2008 IEEE Intelligent Vehicles Symposium
Eindhoven University of Technology
Eindhoven, The Netherlands, June 4-6, 2008

Reinforcement Learning of Adaptive Longitudinal Vehicle Control for Dynamic Collaborative Driving

Luke Ng, Christopher M. Clark, and Jan P. Huissoon

Abstract— **Dynamic collaborative driving involves the motion coordination of multiple vehicles using shared information from vehicles instrumented to perceive their surroundings in order to improve road usage and safety. A basic requirement of any vehicle participating in dynamic collaborative driving is longitudinal control. Without this capability, higher-level coordination is not possible. This paper focuses on the problem of longitudinal motion control. A detailed nonlinear longitudinal vehicle model which serves as the control system design platform is used to develop a longitudinal adaptive control system based on Monte Carlo Reinforcement Learning. The results of the reinforcement learning phase and the performance of the adaptive control system for a single automobile as well as the performance in a multi-vehicle platoon is presented.**

I. Introduction

IN major cities throughout the world, urban expansion is leading to an increase of vehicle traffic flow. One solution is to build more roads; another is to automate the process of driving. Dynamic Collaborative Driving is an automated driving approach where multiple vehicles dynamically form groups and networks, sharing information in order to build a dynamic representation of the road to coordinate efficient road travel while maintaining safety.

Ultimately our research goal is to create a decentralized control system capable of performing dynamic collaborative driving which is scalable to a large number of vehicles, can be used on any vehicle and in any environment. However, before we can deal with the issue of coordination, basic control of the vehicle must be achieved. The focus of this paper is longitudinal motion control, commonly referred to as adaptive cruise control (ACC). The use of *adaptive* in ACC is a misnomer as it does not refer to the type of control but is used to indicate that distance control is present in addition to speed control.

Ioannou and Chien [1] describe an autonomous intelligent adaptive cruise control system (AICC) for automatic vehicle following using a linear vehicle following model. Studies by Hedrick in the mid 1990s at UC Berkeley [2], [3] focused on using sliding mode control to address the nonlinearities of longitudinal vehicle dynamics. More recently, Zhang and Ioannou [4] proposed an adaptive control approach to vehicle following with using a simplified first order linear vehicle model.

Due to the high costs associated with procuring large numbers of vehicles and the safety issues involved, full-scale vehicle studies can only be conducted through large scale government research projects in association with governments and automobile manufacturers such as Demo '97 [5][6][7], and in Japan during Demo 2000 [8]. In Canada, smaller projects have used small mobile robots to model cars [9], however the cost and complexity associated with these mobile robot studies can also be quite high. In addition the vehicle dynamics of a mobile robot platform are significantly different from those of full-sized automobiles thereby limiting the applicability of those results.

Alternatively, simulation studies can be developed faster, they are more flexible, cost effective, have better repeatability and explore situations not easily achieved in reality. In 1989, the National Highway Traffic Safety Administration (NHTSA) began researching the use and construction of a new state-of-the-art driving simulator, the National Advanced Driving Simulator (NADS) [10]. Since then, NADS has been used as a substitute for actual vehicle testing. NHTSA's Vehicle Research and Test Center (VRTC) provides vehicle data for a number of vehicles (i.e. the 1997 Jeep Cherokee [11]), which can be used to validate simulations. With the adoption of high fidelity simulation on modern computers, simulation has become the dominant method of study in this field.

II. Vehicle Model

The basis of our vehicle simulation has its roots going back to the late 1980's. A significant amount of research was conducted at the Vehicle Dynamics Laboratory at the University of California at Berkeley by Hedrick under the PATH project. His group developed a complex numerical automobile model used to design and evaluate the performance of various controllers under certain driving conditions [12] [13].

The vehicle model in Figure 1, adopts many of the models used by Hedrick's group for key subsystems such as the engine, transmission, suspension and tires. However, in order to have a simulation which can be subjected to reinforcement learning, these separate models have been

March 15, 2007. This work was funded by AUTO21 Canada.

Luke. Ng is a PhD candidate studying mobile robotics in the Dept. of Mechanical and Mechatronics Engineering, University of Waterloo, ON, Canada N2L 3G1 l4ng@engmail.uwaterloo.ca

Christopher M. Clark is an Assistant Professor at the Computer Science Department, California Polytechnic State University, San Luis Obispo, CA, USA 93407 cmclark@calpoly.edu

Jan P. Huissoon is a Professor and Deputy Chair of the Dept. of Mechanical and Mechatronics Engineering, University of Waterloo, ON, Canada N2L 3G1, jph@mecheng1.uwaterloo.ca.

integrated to provide system performance throughout its entire operating range. Most of the subsystem models are nonlinear such as the *Engine*, *Transmission* [14], *Drivetrain* [14] and *Tire* models [14], [15]. Other subsystems such as the *Throttle* and *Brake Actuator* model are first order linear systems [14]. The *Brake System* is modeled with a linear function [14] while the *Suspension* is second order linear systems [16]. Figures 2 through 4 show the vehicle model's velocity responses to various throttle and brake step inputs.

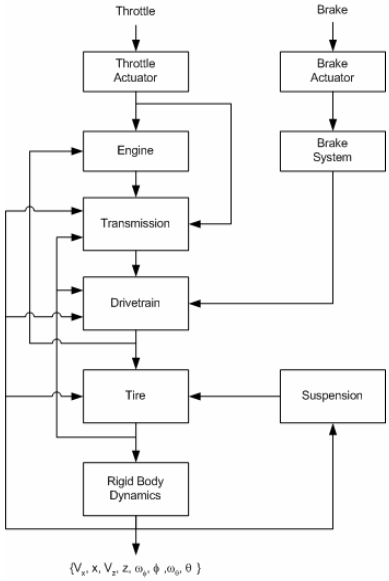


Figure 1 Overview of the vehicle model

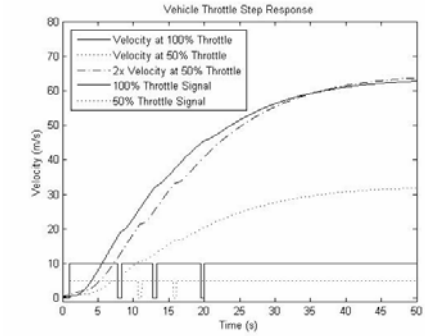


Figure 2 Vehicle model velocity responses to throttle step inputs

III. DESIGN

The outputs of the longitudinal controller are i) the throttle angle, which controls the fuel/air mixture for the combustion process within the engine and ii) the brake pedal position, which applies a braking torque to each wheel. In Figure 2 the vehicle response to a throttle step input can be characterized as a second order over-damped response with a slight delay. The vehicle response to a 50% throttle step input is also shown in the figure. By comparing the 50% response multiplied by a factor of two with the 100% response we see that the vehicle model's response with respect to the throttle is clearly non-linear.

In Figure 3 the vehicle model velocity response to a step input demonstrates that during braking, Coulomb friction dominates the system. The vehicle response to a 50% brake step input is also shown in the figure and is clearly not half of the 100% signal indicating that the modeled braking system is non-linear.

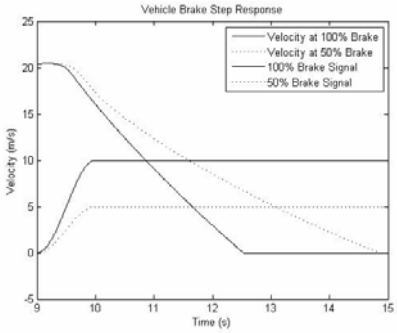


Figure 3 Vehicle model velocity responses to brake step inputs

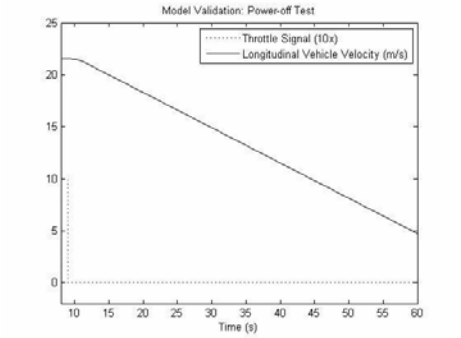


Figure 4 Vehicle model velocity response to power-off condition

Figure 4 shows the vehicle model response when the throttle is disengaged, this can be considered a step input from 1 to 0. The throttle power off resembles the brake system's response although more gradual. It demonstrates the Coulomb friction as well and can be considered a nonlinear response.

To address each of these nonlinear responses, different control systems are required depending on the operating conditions. One approach is to divide the control space into regions within which the behavior of the plant approximates linearity. A patch-work of linear controllers would then be able to address the entire operating envelope. These linear controllers would all have the same form, but their gains would differ depending on the operating conditions. This common linear controller along with its collection of gains is considered a form of adaptive control referred to as gain scheduling [17]. The difference in our implementation of gain scheduling, is that the tedious task of determining each gain is achieved using a machine learning algorithm called *Monte Carlo ES* reinforcement learning.

A. Reinforcement Learning

Reinforcement learning (RL) is a machine learning approach where a software agent senses the *environment* through its *states s* and responds to it through its *actions a* under the control of a *policy*, $a = \pi(s)$. This policy is improved iteratively through its experiences with the

environment through a *reinforcement learning algorithm* which in this paper is called *Monte Carlo ES* (Figure 5). The environment provides the agent with numerical feedback called a *reward* for the current state, $r = R(s)$. The environment also supplies the next state based on the current state and the actions taken using the *transition function* σ, $s' = \sigma(s, a)$. In this study, the transition function is provided by the vehicle model. The control problem is formulated into mathematical framework known as a finite Markov Decision Process (MDP) [18] by defining $\{s, a, \pi, R(s), \sigma(s,a)\}$. The key feature of an MDP is that to be considered Markov, its current state must be independent of previous states. This is so that for each visit to a state, the software agent is given a path independent reward. Subsequent actions will result in new states giving rise to different rewards.

The challenge of reinforcement learning is to determine the actions which result in the maximum reward for every possible state, this state to action mapping is called the optimal policy π^* or the controller. For the current state, actions that result in more favorable future states lead to higher rewards. The favorability of a certain action given the current state is known as the *Q*-Value. As an agent experiences its environment, it updates the *Q*-Value for each state-action (s,a) pair it visits according to its reinforcement learning algorithm. As it repeatedly visits every (s,a), it updates the policy so that the highest valued (s,a) will dominate. The optimal policy is reached when every state-action pair results in the highest reward possible; that is when the *Q*-Value function has been maximized. The convergence of this maximization process requires that all states and actions be visited infinitely in order for estimates of the *Q*-value to reach their actual values. To ensure this convergence criterion, policies leading to π^* are ε-soft, meaning that there is a ε probability that a random action is selected, thus all actions and states will be reached as $t \to \infty$.

Initialize, for all $s \in S$, $a \in A(s)$:
  $Q(s, a) \leftarrow$ arbitrary
  $\pi(s) \leftarrow$ arbitrary
  $Returns(s, a) \leftarrow$ empty list

Repeat forever:
  (a) Generate an episode using exploring starts π
  (b) For each pair (s, a) appearing in the episode
    $R \leftarrow$ return following the first occurrence of (s,a)
    Append R to $Returns(s, a)$
    $Q(s, a) \leftarrow$ average($Returns(s, a)$)
  (c) For each s in the episode:
    $\pi(s) \leftarrow argmax_a\, Q(s, a)$

Figure 5 Monte Carlo ES algorithm

The key to the process of improvement is the reward function which expresses the desirability of being in a current state. It is the method of communicating to the agent the task to be performed. The challenge of the designer is to be able to come up with a reward function that captures the essence of the task so that learning can be achieved. Monte Carlo reinforcement learning algorithms find the optimal policy using the averaged sample returns experienced by the agent at the end of each episode [19].

B. Longitudinal Control

Simply stated, longitudinal control of a vehicle is to be able to follow another vehicle in traffic without colliding into it. That is, the controller must maintain a relative speed of zero with the vehicle ahead while maintaining a fixed distance behind the forward vehicle; this fixed distance will be referred to as Δx_i. During the process of control, the vehicle's relative speed, $Vx_{i-1} - Vx_i$ and range, $x_{i-1} - x_i$, to the vehicle ahead will provide feedback to the control system. Figure 6 shows how multiple vehicle's are linked to provide longitudinal control for multiple vehicles.

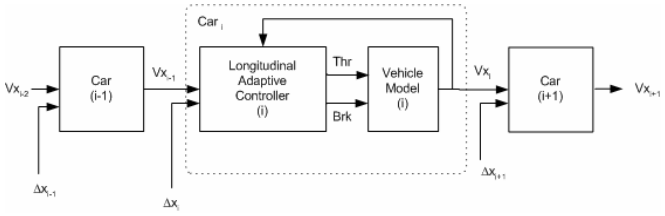


Figure 6 Overview of longitudinal control system

Two parallel control systems are used, one for throttle control, and one for brake control. These two throttle and brake controllers are a combination of a digital Proportional-Derivative (PD) controller for V_{rel}, and a digital Proportional-Integral (PI) controller for X_{rel}. The difference equation which provides the throttle/brake command m_n is shown below

$$m_n = m_{n-1} + k_{pv}(v_n - v_{n-1}) + \frac{k_{dv}}{\Delta T}(v_n - 2v_{n-1} + v_{n-2}) + k_{px}(x_n - x_{n-1}) + k_{ix}\Delta T\, x_n \quad (1)$$

where n is the current iteration of the control cycle, v is V_{rel}, x is X_{rel}, and ΔT is the period of the control cycle. Moreover, k_{pv}, k_{dv}, k_{px}, and k_{ix} are gains that are functions of MDP state variables s_1, s_2, and s_3 as described in Table 1.

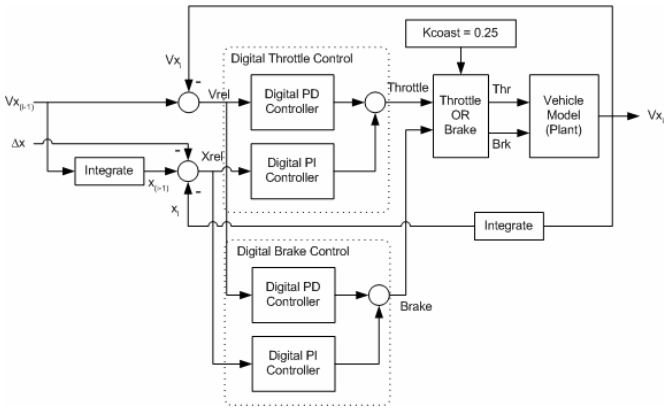


Figure 7 Block diagram of longitudinal controller

This allows simultaneous regulation of the relative speed as

well as the range while reducing the steady state range error through the integral control of the range. The results of both the throttle and brake controllers are fed into a logic element controlled by the gain K_{coast} which decides whether throttle control or brake control is to be used. In this case, K_{coast} is set to 0.25; that is throttle values less than -0.25 utilize the braking system rather than coasting. The logic for this element is shown below.

$$\begin{aligned} &\text{if } (throttle > 0) \\ &\quad cmd_{throttle} = throttle,\ cmd_{brake} = 0 \\ &\text{else if } (throttle < -K_{coast}) \\ &\quad cmd_{throttle} = 0,\ cmd_{brake} = \text{abs}(brake) \\ &\text{else} \\ &\quad cmd_{throttle} = 0,\ cmd_{brake} = 0 \end{aligned} \tag{2}$$

Table 1 States of the longitudinal control problem MDP

State	*Description*	*Digitization Sets*
s_1	Vx_0: initial speed	{ 5, 10, 15, …, 40} m/s
s_2	Vx_{i-1}: target speed	{ 5, 10, 15, …, 40} m/s
s_3	$\Delta x_f - \Delta x_0$: change in spacing	{-100, -90, -80, …, 80, 90, 100} m

For a given operating point, there are eight parameters or gains which must be provided in a lookup table or schedule. By formulating the control problem into a MDP, the gain schedule can be learned using reinforcement learning. The episode is defined as starting at the onset of a change in Vx_{i-1} and ending when $Vx_i = Vx_{i-1}$ or when Vx_{i-1} has been changed. This follows the logic that when a new velocity is required, a set of gains should be selected from the gain schedule and applied for the duration of that command. The goodness of a set of gains can therefore only be assessed once the command is complete, thus the MDP is episodic in nature and the Monte Carlo ES reinforcement learning algorithm described in Figure 5 is used to learn the gain schedule.

Table 2 Actions of the longitudinal control problem MDP

Action	*Description of Gains*	*Digitization (n_s=100)*
a_1	K_{px}:Throttle$_x$ Proportional	{0.1, 0.2, …, 9.9}
a_2	K_{ix}: Throttle$_x$ Integral	{0.01, 0.02, …, 0.99}
a_3	K_{pv}:Throttle$_V$ Proportional	{0.1, 0.2, …, 9.9}
a_4	K_{dv}:Throttle$_V$ Derivative	{0.01, 0.02, …, 0.99}
a_5	K_{px}: Brake$_x$ Proportional	{0.1, 0.2, …, 9.9}
a_6	K_{ix}: Brake$_x$ Integral	{0.01, 0.02, …, 0.99}
a_7	K_{pv}: Brake$_V$ Proportional	{0.1, 0.2, …, 9.9}
a_8	K_{dv}: Brake$_V$ Derivative	{0.01, 0.02, …, 0.99}

The choice in the selection of states lies in the nonlinear nature of the throttle plant. At different initial speeds the throttle responds differently. Therefore, the controller gains will differ from a given initial speed to a final speed. In addition the distance required to achieve this acceleration/deceleration which is reflected in the change in vehicle spacing is also an independent variable for the gain schedule. These three parameters are used as states (Table 1). The actions are the eight values which represent the gains used in the digital control system (Table 2).

The reward is a discrete function of the feedback variables, the current normalized relative speed and normalized relative velocity of the vehicle and is expressed below.

$$R_{Total} = R_V(V_{rel}) + R_X(X_{rel}) \tag{3}$$

$$R_X(X_{rel1}) = \begin{cases} 1 & if\ X_{rel} \le 0.1 \\ -1 & if\ X_{rel} < 0 \end{cases} \tag{4}$$

$$R_V(V_{rel1}) = 1 \quad if\ |V_{rel}| < 0.1 \tag{5}$$

For a given episode, the solution which maximizes the reward, or minimizes the X_{rel} and V_{rel} without colliding with the vehicle ahead ($X_{rel} < 0$) will be favored. These favored solutions will be further explored to determine the optimal solution.

IV. Experimental

A. Reinforcement Learning (RL)

The purpose of the *RL* experiments is to obtain an optimal policy for the longitudinal control of the vehicle. An experiment consists of 300 episodes where ε of the ε-soft greedy policy is set to 0.25 for a particular combination of the 3 states. For each episode the agent must follow another vehicle placed ahead of it which is traveling at a constant speed. Once the leading vehicle has reached the end of the test track, the episode is complete. The distance of the test track is dependent on the speed of the lead vehicle using the following equation.

$$x_{\max} = (1 + 0.2\,v_{lead}) \times 1000 \quad m \tag{6}$$

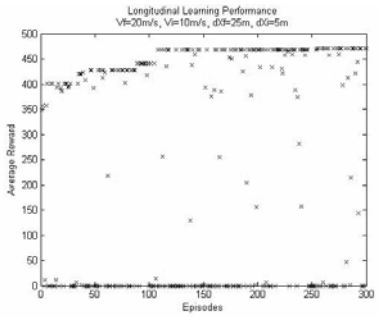


Figure 8 Performance of a typical *RL* experiment

During each step of an episode, a reward is generated (See Equations 3, 4 and 5), this reward is accumulated during the course of an episode to measure the controller's tracking performance using a particular set of actions. Since it is possible to collide with the vehicle ahead during an episode, it would be beneficial if the reward were averaged to reflect how far the vehicle traveled during the course of the episode. Therefore, the average reward for the course of the entire episode is provided by the following equation.

$$R_{avg} = \frac{\sum_{i=0}^{final} R_i}{x_{\max} - x_{final}} \tag{7}$$

Figure 8 shows the average reward as the agent progresses through the learning cycle for a particular state combination. The learning performance is similar for all combinations. One can observe the steady increase in the average reward which eventually reaches a plateau.

For each of the 1344 state combinations, an *RL* experiment is performed to generate the optimal policy π*. This policy is a collection of eight four-dimensional discrete hyperspaces, one for each gain of the longitudinal controller; that is four for the throttle controller and four for the brake controller.

$$\pi^* = \begin{cases} \overset{Throttle}{k_{px}(v_f, v_0, \Delta x_f - \Delta x_0)} & \overset{Brake}{k_{px}(v_f, v_0, \Delta x_f - \Delta x_0)} \\ k_{ix}(v_f, v_0, \Delta x_f - \Delta x_0) & k_{ix}(v_f, v_0, \Delta x_f - \Delta x_0) \\ k_{pv}(v_f, v_0, \Delta x_f - \Delta x_0) & k_{pv}(v_f, v_0, \Delta x_f - \Delta x_0) \\ k_{dv}(v_f, v_0, \Delta x_f - \Delta x_0) & k_{dv}(v_f, v_0, \Delta x_f - \Delta x_0) \end{cases} \tag{8}$$

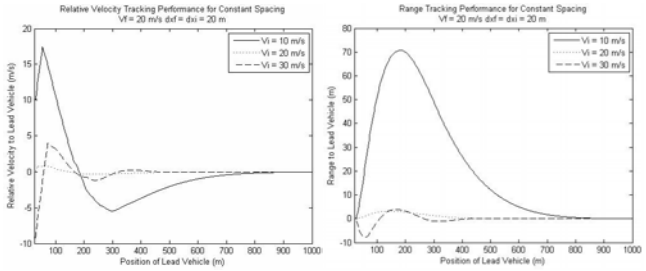


Figure 9 Speed control experiments at 20 *m*/*s*

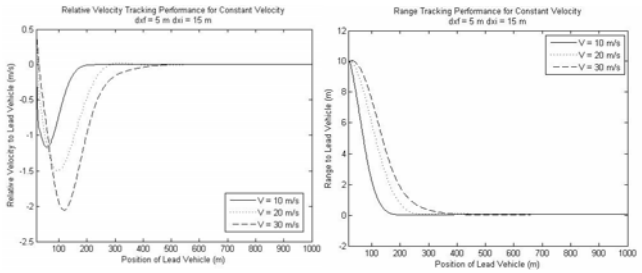


Figure 10 Negative range control experiments

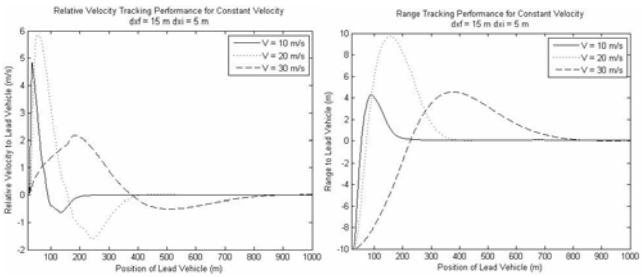


Figure 11 Positive range control experiments

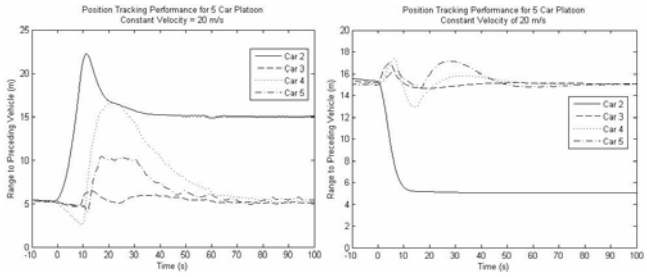
B. *Controller Performance*

These experiments demonstrate the tracking performance of the optimal policy at various operating points. Three control situations are shown which form the basis of platoon maneuvers which allow vehicles to enter or exit formations. The first of these, shown in Figure 9 is speed control. The final speed is 20 *m*/*s* and V_{rel} and X_{rel} are plotted respectively for 3 cases where the vehicle must increase, maintain or decrease its speed.

The second, shown in Figure 10 is negative range control. The initial range is 15 *m* and the vehicle must move to a final range of 5 *m* under while maintaining a constant speed. V_{rel} and X_{rel} are plotted respectively for speeds of 10, 20 and 30 *m*/*s*. The final control situation, shown in Figure 11 is positive range control. The initial range is 5 *m* and the vehicle must move to a final range of 15 *m* under while maintaining a constant speed. V_{rel} and X_{rel} are plotted respectively for speeds of 10, 20 and 30 *m*/*s*.

C. *Multi-Vehicle Performance*

These experiments show the operation of the control system within a five car formation or platoon. Four control situations have been chosen to demonstrate the range tracking performance of the optimal policy for each of the four following vehicles.



(a) Open (b) Close

Figure 12 Multi-vehicle range control experiments

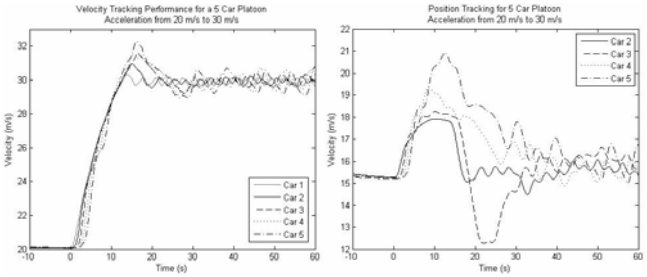


Figure 13 Acceleration experiment (20-30 *m*/*s*)

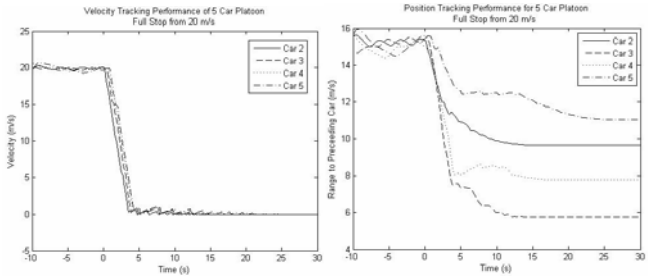


Figure 14 Emergency Stop experiment (20-0 *m*/*s*)

Figure 12 shows the results of a five car formation moving at a constant speed of 20 *m*/*s* for two situations. For the experiment in Figure12a the inter-vehicle spacing is set to 5*m* between each car. At time *t* = 0 s, *Car 2* is instructed to open the space in front to 15 *m*. The results show *Car 2* overshooting the 15 *m* to roughly 22 *m*, in 35 *s* the car has reached a steady-state separation of 15 *m*, the following cars reach the steady-state by 70 *s*. In the experiment shown in

Figure 12b the inter-vehicle spacing is set to 15 *m* between each car. At time $t = 0$ *s*, *Car 2* is instructed to close the space in front to 5 *m*. The results show *Car 2* reaching 5 *m* in 10 *s* without overshoot; the following cars reach 5 *m* in 50 *s*.

Figure 13 shows the results of a five car formation trying to maintain constant spacing while accelerating from 20-30 *m*/*s*. The inter-vehicle spacing is set to 20 *m* between each car. At time $t = 0$ *s*, *Car 1* is to accelerate to 30 *m*/s. The results show a close tracking of the velocity with the presence of oscillations. The position tracking exhibits oscillations which decrease with time, however, no collisions occur. Figure 14 shows an emergency stop situation with a 15 *m* inter-vehicle spacing. The tracking of the velocity and is excellent with a very steep deceleration. All vehicles stop without colliding into the vehicle ahead.

V. Conclusion

In this paper the nonlinear nature of the vehicle dynamics is shown. Due to the nonlinearities present in the engine model, the transmission model, and the tire model a complex nonlinear model results. From this, we conclude that linearization of the longitudinal model may not be suitable for the entire operating range of the vehicle. The linear controllers resulting from using a simplified linear model of the vehicle dynamics in the design process may only be adequate for a particular operating point.

The use of a more accurate nonlinear vehicle dynamics model in the design process should result in better nonlinear control systems for longitudinal control. In this paper, an adaptive control system using gain scheduling is introduced whereby the gains are learned using reinforcement learning. Even with a simple reward function (Eq 3, 4, and 5), it is possible for Monte Carlo reinforcement learning to converge upon an optimal policy within 300 episodes for a particular operating regime; therefore, the MDP properly describes the task to be learned.

When the learned optimal policies are combined to provide an adaptive control surface or a gain schedule, nonlinear control is achieved throughout the operating range. The performance of the controller at specific operating points shows accurate tracking of both velocity and position in most cases. When the adaptive controller is deployed in a multi-vehicle convoy or platoon, the tracking performance is less smooth. As the second car attempts to track the leader, it oscillates. This oscillation is passed to the following cars, as we move farther in the formation, the oscillations decrease, implying stability. The performance of the adaptive controller in a multi-vehicle convoy or platoon shows promise and forms the basis of higher level platoon maneuvers.

References

[1] Ioannu P.A. and Chien C.C. "Autonomous Intelligent Cruise Control". *IEEE Transactions on Vehicular Technology*, Vol 42, No. 4, Nov, pp 657-672. 1993.

[2] Maciuca D.B., Hedrick, J.K. "Advanced Nonlinear Brake System Control for Vehicle Platooning". *Proceedings of the Third European Control Conference (ECC 1995)*, Rome, Italy. 1995.

[3] Swaroop D., Hedrick J.K. "Direct Adaptive Longitudinal Control for Vehicle Platoons". *IEEE Conference on Decision and Control*, December. 1994.

[4] Zhang J. and Ioannou P.A. "Adaptive Vehicle Following Control System with Variable Time Headways". *Proceedings of 44th IEEE Conference on Decision and Control and 2005 European Control Conference*. CDC-ECC '05. pp 3880 – 3885. 2005.

[5] Raza H. and Ioannou P. "Vehicle following control design for automated highway systems". *Proceedings of 1997 IEEE 47th Vehicular Technology Conference*, Phoenix, AZ, USA, Vol 2, pp 904-908. 1997.

[6] Rajamani R., Tan H.S., Law B.K., Zhang W.B. "Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons". *IEEE Transactions on Control Systems Tech*. Vol 8, Issue 4, July, pp 695-708. 2000.

[7] Thorpe C., Jochem T., Pomerleau, D. "The 1997 automated highway free agent demonstration". *Proceedings of IEEE Conference on Intelligent Transportation System*, 1997. ITSC 97. Boston, MA, USA, pp 495-501. 1997.

[8] Kato, S., Tsugawa S., Tokuda, K., Matsui T. Fujii, H. "Cooperative Driving of Automated Vehicles with Inter-vehicle Communications". *IEEE Transactions on Intelligent Transportation Systems*. Volume: 3, Issue: 3, pp 155- 161. 2002.

[9] Michaud, F., Lepage, P., Frenette, P., Létourneau, D., Gaubert, N. "Coordinated maneuvering of automated vehicles in platoon". *IEEE Transactions on Intelligent Transportation Systems, Special Issue on Cooperative Intelligent Vehicles*, 7(4):437-447. 2006.

[10] Haug, E. J. "Feasibility Study and Conceptual Design of a National Advanced Driving Simulator", NHTSA Contract DTNH22-89-07352, *Report No. DOT-HS-807-597*, March. 1990.

[11] Salaani, M. K., Heydinger, G. J. "Model Validation of the 1997 Jeep Cherokee for the National Advanced Driving Simulator". *SAE Paper No. 2000-01-0700*, March, 2000.

[12] Pham H., Tomizuka M., Hedrick K. "Integrated Maneuvering Control for Automated Highway Systems Based on a Magnetic Reference Sensing System". *Research Reports: UCB-ITS-PRR-97-28*, Institute of Transportation Studies California Partners for Advanced Transit and Highways (PATH), University of California, Berkeley, USA. 1997.

[13] Pham H., Hedrick K., Tomizuka M. "Combined lateral and longitudinal control of vehicles for IVHS". *Proceedings of the 1994 American Control Conference*, Vol.2, pp 1205 – 1206. 1994.

[14] McMahon, D. H; Hedrick, J K. "Longitudinal model development for automated roadway vehicles". *Research Report: UCB-ITS-PRR-89-05*. Institute of Transportation Studies California Partners for Advanced Transit and Highways (PATH), University of California, Berkeley, USA. 1989.

[15] Bakker, E., Nyborg, L., Pacejka, H. B. "Tyre Modeling for Use in Vehicle Dynamics Studies", *SAE Technical Paper Series*, No 870421. 1987.

[16] Peng H., Zhang W.B., Arai A., Lin Y., Hessburg T., Devlin P., Tomizuka M., Shladover S. "Experimental Automatic Lateral Control System for an Automobile", *Research Reports: UCB-ITS-PRR-92-11*, Institute of Transportation Studies California Partners for Advanced Transit and Highways (PATH), University of California, Berkeley, USA. 1992.

[17] Aström K. J., Wittenmark B. *Adaptive Control*, Addison-Wesley. 1994.

[18] Bellman R. E. "A Markov decision process". *Journal of Mathematical Mech.*, Vol 6 pp 679-684. 1957.

[19] Sutton, R.S. and Barto A.G. *Reinforcement Learning: An Introduction*. The MIT Press. Cambridge. USA. 1998.