W. J., van Arem, B., & Netten, B. D. (2010). Effects of cooperative adaptive cruise control on traffic flow stability. *13th International IEEE Conference on Intelligent Transportation Systems*, 759–764.

Scheiner, N., Kraus, F., Wei, F., Phan, B., Mannan, F., Appenrodt, N., Ritter, W., Dickmann, J., Dietmayer, K., & Sick, B. (2020). Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2068–2077.

Schulman, J., Moritz, P., Levine, S., Jordan, M., & Abbeel, P. (2015). High-Dimensional Continuous Control Using Generalized Advantage Estimation. *arXiv:1506.02438*.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms. *arXiv:1707.06347*.

Shang, M., & Stern, R. E. (2021). Impacts of commercially available adaptive cruise control vehicles on highway stability and throughput. *Transportation Research Part C: Emerging Technologies*, *122*, 102897.

Shi, H., Zhou, Y., Wang, X., Fu, S., Gong, S., & Ran, B. (2022). A deep reinforcement learning-based distributed connected automated vehicle control under communication failure. *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/mice.12825

Shi, H., Zhou, Y., Wu, K., Wang, X., Lin, Y., & Ran, B. (2021). Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment. *Transportation Research Part C: Emerging Technologies*, *133*, 103421. https://doi.org/10.1016/j.trc.2021.103421

Shladover, S. E. (1978). Longitudinal Control of Automated Guideway Transit Vehicles Within Platoons. *Journal of Dynamic Systems, Measurement, and Control*, *100*(4), 302–310. https://doi.org/10.1115/1.3426382

Shladover, S. E., Nowakowski, C., Lu, X.-Y., & Ferlis, R. (2015). Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record: Journal of the Transportation Research Board*, *2489*(1), 145–152.

Shladover, S. E., Su, D., & Lu, X.-Y. (2012). Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transportation Research Record*, *2324*(1), 63–70.

Spiliopoulou, A., Manolis, D., Vandorou, F., & Papageorgiou, M. (2018). Adaptive cruise control operation for improved motorway traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(22), 24–35.

Sun, J., Zheng, Z., & Sun, J. (2020). The relationship between car following string instability and traffic oscillations in finite-sized platoons and its use in easing congestion via connected and automated vehicles with IDM based controller. *Transportation Research Part B: Methodological*, *142*, 58–83. https://doi.org/10.1016/j.trb.2020.10.004

Swaroop, D., Hedrick, J., Chien, C. C., & Ioannou, P. (1994). A Comparision of Spacing and Headway Control Laws for Automatically Controlled Vehicles. *Vehicle System Dynamics*, *23*(1), 597–625. https://doi.org/10.1080/00423119408969077

Treiber, M., & Kesting, A. (2013). *Traffic flow dynamics: data, models and simulation.* Springer-Verlag Berlin Heidelberg.

Treiber, M., Kesting, A., & Helbing, D. (2006). Delays, inaccuracies and anticipation in microscopic traffic models. *Physica A: Statistical Mechanics and its Applications*, *360*(1), 71–88.

Treiber, M., Kesting, A., & Helbing, D. (2007). Influence of Reaction Times and Anticipation on Stability of Vehicular Traffic Flow. *Transportation Research Record: Journal of the Transportation Research Board*, *1999*(1). https://doi.org/10.3141/1999-03

Tsai, C.-C., Hsieh, S.-M., & Chen, C.-T. (2010). Fuzzy Longitudinal Controller Design and Experimentation for Adaptive Cruise Control and Stop&Go. *Journal of Intelligent & Robotic Systems*, *59*(2), 167–189. https://doi.org/10.1007/s10846-010-9393-z

Van Brummelen, J., O'Brien, M., Gruyer, D., & Najjaran, H. (2018). Autonomous vehicle perception: The technology of today and tomorrow. *Transportation Research Part C: Emerging Technologies*, *89*, 384–406.

van Arem, B., Hogema, J. H., & Smulders, S. A. (1996). The impact of autonomous intelligent cruise control on traffic flow. *In Intelligent Transportation: Realizing the Future. Abstracts of the Third World Congress on Intelligent Transport Systems ITS America.*

van Arem, B., van Driel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on intelligent transportation systems*, *7*(4), 429–436.

VanderWerf, J., Shladover, S., Kourjanskaia, N., Miller, M., & Krishnan, H. (2001). Modeling effects of driver control assistance systems on traffic. *Transportation Research Record: Journal of the Transportation Research Board*, *1748*(1), 167–174.

VanderWerf, J., Shladover, S. E., Miller, M. A., & Kourjanskaia, N. (2002). Effects of adaptive cruise control systems on highway traffic flow capacity. *Transportation Research Record: Journal of the Transportation Research Board*, *1800*(1), 78–84.

Vargas, J., Alsweiss, S., Toker, O., Razdan, R., & Santos, J. (2021). An Overview of Autonomous Vehicles Sensors and Their Vulnerability to Weather Conditions. *Sensors*, *21*(16), 5397. https://doi.org/10.3390/s21165397

Wang, M., Daamen, W., Hoogendoorn, S., & Van Arem, B. (2014). Potential impacts of ecological adaptive cruise control systems on traffic and environment. *IET Intelligent Transport Systems*, *8*(2), 77–86.

Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2014a). Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. *Transportation Research Part C: Emerging Technologies*, *40*, 290–311.

Wang, M., Daamen, W., Hoogendoorn, S. P., & van Arem, B. (2014b). Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. *Transportation Research Part C: Emerging Technologies*, *40*, 290–311.

Wang, M., Hoogendoorn, S. P., Daamen, W., van Arem, B., Shyrokau, B., & Happee, R. (2018). Delay-compensating strategy to enhance string stability of adaptive cruise controlled vehicles. *Transportmetrica B: Transport Dynamics*, *6*(3), 211–229.

Wang, M., Li, H., Gao, J., Huang, Z., Li, B., & van Arem, B. (2017). String stability of heterogeneous platoons with non-connected automated vehicles. *2017 IEEE 20th Interna-*

tional Conference on Intelligent Transportation Systems, 1–8. https://doi.org/10.1109/ITSC.2017.8317792

Wei, H., Zheng, G., Gayah, V., & Li, Z. (2021). Recent Advances in Reinforcement Learning for Traffic Signal Control. *ACM SIGKDD Explorations Newsletter*, *22*(2), 12–18. https://doi.org/10.1145/3447556.3447565

Wei, J., Dolan, J. M., Snider, J. M., & Litkouhi, B. (2011). A point-based mdp for robust single-lane autonomous driving behavior under uncertainties. *2011 IEEE International Conference on Robotics and Automation*, 2586–2592.

Welch, G., & Bishop, G. (2006). *An Introduction to the Kalman Filter* (tech. rep.). Department of Computer Science, University of North Carolina at Chapel Hill Chapel Hill.

Wilmink, I. R., Klunder, G. A., & van Arem, B. (2007). Traffic flow effects of integrated full-range speed assistance (IRSA). *2007 IEEE intelligent vehicles symposium*, 1204–1210.

Wilson, R. E., & Ward, J. A. (2011). Car-following models: fifty years of linear stability analysis–a mathematical perspective. *Transportation Planning and Technology*, *34*(1), 3–18.

Wu, C., Lin, Y., & Eskandarian, A. (2019). Cooperative Adaptive Cruise Control With Adaptive Kalman Filter Subject to Temporary Communication Loss. *IEEE Access*, *7*, 93558–93568. https://doi.org/10.1109/ACCESS.2019.2928004

Wu, S., Decker, S., Chang, P., Camus, T., & Eledath, J. (2009). Collision Sensing by Stereo Vision and Radar Sensor Fusion. *IEEE Transactions on Intelligent Transportation Systems*, *10*(4), 606–614. https://doi.org/10.1109/TITS.2009.2032769

Xiao, L., Wang, M., & Van Arem, B. (2017). Realistic car-following models for microscopic simulation of adaptive and cooperative adaptive cruise control vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, *2623*(1), 1–9.

Xiao, L., & Gao, F. (2010). A comprehensive review of the development of adaptive cruise control systems. *Vehicle system dynamics*, *48*(10), 1167–1192.

Xiao, L., & Gao, F. (2011). Practical string stability of platoon of adaptive cruise control vehicles. *IEEE Transactions on intelligent transportation systems*, *12*(4), 1184–1194.

Yamamura, Y., Seto, Y., Nishira, H., & Kawabe, T. (2008). An ACC Design Method for Achieving Both String Stability and Ride Comfort. *Journal of System Design and Dynamics*, *2*(4), 979–990. https://doi.org/10.1299/jsdd.2.979

Zec, E. L., Mohammadiha, N., & Schliep, A. (2018). Statistical sensor modelling for autonomous driving using autoregressive input-output HMMs. *2018 21st International Conference on Intelligent Transportation Systems*, 1331–1336.

Zhao, D., Wang, B., & Liu, D. (2013). A supervised actor–critic approach for adaptive cruise control. *Soft Computing*, *17*(11), 2089–2099.

Zheng, T., Yao, Y., He, F., & Zhang, X. (2019). An RNN-based Learnable Extended Kalman Filter Design and Application. *2019 18th European Control Conference*, 3304–3309. https://doi.org/10.23919/ECC.2019.8796088

Zhou, H., Laval, J., Zhou, A., Wang, Y., Wu, W., Qing, Z., & Peeta, S. (2022). Review of Learning-Based Longitudinal Motion Planning for Autonomous Vehicles: Research Gaps Between Self-Driving and Traffic Congestion. *Transportation Research Record: Journal of the Transportation Research Board*, *2676*(1), 324–341. https://doi.org/10.1177/03611981211035764

Zhou, H., Zhou, A., Li, T., Chen, D., Peeta, S., & Laval, J. (2021). Low-level Controllers Play a Decisive Role in the String Stability of Adaptive Cruise Control. *2021 IEEE Interna-*

tional Intelligent Transportation Systems Conference, 891–898. https://doi.org/10.1109/ITSC48978.2021.9564679

Zhou, J., & Peng, H. (2005). Range policy of adaptive cruise control vehicles for improved flow stability and string stability. *IEEE Transactions on intelligent transportation systems*, *6*(2), 229–237.

Zhou, M., Qu, X., & Li, X. (2017). A recurrent neural network based microscopic car following model to predict traffic oscillation. *Transportation Research Part C: Emerging Technologies*, *84*, 245–264. https://doi.org/10.1016/j.trc.2017.08.027

Zhou, T., Yang, M., Jiang, K., Wong, H., & Yang, D. (2020). MMW Radar-Based Technologies in Autonomous Driving: A Review. *Sensors*, *20*(24), 7283.

Zhou, Y., Ahn, S., Chitturi, M., & Noyce, D. A. (2017). Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty. *Transportation Research Part C: Emerging Technologies*, *83*, 61–76.

Zhu, M., Wang, X., & Wang, Y. (2018). Human-like autonomous car-following model with deep reinforcement learning. *Transportation Research Part C: Emerging Technologies*, *97*, 348–368.

Zhu, M., Wang, Y., Pu, Z., Hu, J., Wang, X., & Ke, R. (2020). Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving. *Transportation Research Part C: Emerging Technologies*, *117*, 102662. https://doi.org/10.1016/j.trc.2020.102662

Zhu, Z., & Zhao, H. (2021). A Survey of Deep RL and IL for Autonomous Driving Policy Learning. *IEEE Transactions on Intelligent Transportation Systems*, 1–23. https://doi.org/10.1109/TITS.2021.3134702