博士論文

Driving Characteristics in Mixed Traffic and Replicability of Car-Following Models

（車種混合交通における運転特性と追従モデルの再現性）

長濱　章仁

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Importance of Study on Mixed Traffic

Traffic jams are increasingly becoming a serious problem in worldwide owing to the growth of the economies [1] and populations [2] in various countries. Although widespread usage of motor vehicles enhances economic activity, and improves convenience and quality of life, many problems, such as air pollution, fuel depletion, and economic losses, arise from traffic jams.

Homogeneous traffic flows, i.e., traffic flows comprising one type of vehicle, have been modeled by many researchers. As reviewed in [3], there are several types of models, e.g., microscopic follow-the-leader models, cellular automaton, macroscopic traffic models, gas-kinetic traffic models, etc. The details of microscopic models are reviewed in [4].

Although the differences between these models will be described in Section 3.1, insights obtained from these models have enabled researchers to discuss theories for reducing of traffic jams without infrastructure construction, e.g., road widening. For example, Nishi *et al.* proposed a theoretical framework called "jam-absorption driving" [5], and suggested theoretical conditions to remove a jam while not creating another jam because of the jam-absorption driving. This theoretical framework was verified by Taniguchi *et al.* [6]. They successfully removed a jam in a test circuit using human-driven cars, and the feasibility of jam reduction without infrastructure construction was proven. Researchers have also realized that traffic flow is stabilized when drivers refer multiple leaders and even followers using these traffic models [7, 8, 9, 10, 11, 12].

In addition, researchers have focused on increasing road capacity through autonomous vehicles and/or connected vehicles, although discussion has also been presented from the safety perspective [13]. In an early work, Varaiya *et al.* presented a rough system architecture for an intelligent vehicle/highway system [14]. Hanebutte *et al.* provided a simulation framework for the intelligent vehicle/highway system as an autonomous intelligent cruise control [15]. Ferlis demonstrated an automated highway system on l-15 in San Diego, CA [16]. In the demonstration, eight fully automated vehicles were driven in a tight formation, i.e., 6.5 m apart at a velocity of approximately 105 km/h. The demonstration showed the feasibility of high traffic flow realized by automated systems. These days, with the introduction of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications to the automated systems [17], the potential performance

of cooperative adaptive cruise control (CACC) for traffic flow is under discussion. Naus *et al.* investigated the conditions for string stability in a system with CACC vehicles and clarified that, under a velocity-dependent intervehicle spacing policy, V2V wireless communication enabled small distance between vehicles maintaining string stability [18]. Wang *et al.* introduced multi-anticipative adaptive cruise control under cooperative control strategy and confirmed that, even when human drivers were mixed into the traffic, smoother deceleration and responsive acceleration which relate to traffic stability were realized [19]. Talebpour *et al.* examined the effects of connected and autonomous vehicles and found that autonomous vehicles were more effective than connected vehicles alone in the prevention of shockwave propagation [20]. Ploeg *et al.* investigated the effectiveness of CACC for the stable traffic using real passenger cars [21].

Regardless of whether the reduction of traffic jams is realized by human drivers or CACC, the traffic jam reduction relying on the improvement of driving methods requires no or a small amount of infrastructure construction and contributes to low-cost traffic improvement. Indeed, there are many types of problems related to driver education and autonomous technology, and jam reduction relying on driving methods will not yield traffic flows that exceed road capacity. However, the feasibility of introducing jam reduction methods is worth investigating, especially in developing countries where insufficient resources are allocated for traffic improvement despite the increasing number of vehicles. Jam reduction relying on driving methods will make the most of limited traffic resources in these countries.

However, the traffic situation in some developing countries are totally different from the traffic targeted by the models mentioned above. This is primarily because of the various types of vehicles in traffic. We call traffic composed of several types of vehicles as "mixed traffic." For instance, Figure 1.1 shows mixed traffic observed in Mumbai, India, in January 2017. Not only normal passenger cars and trucks but also motorcycles, bicycles, and motorized three-wheelers called auto-rickshaws were observed at a certain ratio. Hsu *et al.* investigated the proportions of vehicle types in Taiwan, Malaysia, and Vietnam ,and clarified the following [22]:

1. The number of utilized vehicles are increasing in these countries.

2. Mixed traffic occurs in Taiwan and Malaysia.

3. Motorcycles are dominant in traffic in Vietnam.

A portion of the statistics which they investigated in [22] is shown in Figure 1.2.

In order to introduce jam reduction by driving improvement in these developing countries, it is necessary to evaluate the effects of driving changes on jam reduction. Therefore, as the first step of the evaluation, we definitely need to analyze microscopic phenomena in current mixed traffic by modeling of such mixed traffic.

Figure 1.1: Mixed traffic observed in Mumbai, India.

(a) Case of Taiwan.

(b) Case of Vietnam.

(c) Case of Malaysia.
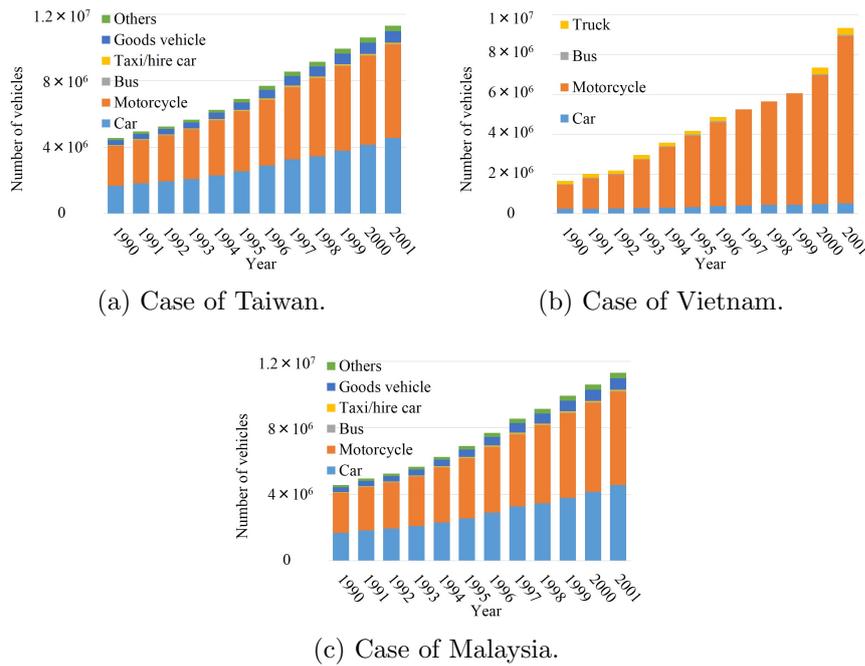
Figure 1.2: Number of vehicles and their types in some Asian countries.

1.2 Previous Studies on Mixed Traffic

In mixed traffic, vehicles of different sizes and performance behave differently and are supposed to cause "heterogeneity" of local density, velocities, accelerations, and locations of each type of vehicle. This heterogeneity is the obstacle that "uniform" traffic models cannot address. To tackle this heterogeneity, researchers are adopting three types of approaches.

1. Analysis of driving behaviors according to physical values, i.e., velocity, acceleration, deceleration, distance gap between vehicles, etc.

2. Analysis of driving behaviors based on parameters of mathematical models which relate to driving behaviors, or analysis of macroscopic characteristics, e.g., stability, relationship between the density and flow, etc., based on simulations by fitted mathematical models

3. Increasing the number of vehicle types in consideration

As an example of Approach 1, Sayer *et al.* clarified that drivers tend to have a shorter distance gap, or space headway, between vehicles when following a heavy vehicle (HV) than that when following a normal passenger car [23]. Sarvi measured the space and time headway in the car-following-car (C-C), car-following-truck (C-T), and truck-following-car (T-C) cases and discovered different trends [24]. Aghabayk *et al.* analyzed the characteristics of acceleration and reaction time for the C-C, C-HV (car-following-heavy vehicle), HV-C, and HV-HV cases. They also focused on the time and space headway. They found that the tendencies for HV-C and C-HV cases were reversed at a velocity of 30 km/h. The largest acceleration deviations were observed in the C-C case, followed in order by C-HV, HV-C, and HV-HV [25]. All these studies considered the combination of a leading vehicle (leader) and a following vehicle (follower) but classified vehicles as either normal cars and trucks or heavy vehicles. Moreover, these studies did not discuss the factors affecting these driving changes.

Although there are some studies on mixed traffic based on macroscopic models [26, 27]; here, we introduce studies based on microscopic models as examples for Approach 2. Munigety *et al.* tried to replicate behavioral difference of various types of vehicles using the spring-mass-damper model and investigated relationship between the flow and velocity in the cases of homogeneous and mixed traffic [28]. Based on the parameter changes of the car-following models, Ossen *et al.* concluded that the desired time headway of a passenger car when following a truck is smaller than that when following another passenger car. They also argued that "robust" car-following is realized by trucks owing to their preference for constant velocity [29]. Furthermore, Chen *et al.* insisted that a stop-and-go wave is attenuated because of the combination of HV-C [30]. They analyzed the parameters of the asymmetric behavioral model [31] of the C-HV, C-C, and HV-C cases. Mason *et al.* analyzed the stability of a system comprising normal passenger cars and trucks using the optimal velocity model [32]. They confirmed that nonlinear waves can be amplified or attenuated based on stability condition extended to a multi-vehicle system. Yang *et al.* analyzed fundamental diagrams and stability with changes in the parameters of the intelligent driver model [33]. They argued that vehicle order in a platoon affects the propagation of a shockwave [34]. In other words, the T-T combination amplified the shockwave, while the C-C combination attenuated it. The proportions of combinations, i.e., C-C, C-T, T-C, and T-T, determined the fundamental diagrams. These studies also classified vehicles into passenger cars and heavy vehicles. In addition, in some of these studies, it was not confirmed sufficiently whether driving differences between these vehicles were properly replicated. This can be attributed to the simplicity of models used in the studies, or because the important driving differences to be replicated were not chosen and the objective functions of model fitting referred only to the mean squared error.

For Approach 3, researchers have tried to introduce small motorized vehicles (e.g., motorcycles and motorized three-wheelers) and non-motorized vehicles (e.g., bicycles, carts, pedestrians, etc). In particular, the characteristic maneuvers of motorcycles have attracted the attention of many researchers. Lee categorized the unique behaviors of a motorcycle and proposed "oblique- and lateral-headway"

and "longitudinal-headway" models to replicate the behaviors of motorcycles. Their main focus was on passing and following behaviors with lateral movement [35]. Wong *et al.* evaluated motorcycles based on a consideration of driving aggressiveness and concluded that these motorcycles should be clearly distinguished from other vehicles in microscopic models from the parameter perspective [36]. Shiomi *et al.* developed a microscopic model to replicate lane-free movements frequently observed in the case of small vehicles [37]. Their model was based on discrete choices for several objectives, e.g., current direction, destination, following other vehicles, collision avoidance, etc. From the estimated coefficients for respective objective functions, they clarified the asymmetry of attentions between passenger cars and motorcycles. In contrast, Oketch focused on not only motorcycles but also other non-motorized vehicles and pedestrians. He simulated traffic with nine parameter-sets of vehicle types and fuzzy rules for lane changes. The model included non-motorized vehicles and pedestrians. He confirmed the consistency of this simulation with real traffic in Nairobi, Kenya [38].

1.3 Objective and Organization of the Thesis

As mentioned above, jam reduction by improvement of driving methods is expected to be an effective solution, especially for some developing countries that can not allocate sufficient resources to traffic improvement. From the literature review in Section 1.2, the flow of research on mixed traffic can be depicted as Figure 1.3. While introducing various types of vehicles, researchers analyzed driving differences and tried to adjust model parameters for the final objectives, i.e., replication and clarification of microscopic phenomena in mixed traffic.
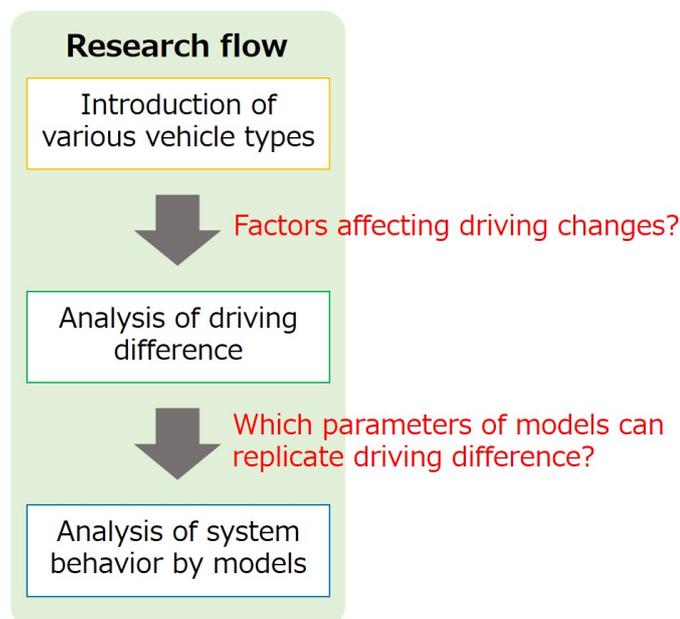
Figure 1.3: Flow of research on mixed traffic.

In this flow, two questions are raised, and answering these questions is the objective of this thesis.

1. What vehicle factors affect driving of respective vehicles?

2. Can we apply driving differences between each type of vehicles into existing models by adjusting parameter values? If so, which parameters should be adjusted to replicate the driving differences?

Answers to Question 1 will help us replicate microscopic phenomena in mixed traffic currently observed in developing countries by taking many new types of vehicles into consideration based on vehicle characteristics. It also helps us understand the vehicle factors that psychologically influence a human driver. Psychological investigation based on this thesis will contribute to the development of an autonomous driving strategy that is comfortable for drivers of various types of vehicles. A comfortable autonomous driving strategy will be constantly essential in the coming eras of completely automated vehicle traffic and half automated traffic, i.e., traffic comprising manual and autonomous driving vehicles. Furthermore, even for current traffic in developed countries, the psychological investigation contributes to understanding driving trends for each type of vehicles. Insights from the investigation can be utilized to foster the self-understanding of drivers in safety training, etc.

Answering Question 2 will provide important parameters in existing models to replicate driving differences between respective vehicles. It also clarifies the performance limitations of models. In other words, we can identify features that cannot be replicated by any parameter when a certain existing model is used. Based on these insights, the limitations of macroscopic analysis of the traffic with some existing models will be also clarified. Furthermore, clarifying the performance of each parameter will be the basis for development of more accurate models replicating microscopic behaviors in mixed traffic.

In order to answer Question 1, we focus on the driving of following vehicles in a platoon comprising one leading vehicle and one following vehicle. In particular, by changing the vehicle types of leaders and followers, we develop multiple regression models of follower driving, with explanatory variables that include not only vehicle characteristics of leaders and followers but also their driving. Multiple regression models with standardized variables enable us to evaluate the effectiveness of various factors as well as determine the factors that are effective.

In order to answer Question 2, we compare trajectory features caused by vehicle types in the experiments and by changing the values of parameters in the models. By adjusting the lengths of time series in experiments and simulations, we clarify the range at which characteristic trajectories were observed in the observation and confirm whether the simulated trajectories were varied by parameter changes in the characteristic ranges.

In this thesis, as with measurements in [39, 40, 41], we did not utilize field data, but instead used trajectories recorded over a test circuit. This research method was chosen since we were primarily aiming to find clear dependencies without

having any disturbances, such as lane changes, that can affect the model. In addition, exploring the cause of driving changes in respective vehicle combinations, which were not discussed in [39, 40, 41], is enabled by detailed information of on respective vehicles utilized in our measurements.

In Chapter 2, we investigate the factors affecting follower driving. In Section 2.1, we observe the trajectories of platoons comprising two vehicles chosen among motorcycles, normal passenger cars, and trucks using GPS antennas attached to the respective vehicles in a test circuit. We also show the tendencies of observed features of followers' driving in respective vehicle combinations. In Section 2.2, while introducing explanatory variables, i.e., characteristics of vehicles, we develop and discuss regression models for respective followers' driving to find the dominant factors affecting driving behavior.

In Chapter 3, we introduce and select traffic models which we will investigate before the comparison of observed trajectories and simulated trajectories in Section 3.1. Then, we proceed with the comparison in Section 3.2. We describe how to extract the characteristic trajectories of each vehicle from the observed trajectories in Section 3.2.2. Features extracted from trajectories will be shown in Section 3.2.3 On the other hand, trajectories simulated by several types of models are introduced in Section 3.3. Finally, we compare the observed and simulated trajectories based on features extracted from observed trajectories. In Chapter 5, we summarize and conclude discussions in the previous chapters.

Chapter 2

Dependence of Driving Characteristics on Vehicles and Drivings in a Platoon

Abstract   In order to investigate the effective factors on follower drivings, we applied the multiple regression analysis and found that

1. the maximum velocity and acceleration of the following vehicle are mainly affected by the drivings of the leading vehicle, and

2. the maximum deceleration, distance gaps when the platoons starts, the maximum distance gap during a trial, and delay in maximum acceleration timing are also affected by the vehicle characteristics of the leading and following vehicles.

From these findings, it is possible to mention that the maximum velocity in the car-following models need not to be modified by vehicle types.

   In addition, psychological effects were implied in the maximum acceleration, deceleration, and distance gaps. While some effects could be understood by safety mind and comfortableness of the drivers, the other ones conflicted with the safety mind.

   Section 2.1 comprises experiment configuration, pretreatment for observed data, introduction of focused features called driving characteristics in data, and introduction of these row features. These row features are analyzed in Section 2.2 by the multiple regression analysis. As preparation for the analysis, we first introduce the vehicle characteristics which can affect the driving characteristics. Following variable selection from the perspectives of multi-collinearity and emergence timings of the driving characteristics, we introduce obtained regression models for each driving characteristics based on Akaike information criterion. We have published these results in [42].

2.1 Experiment

2.1.1 Experiment Configuration

To measure the dependence of the driving characteristics upon vehicles and their drivings in a platoon, we observed the following behavior in platoons comprising two vehicles: a leader and a follower. These vehicles were selected from motorcycles, normal passenger cars, and trucks. The motorcycles were of the 50-cc-scooter type, i.e., Suzuki's Let's G model and Let's 4 model, two Honda's TACT model, one TACT basic model, and one Dio model. The normal passenger cars were represented by a Toyota's Corolla Axio while the trucks were the Isuzu Elf with a container.

We conducted our experiment and data collection at a test course located at the Japan Automobile Research Institute. The test course is comprised of several types of courses including a straight course and a oval circuit as shown in Figure 2.1. In order to obtain a sufficient sample size, we conducted the same experiment over two days. The straight course and the oval circuit were utilized on respective days.

For the straight course shown in Figure 2.1, the first experiment was held in December, 2015. The course configuration is shown in Figure 2.2a. The test area is a straight road that includes a braking area. The start cone indicates where the leader commences acceleration; the braking cone indicates where the leader commences deceleration. The stop cone indicates where the leader stops. We provided driving instructions to the leaders and followers, as shown in Figure 2.2b. The leaders were instructed to start their acceleration from the start cone while imagining that they were driving on public roads as if commuting. After reaching 60 km/h, they maintained a velocity of 60 km/h until they reached the braking cone. We also instructed the leader to start braking from the braking cone in order to stop at the stop cone with constant braking. The length of the braking areas were set to 70 m and 116 m in order to vary the decelerations of the leaders. On the other hand, we instructed followers only to follow the leaders while imagining that they were commuting. Every 0.1 s, the positions of the respective vehicles were recorded to an accuracy of 60 cm by GPS antennas, namely Hemisphere A100, on top of the vehicles. The positions of the GPS antennas were 1.5 m, 2.2 m and 1.0 m from front bumpers of motorcycles, passenger cars, and truck, respectively. Their positions were considered when calculating the vehicle positions in Section 2.1.2. All measured data were recorded to tablets located at respective vehicles.

For the oval circuit shown in Figure 2.1, the second experiment was held in September, 2016. We divided the oval circuit into Sections A to I, as shown in Figure 2.1. The sections marked in red arrows were used for measurement since Section F and Section I include sags and small curves. Figure 2.3 shows the internal configuration of each section. The braking cones and stop cones have the same purpose as discussed above in Figure 2.2. The stop cone and line in a section become start indication for the next section. In Table 2.1, the length of test sections L and braking area L_b are listed. The driving instructions to

leaders and followers were as for the previous experiment. We instructed leaders to maintain a velocity of 50 km/h until they reached the braking cone because of the curves and length limitation of the sections. In addition, we instructed the leaders regarding the acceleration quickness as shown in Table 2.1 in order to vary accelerations of the leaders. The lengths of the braking area were also varied to obtain various deceleration data for the leaders.

In Table 2.2, the ages of all drivers are listed. They were all males and had at least three years of driving experience of the vehicles which we assigned. Vehicles which they drove respectively were indicated as the "driving vehicle" in Table 2.2. The designators MC, C, and T represent motorcycles, passenger cars, and trucks, respectively. We called the drivers from TD1 to TD7 as the test drivers. The drivers numbered from #8 to #15 were called the leading drivers. Note that Driver #10 and Driver #11 are the same drivers to TD6 and TD7, respectively.

In Table 2.3, the tested combinations and trial numbers are listed. A total of 123 trajectories were collected over both days.
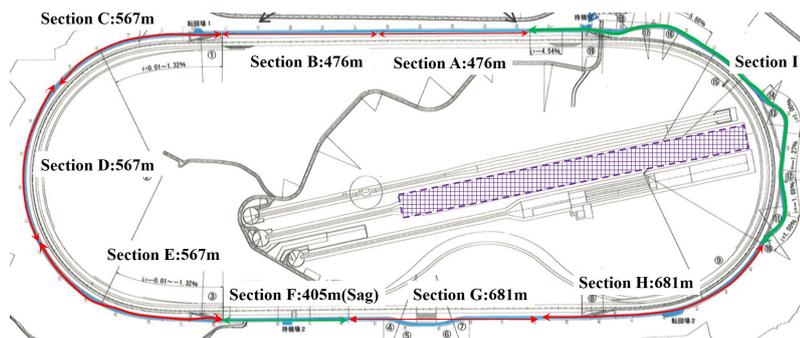
Figure 2.1: Test courses of the Japan Automobile Research Institute. The straight course is denoted by the purple cross stripes and the oval circuit is marked by the red and green arrows.
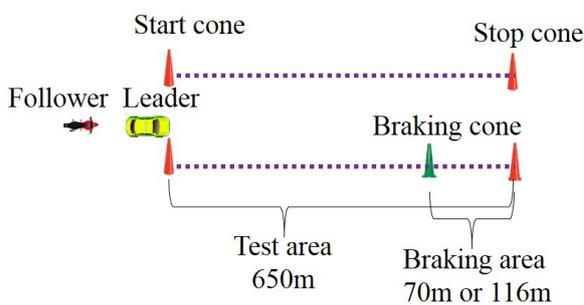
(a) A schematic of the test course.
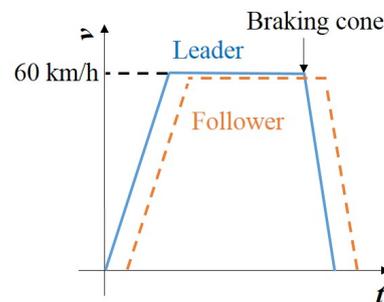
(b) Velocity profiles of a leader and a follower.

Figure 2.2: Configuration of the experiment in December, 2015.
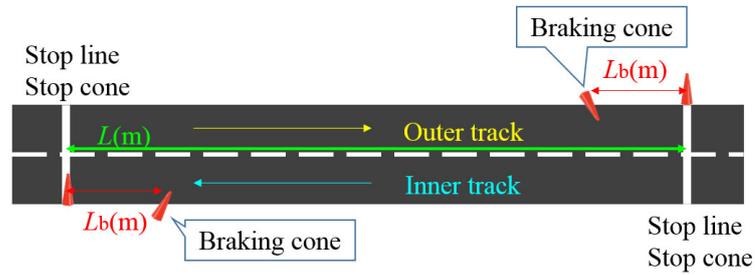
Figure 2.3: Configuration of the test course used in the experiment in September, 2016.

Table 2.1: Lengths of the respective sections. Acceleration instructions and expected decelerations of the leaders are also indicated.

Section name (acceleration instruction to the leaders)	L (m)	L_b (m) (expected deceleration of the leaders)
A (Accelerate quickly)	467	82 (1.2 m/s^2)
B (Accelerate quickly)	467	82 (1.2 m/s^2)
C (Accelerate normally)	567	100 (1 m/s^2)
D (Accelerate normally)	567	100 (1 m/s^2)
E (Accelerate normally)	567	100 (1 m/s^2)
G (Accelerate slowly)	681	196 (0.5 m/s^2)
H (Accelerate slowly)	681	196 (0.5 m/s^2)

Table 2.2: Ages of drivers.

	Name	Driving vehicle	Age	Experiment day
Test driver (follower)	TD1	MC	22	Dec., 2015
	TD2	C	22	
	TD3	T	35	
	TD4	C, T	43	Sep., 2016
	TD5	C	37	
	TD6	MC	23	
	TD7	MC	54	
Leading driver	#8	MC	23	Dec., 2015
	#9	MC	22	
	#10 (TD6)	MC	23	Sep., 2016
	#11 (TD7)	MC	54	
	#12	C	23	Dec., 2015
	#13	C	44	Sep., 2016
	#14	T	44	Dec., 2015
	#15	T	42	Sep., 2016

Table 2.3: Measured vehicle order and the number of measured trial.

Follower	Leader	Experiment day	Test driver (follower)	Trial number	Trial number of combination
MC	MC	Dec., 2015	TD1	2	9
		Sep., 2016	TD6	7	
	C	Dec., 2015	TD1	2	15
		Sep., 2016	TD6	13	
	T	Dec., 2015	TD1	2	7
		Sep., 2016	TD7	5	
C	MC	Dec., 2015	TD2	2	15
		Sep., 2016	TD4	6	
			TD5	7	
	C	Dec., 2015	TD2	2	18
		Sep., 2016	TD4	6	
			TD5	10	
	T	Dec., 2015	TD2	2	24
		Sep., 2016	TD4	10	
			TD5	12	
T	MC	Dec., 2015	TD3	4	10
		Sep., 2016	TD4	6	
	C	Dec., 2015	TD3	4	10
		Sep., 2016	TD4	6	
	T	Dec., 2015	TD3	4	15
		Sep., 2016	TD4	11	

2.1.2   Data Pretreatment

We applied linear interpolation to the position data measured from GPS antennas because these positions had missing values. Due to the accuracy and interpolation, the velocities and accelerations obtained from the position data were not sufficiently smooth to obtain the driving characteristics. Therefore, using the Python library package, pykalman, we applied a Kalman smoother to the velocities and differentiated the velocities to obtain the accelerations. The Kalman smoother estimates time series data from all the observed data [43]. In Figure 2.4a, the observed positions for a certain measurement period are indicated by blue dots. The velocities acquired from the differentiation of these positions are indicated by blue dots in Figure 2.4b. There are constant velocities caused by linear interpolation and discontinuous velocities caused by GPS accuracy. The velocity smoothed by the Kalman smoother is shown by a red line in Figure 2.4b, and the position and acceleration calculated from the smoothed velocity are shown by a red line in Figure 2.4a and 2.4c. The velocity became sufficiently smooth to be differentiated and the positions calculated from the smoothed velocities agreed well with their original values.

(a) Time evolution of the position data.

(b) Time evolution of the velocity data.

(c) Time evolution of the acceleration data.
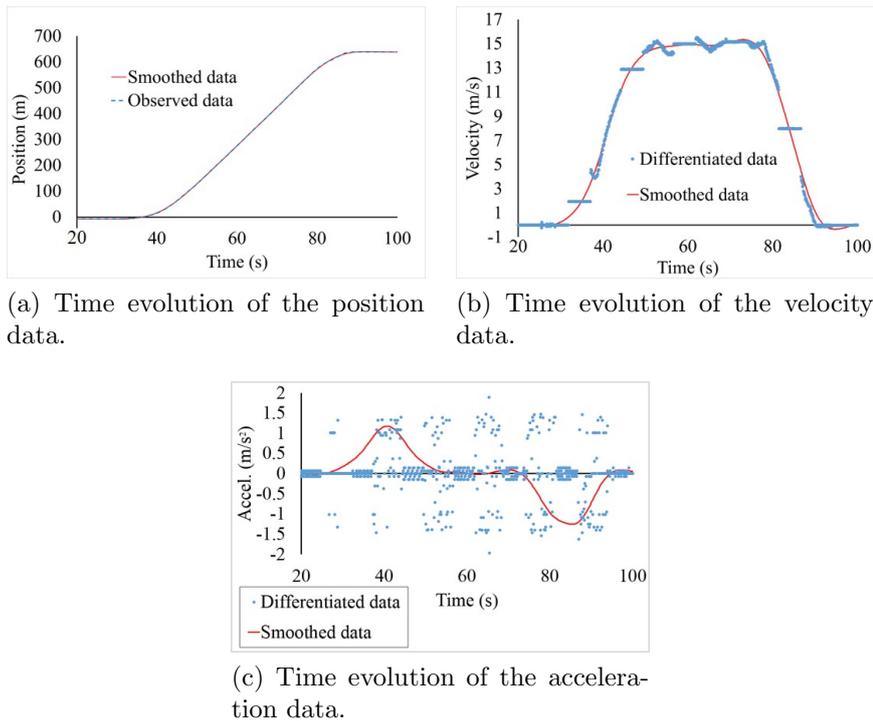
Figure 2.4: Comparison of Kalman-smoothed and observed data.

In addition, we discuss distance gaps, i.e., distance between vehicles, in following sections. The distance gaps calculated from measured positions in the oval circuit should be corrected because there are some sections containing curves. As shown in Figure 2.5, we utilized the total trajectory length for the leader between point P and Q. Where P is the nearest point to current front bumper of the fol-

lower on the leader's trajectory and Q is the real-time position of the rear bumper of the leader. Since the corrected distance gaps contained some noise, we applied FIR low-pass filter, which cut-off frequency was 0.67 Hz and tap number was 91. Characteristic distance gaps discussed in following sections were collected from the smoothed distance gaps.
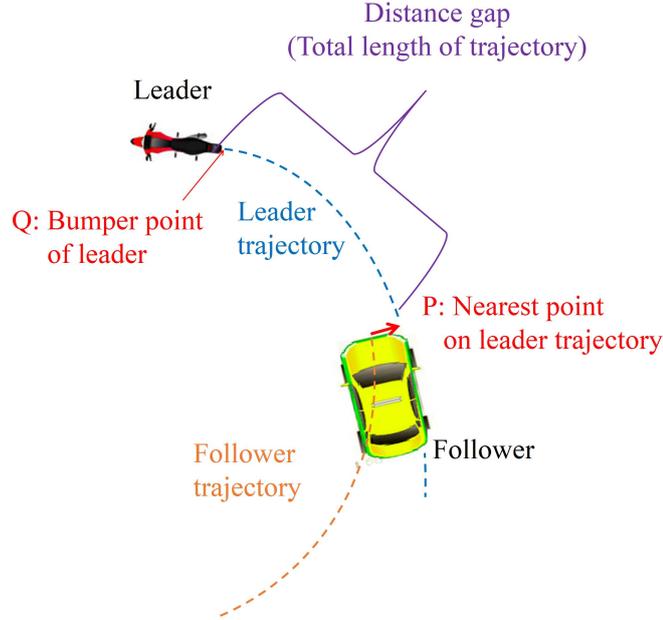
Figure 2.5: Distance gap calculation in a corner.

2.1.3   Introduction of Driving Characteristics

As objective and explanatory variables of the regression analysis, we introduce some characteristic values of the velocity and acceleration, as shown in Figure 2.6. $V_{\max}$ is the maximum velocity while $a_{\max}$ and $a_{\min}$ are the maximum and minimum accelerations, respectively. To determine the delay of the velocity trajectories between a leader and a follower, we introduce ϵ_{a}, which represents the time difference between the leader's and follower's $a_{\max}$ values. These characteristics were extracted from the smoothed data by the Kalman smoother as discussed in Section 2.1.2.

We also introduce S_{start} as the distance gap between rear bumper of the leader's vehicle and front bumper of the follower's vehicle when both vehicles are at their starting points. The $S_{\max}$ is the maximum distance gap when they are moving. These characteristics related to the distance gaps were extracted from the smoothed distance gap by the FIR low-pass filter as discussed in Section 2.1.2.

Note that we utilize the superscripts of "f" and "l" to indicate followers and leaders, respectively. For example, $a_{\max}^{\mathrm{l}}$ indicates the leader's maximum acceleration and $V_{\max}^{\mathrm{f}}$ indicates the follower's maximum velocity. The ϵ_{a} and distance gaps do not have these superscripts because they are determined by relationship between the leader and the follower.
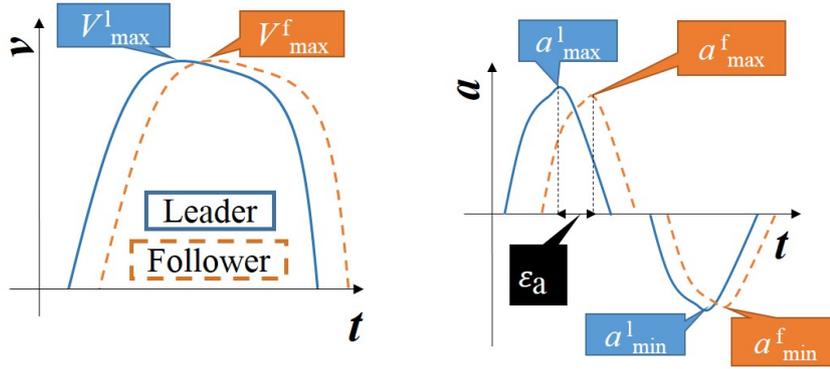
(a) Driving characteristics determined on the velocity evolution.

(b) Driving characteristics determined on the acceleration evolution.

Figure 2.6: Schematics of the driving characteristics.

2.1.4 Driving Characteristics Measured in Various Vehicle Combinations

In this section, we compare driving characteristics in various vehicle combinations. In Figure 2.7, the $V_{\mathrm{max}}^{\mathrm{f}}$ values for each vehicle combination are shown by the colored boxes. M, C and T stand for the motorcycle, car, and truck. The left boxes marked in checkered pattern in Figure 2.7a indicate the average values of three colored boxes on the right side of them. For example, the nearest blue plaid box in Figure 2.7a indicates the average $V_{\mathrm{max}}^{\mathrm{f}}$ value when the motorcycle is the follower. The outermost yellow boxes in Figure 2.7a show the average of all the followers, motorcycles, cars, and trucks. For example, the center yellow box indicates the average $V_{\mathrm{max}}^{\mathrm{f}}$ when the leader is the car. Figure 2.7b and 2.7c are extracted views from Figure 2.7a. To compare the average $V_{\mathrm{max}}^{\mathrm{f}}$ value to the $V_{\mathrm{max}}^{\mathrm{l}}$ value in each leader case, we plotted them in Figure 2.7c.

From Figure 2.7b, the $V_{\mathrm{max}}^{\mathrm{f}}$ value was the greatest when the follower was the truck (p-value < 0.10 in Dunnett's test). This implies that the trucks had tendency to have higher velocity than other types of vehicles. On the other hand, $V_{\mathrm{max}}^{\mathrm{f}}$ when the motorcycle was the leader was smaller than remaining cases shown in Figure 2.7c (p-value < 0.005 in Dunnett's test). As the yellow boxes and the orange dotted line in Figure 2.7c obey similar trends, it is possible that $V_{\mathrm{max}}^{\mathrm{f}}$ is affected by $V_{\mathrm{max}}^{\mathrm{l}}$. This hypothesis is natural because the platoon is in the car-following phase. Note that it is also possible that the yellow boxes in Figure 2.7c will obey a certain trend because the followers are affected by the leader types.

(a) $V_{\max}^{\mathrm{f}}$ for all vehicle combinations.

(b) $V_{\max}^{\mathrm{f}}$ for each follower case.

(c) Comparison of $V_{\max}^{\mathrm{f}}$ and $V_{\max}^{\mathrm{l}}$ for each leader case.
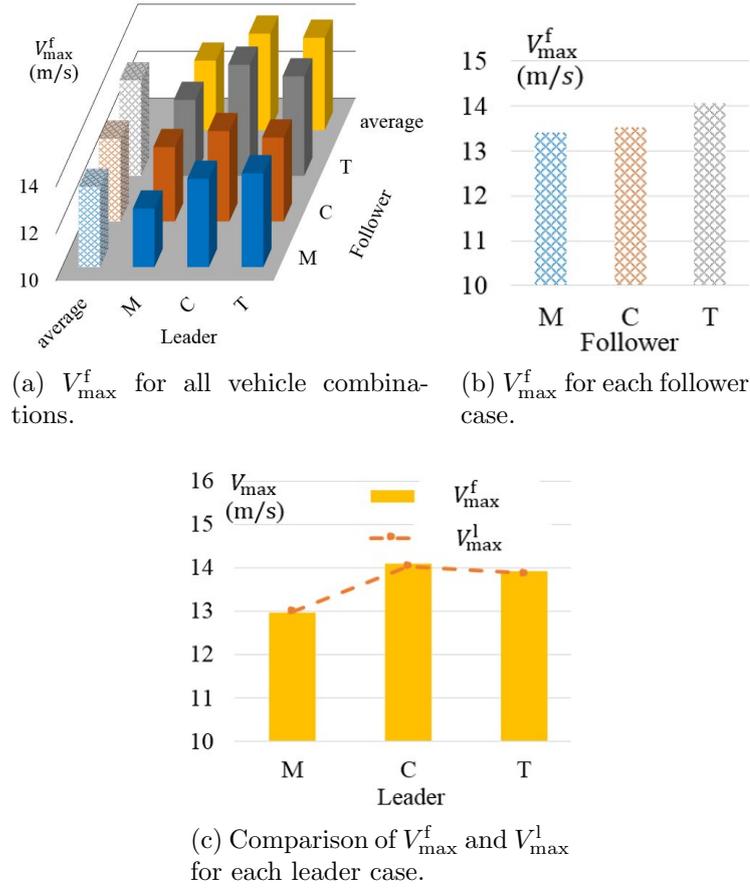
Figure 2.7: The maximum velocities of followers, i.e., $V_{\max}^{\mathrm{f}}$.

In Figure 2.8 and 2.9, $a_{\max}^{\mathrm{f}}$ and $a_{\min}^{\mathrm{f}}$ are shown for each vehicle combination. The $a_{\max}^{\mathrm{f}}$ value was the smallest value when the follower was the truck in Figure 2.8b. The p-values in Dunnett's were less than 0.05. From Figure 2.8c, it can be observed that $a_{\max}^{\mathrm{f}}$ is not affected by leaders' types. The p-value was 0.78 in one-way ANOVA. Since the plaid colored boxes in Figure 2.8b obey a similar trend as with the dotted line in Figure 2.8c, the trucks may originally have a smaller acceleration than the motorcycles and cars. This trend may be caused by the weight of the trucks.

Regarding $a_{\min}^{\mathrm{f}}$, average values to respective followers, i.e., the plaid colored boxes in Figure 2.9b did not have significant difference (p-value = 0.31 in one-way ANOVA). On the other hand, in Figure 2.9c, $a_{\min}^{\mathrm{f}}$ had the greatest value when the leader was the car (p-value < 0.05 in Dunnett's test) and had the same trends to $a_{\min}^{\mathrm{l}}$. Although this implies that $a_{\min}^{\mathrm{f}}$ is affected from $a_{\min}^{\mathrm{l}}$ or from the vehicle types of the leaders, the primary effects are still unclear. It will be clarified by the multi-regression analysis in Section 2.2.2.
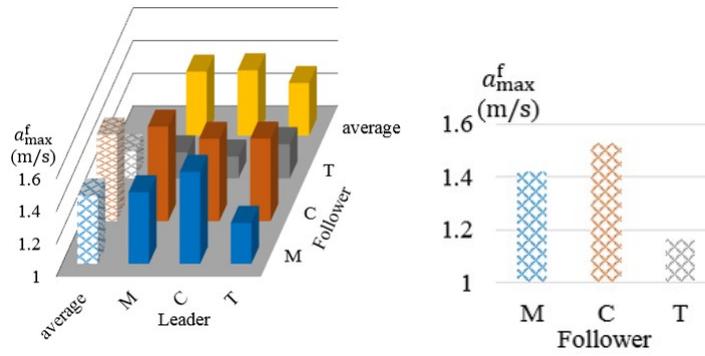
(a) $a_{\max}^{\mathrm{f}}$ for all vehicle combinations.

(b) $a_{\max}^{\mathrm{f}}$ for each follower case.
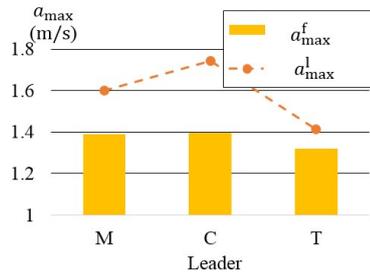
(c) Comparison of $a_{\max}^{\mathrm{f}}$ and $a_{\max}^{\mathrm{l}}$ for each leader case.

Figure 2.8: The maximum accelerations of followers, i.e., $a_{\max}^{\mathrm{f}}$.
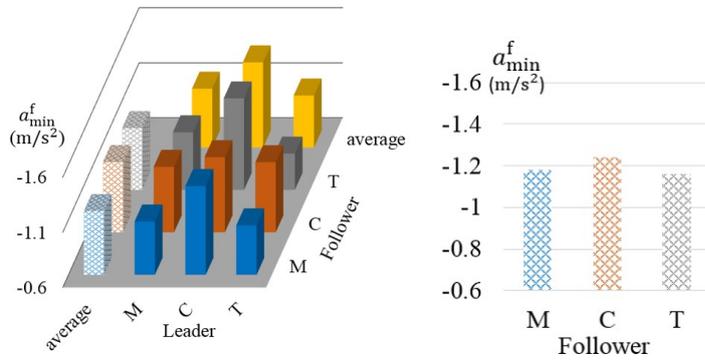
(a) $a_{\min}^{\mathrm{f}}$ for all vehicle combinations.

(b) $a_{\min}^{\mathrm{f}}$ for each follower case.
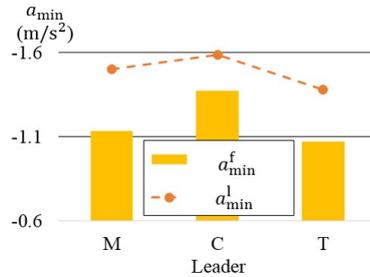
(c) Comparison of $a_{\min}^{\mathrm{f}}$ and $a_{\min}^{\mathrm{l}}$ for each leader case.

Figure 2.9: The maximum decelerations of followers, i.e., $a_{\min}^{\mathrm{f}}$.

Based on the data shown in Figure 2.10b, we found that the values of S_{start} became small when the follower was the motorcycle (p-value < 0.001 in Dunnett's test). It would be because the following drivers on the motorcycles could percept their vehicle size more easily than other types of vehicles. In Figure 2.10c, we did not recognize a clear relationship between the leaders' types (p-value $= 0.14$ in one-way ANOVA).

(a) S_{start} for all vehicle combinations.

(b) S_{start} for each follower case.

(c) S_{start} for each leader case.
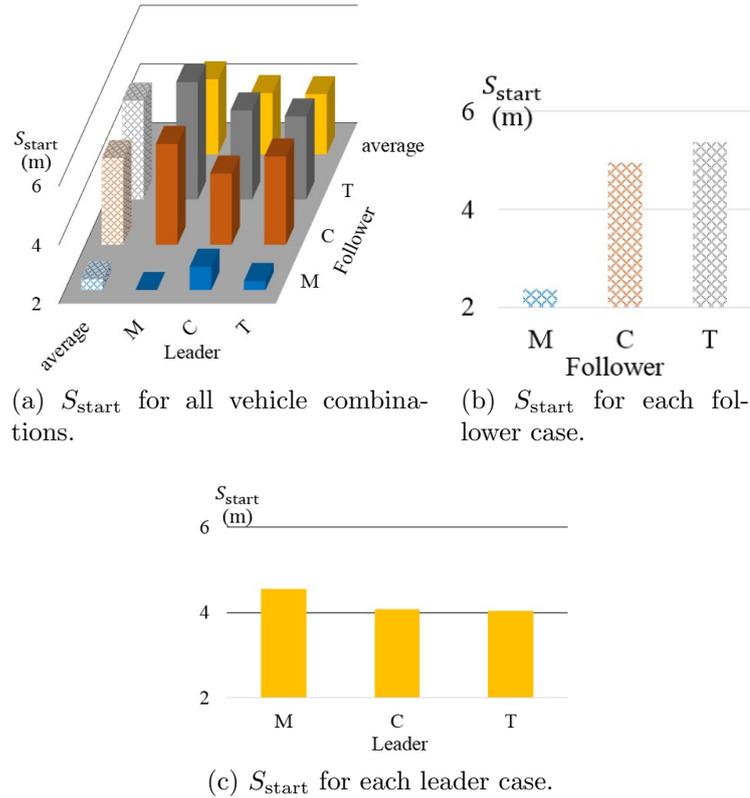
Figure 2.10: The distance gaps when the platoons start, i.e., S_{start}.

From Figure 2.11b and 2.11c, we could find the values of S_{max} became small when the follower was the motorcycle (p-value < 0.001 in Dunnett's test) and could not recognize a clear relationship between the leaders' types (p-value $= 0.60$ in one-way ANOVA). Although these trends are similar to those of S_{start}, we are not sure if these trends were caused by the same factors of vehicles.

In Figure 2.12b, the ϵ_{a} value is the largest when the follower is the truck, followed by the car and motorcycle. The p-values of less than 0.01 in Steel-Dwass' test were obtained between the followers. From Figure 2.12c, The ϵ_{a} value was the smallest when the leader was the truck. The p-value was less than 0.001 in Dunnett's test. These trends are natural because operational delays would be small when the followers are agile or the leaders are unagile.

From driving characteristics in respective combinations, we observed several trends of them and gave some guesses based on vehicle types. In Section 2.2.2, we try to clarify the affecting factors of vehicles and drivings to respective driving characteristics by multiple regression analysis.

(a) $S_{\max}$ for all vehicle combinations.

(b) $S_{\max}$ for each follower case.

(c) $S_{\max}$ for each leader case.
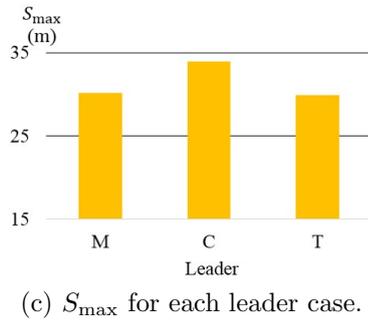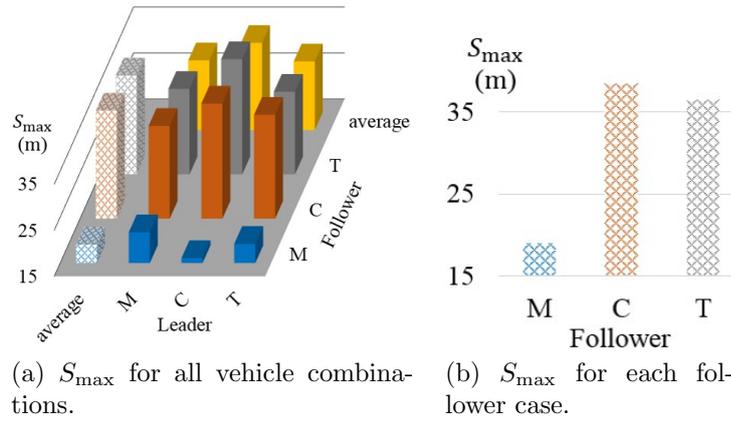
Figure 2.11: The maximum distance gaps in each trial, i.e., $S_{\max}$.

(a) ϵ_{a} for all vehicle combinations.

(b) ϵ_{a} for each follower case.

(c) ϵ_{a} for each leader case.
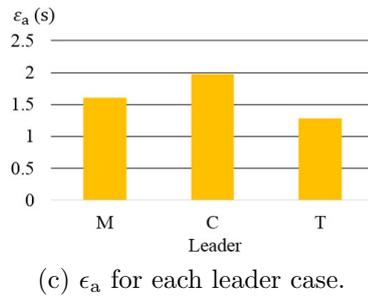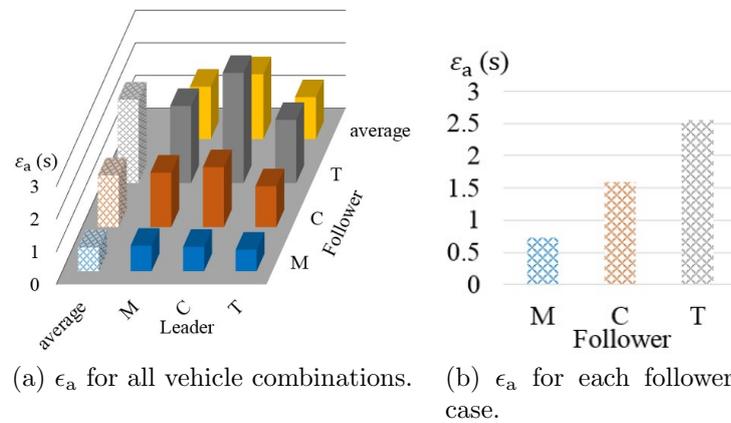
Figure 2.12: Time delays between the maximum accelerations of leaders and followers, i.e., ϵ_{a}.

2.2 Detailed Factors Affecting the Driving Characteristics of the Followers

In Section 2.1.4, even when some characteristics appeared to be dependent upon the vehicle type of a follower and a leader, these characteristics did not necessarily follow the instinctive size order (i.e., the motorcycle, car, and truck). In other words, we found no clear insight into what factors other than the instinctive order of the motorcycle, passenger car, and truck would affect the respective driving characteristics.

It is also possible that some driving characteristics of a leader and a follower as well as both of their vehicle characteristics would affect the related driving characteristics of a follower. Therefore, we applied multiple regression analysis to the driving characteristics of the followers. The explanatory variables are the driving characteristics of the leaders and followers, and the vehicle characteristics related to the size and the engine performance.

In this section, we introduce several vehicle characteristics and select representatives from the perspective of multi-collinearity. Then, we specify candidates for the explanatory variables from the perspective of emergent timings. The candidates for the explanatory variables depend upon objective variables, i.e., target driving characteristics. Finally, we develop and discuss multiple regression models of the respective driving characteristics.

2.2.1 Introduction and Selection of Vehicle Characteristics

To clarify the characteristics of each vehicle, we consider the features listed in Table 2.4. We add passenger's weight of 55 kg to the vehicle weight since the total weight of a motorcycle changes drastically when carrying a rider. The weight of 55 kg is the standard weight for the total weight of a vehicle full of passengers as calculated by Japanese motor companies [44]. The top height is the total height including a passenger, i.e., the same as the regular height in the case of a passenger car and truck. We introduce this value because the observed height of a leading motorcycle drastically changes when supporting a rider. It is also useful to consider the riders' eye height on the following motorcycles. In the case of motorcycles, we calculate the top height as the seat height + 0.92 m based on the age of riders in the experiment. The value of 0.92 m is the average seated height of a Japanese male in their 20s to 50s between 2004 and 2006 [45]. Driver TD6 and TD7 and Drivers #8 and #9, who rode motorcycles in the experiments, were their 20s to 50s. These characteristics of the leaders and followers are defined as the explanatory variables for the multiple regression process with the exception of the power-to-weight ratio (PWR) and torque-to-weight ratio (TWR) of the leaders because the following drivers would not perceive these characteristics.

When we utilize multiple regression analysis, we should avoid the problem of multicollinearity caused by the explanatory variables for which a high correlation is observed. Variables having a correlation above 0.8 are bound by the black

lines in Figure 2.13, from which we find that they are roughly classified into three categories: vehicle height, vehicle size, and engine performance. From these categories, we selected the top height, length, and PWR since they are less correlated with the variables in other categories. These selected variables are used in following multiple regression analysis as explanatory variables for vehicle characteristics. Note that these three variables are the representatives of three categories. For example, if the length is one of the explanatory variables of a regression model in Section 2.2.2, the model does not directly indicate that the length is the affecting factor, but indicates that vehicle size is the affecting factor.

Table 2.4: Definitions of vehicle characteristics and values for each vehicle. SD stands for the standard deviation.

Name	Meaning	MC (average) (SD)	C	T
Width (m)	Vehicle width	0.64 (0.030)	1.7	1.89
Length (m)	Vehicle longitudinal length	1.67 (0.006)	4.3	4.9
Weight (kg)	Vehicle weight + one passenger (55 kg)	128 (5.4)	1145	2605
Height (m)	Vehicle height	1.01 (0.022)	1.46	2.81
Top height (m)	Same to height in case of C and T. In case of MC, seat height + 0.92 m.	1.62 (0.013)	1.46	2.81
Power-to-weight ratio (PWR; kW/kg)	Max. power divided by weight	0.028 (0.0056)	0.07	0.042
Torque-to-weight ratio (TWR; N·m/kg)	Max. torque divided by weight	0.036 (0.0109)	0.119	0.144
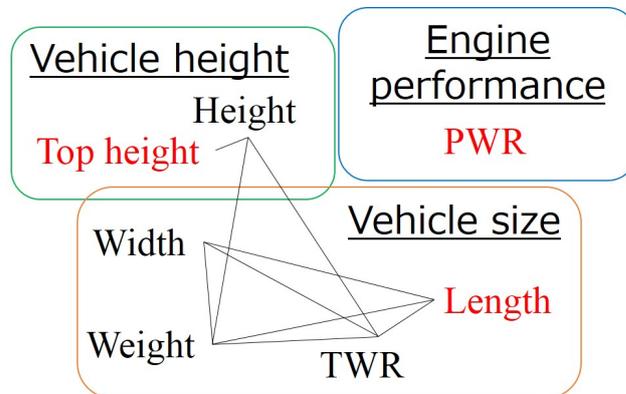
Figure 2.13: Relationship between vehicle characteristics. The black lines indicate correlations above 0.8. Characteristics can be classified into three categories. The red labels are representatives of these categories.

2.2.2 Multiple Regression Analysis of Driving Characteristics

In this section, we apply the multiple regression analysis to the respective driving characteristics of the followers. First, we clarify which characteristics would be candidates for the explanatory variables. Then, we proceed with multiple regression analysis of the respective driving characteristics. Based on the Akaike information criterion (AIC) [46], we chose the best regression models for the respective objective driving characteristics and discuss effective vehicle and driving factors indicated by these models.

2.2.2.1 Candidates of explanatory variables for the respective driving characteristics of followers

In Table 2.5, the candidate explanatory variables for multiple regression of respective driving characteristics are listed. The designator "c" indicates that the characteristic is a candidate of exploratory variables. The designator "—" means that it is not the candidate.

In Table 2.5, the vehicle characteristics of the leaders include leader length and top height. The leader PWR is not included because the follower cannot recognize this parameter. Follower-vehicle characteristics include follower PWR as well as length and top height since follower PWR would directly affect the follower's driving. The day of experiment and driver are dummy variables to remove effects of experiment environment and drivers' tendencies. Because driving tendency of leading drivers are included in leaders' driving characteristics, the dummy variables for drivers are only to test drivers, i.e., TD1 to TD7. These vehicle characteristics and dummy variables are candidates for the explanatory variables for all objective driving characteristics.

Respective driving characteristics are affected by different variables due to their emerging timing. S_{start} would be affected by only the vehicle characteristics, experiment days, and drivers because no driving sequence has yet been conducted. The value ϵ_{a} is affected by S_{start} as well as vehicle characteristics. The values $a_{\text{max}}^{\text{f}}$, $V_{\text{max}}^{\text{f}}$, and S_{max} can be affected by S_{start}, ϵ_{a}, $a_{\text{max}}^{\text{l}}$, and $V_{\text{max}}^{\text{l}}$ because the following drivers seem to decide their velocity and acceleration based on their leaders' velocities and accelerations after a certain delay. Because the following driver commences deceleration and stops finally after the maximum velocity $V_{\text{max}}^{\text{f}}$ and the maximum distance gap S_{max} are observed, we can presume that $a_{\text{min}}^{\text{f}}$ is not affected by S_{start}, $a_{\text{max}}^{\text{l}}$, or $a_{\text{max}}^{\text{f}}$.

Table 2.5: Candidate characteristics for objective variables of multiple regression. "c" indicates that the left characteristic is the candidate. "—" indicates that the left characteristic is not the candidate.

		Objective variables (analyzed driving char.)					
		S_{start}	ϵ_{a}	$a^{\text{f}}_{\text{max}}$	$V^{\text{f}}_{\text{max}}$	S_{max}	$a^{\text{f}}_{\text{min}}$
	S_{start}	—	c	c	c	c	—
	ϵ_{a}	—	—	c	c	c	c
	$a^{\text{l}}_{\text{max}}$	—	—	c	c	c	—
	$V^{\text{l}}_{\text{max}}$	—	—	c	c	c	c
Candidate of explanatory variables	$V^{\text{f}}_{\text{max}}$	—	—	—	—	—	c
	$a^{\text{f}}_{\text{max}}$	—	—	—	—	—	—
	S_{max}	—	—	—	—	—	c
	$a^{\text{l}}_{\text{min}}$	—	—	—	—	—	c
	$a^{\text{f}}_{\text{min}}$	—	—	—	—	—	—
	Leader vehicle char.						
	Follower vehicle char.				c		
	Experiment day						
	Driver						

2.2.2.2 Interpretation of respective regression models

In this section, we calculate the partial regression coefficients for the best models of the respective objective variables.

Models discussed in this section were chosen from the perspective of AIC [46]. AIC is calculated as follows.

$$AIC = -2 \ln \hat{L} + 2k \tag{2.1}$$

$\ln \hat{L}$ is the maximum value of log likelihood and k is the number of the parameters. The value of AIC becomes small when the model fits to observed data well with small number of parameters. Although we discuss the best model from the AIC perspective, the high-ranking models, which AICs are neither equal nor larger than the best AIC + 1.5 ($\Delta AIC < 1.5$), are also checked to confirm if trends of parameter selection and signs of coefficients are not exceptional.

In Figure 2.14a and 2.14b, the best regression models for $V_{\max}^{f}$ and $a_{\max}^{f}$ are shown. Variables located at the starting points of arrows are explanatory variables. Variables located at the end point of the arrows are objective variables. The numbers indicated next to or under the explanatory variables are partial regression coefficients for respective variables. Symbols accompanied by the coefficients indicate the significance level. "◇" means that the significance level is less than 0.001, "‡" means that it is less than 0.01, "†" means that it is less than 0.05, and "•" means that it is less than 0.1.
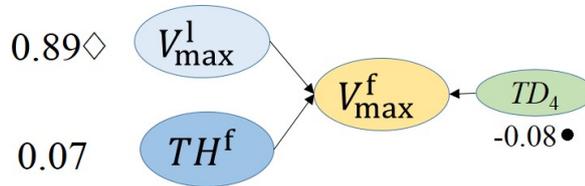
(a) The best multiple regression model for $V_{\max}^{f}$.
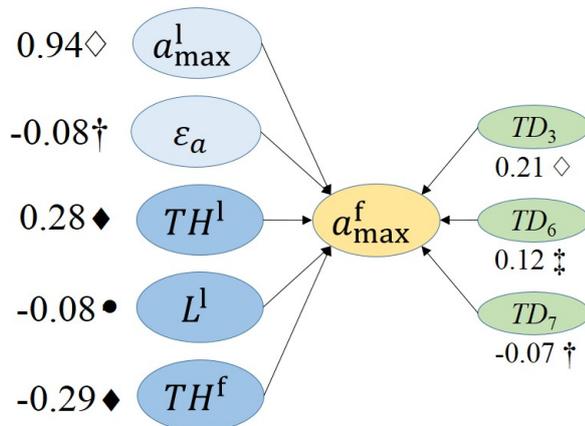
(b) The best multiple regression model for $a_{\max}^{f}$.

Figure 2.14: The best models for followers' velocity and acceleration from the AIC perspective. These models are mainly affected from leaders' drivings.

$V_{\mathrm{max}}^{\mathrm{f}}$ in Figure 2.14a is determined primarily by $V_{\mathrm{max}}^{\mathrm{l}}$ and $R^2 = 0.80$. Thus, the yellow colored boxes and the orange line in Figure 2.7c exhibit the same trends. In the car-following model, we need not change the parameters determining the maximum velocity based on the vehicle types as long as the speed of a traffic stream does not exceed the performance of each vehicle. For reference, we show the non-standardized model below.

$$V_{\mathrm{max}}^{\mathrm{f}} = 0.93 V_{\mathrm{max}}^{\mathrm{l}} + 0.16 TH^{\mathrm{f}} - 0.20 TD_4 + 0.72 \qquad (2.2)$$

The units for the velocity, the top height are (m/s) and (m), respectively. As TD_4 is the dummy variable for TD4 and take 0 or 1, we can conclude that the velocity variation caused by TD4 was 0.20 m/s: when the TD4 was the following driver, the follower velocity became 0.20 m/s smaller than the average of the other cases.

Regarding $a_{\mathrm{max}}^{\mathrm{f}}$ shown in Figure 2.14b, the main factor which determines $a_{\mathrm{max}}^{\mathrm{f}}$ is $a_{\mathrm{max}}^{\mathrm{l}}$. The value of R^2 is 0.87. Indeed, the model also includes the vehicle characteristics of a leader and a follower, the degree of effectiveness of them is one third, at most. The model claims that the maximum acceleration of followers is determined basically by the leaders' acceleration but additionally by the top heights of both the follower's and leader's vehicle, the followers' engine performance seems not to be important factor though. Regarding additional effect from the followers' height, the following drivers accelerate gently when the following drivers' eye point is high. It might be related to the comfortableness when the drivers experience the acceleration. This tendency is consistent with the plaid colored boxes in Figure 2.8b. Regarding additional effect from the leaders' height, following drivers accelerate gently when the leader top height is low. The model seems to replicate the situation that the height differences between yellow colored boxes in Figure 2.8c are not as large as those of the leaders' accelerations shown in the orange dotted line. The non-standardized model of $a_{\mathrm{max}}^{\mathrm{f}}$ is as follows.

$$\begin{aligned} a_{\mathrm{max}}^{\mathrm{f}} = {}& 0.73 a_{\mathrm{max}}^{\mathrm{l}} - 0.024 \epsilon_a + 0.21 TH^{\mathrm{l}} - 0.028 L^{\mathrm{l}} \\ & - 0.23 TH^{\mathrm{f}} + 0.33 TD_3 + 0.15 TD_6 - 0.18 TD_7 + 0.36 \qquad (2.3) \end{aligned}$$

The units for the acceleration, ϵ_a and the length are (m/s^2), (s) and (m), respectively. Drivers TD3 and TD6 tended to have larger acceleration than other drivers while Driver TD7 driving the motorcycle had smaller acceleration. Note that this tendency was extracted from the data which other vehicle and driving characteristics had been already extracted. For example, the maximum acceleration when the follower was the motorcycle became 0.18 m/s^2 smaller than expected one when the driver was TD7.

In Figure 2.15, we show the bar charts of regression coefficients included in high-ranking regression models for $V_{\mathrm{max}}^{\mathrm{f}}$ and $a_{\mathrm{max}}^{\mathrm{f}}$. Each color bar indicates one regression model. The models within $\Delta AIC < 1.5$ are shown from left to right in order of AIC ranking. In other words, the left most model is the best model explained above. Each colored box represents explanatory variable. If a certain explanatory variable has the positive regression coefficient, the colored box is stacked above zero. If the coefficient is negative, the colored box is stacked under zero. In both bar charts, although there are some variations in variable selection,

we can confirm that the leaders' velocity and acceleration are the main factors affecting those of followers.

(a) The models for V_{max}^f within $\Delta AIC < 1.5$ from the best model.

(b) The models for a_{max}^f within $\Delta AIC < 1.5$ from the best model.
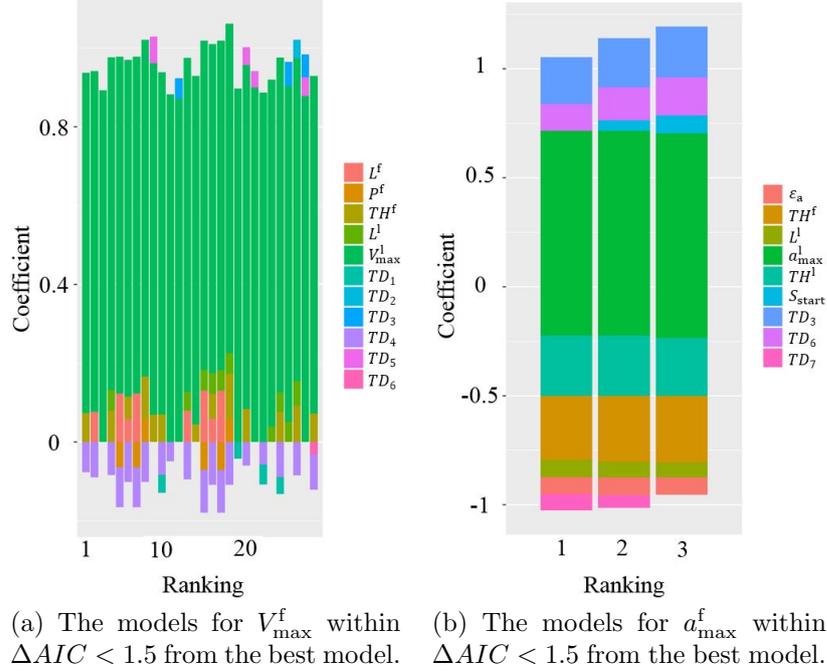
Figure 2.15: The coefficients of high-ranking multiple regression models for the velocity and acceleration.

Different from V_{max}^f and a_{max}^f, the effectivenesses of vehicle characteristics are as strong as a_{min}^l in the high-ranking models of a_{min}^f shown in Figure 2.16a. TH^l, TH^f, L^l, and a_{min}^f are included in most of the models. Figure 2.16b is the best model from the AIC perspective. Although the value of $R^2 = 0.27$, the trend of the variable selection is not extraordinary when we compare the best model with the others. The best model implies that the maximum deceleration becomes mild when the leaders' and followers' heights are high. Height appearance and visibility from the drivers' seat would stimulate safety mind of drivers. On the other hand, the model also implies that the deceleration becomes strong when the size of the leading vehicle becomes large. We can argue that we should consider the types of the following and leading vehicle when the parameters determining the maximum deceleration are calculated. Note that, from the value of R^2, there is the possibility that the randomness or other factors affect the followers' deceleration. If other characteristics of the vehicles affected, the effects occur as those by test drivers of each vehicle. From this reason, the small value of R^2 might be caused the randomness of human operation. In any case, the affecting vehicle and driving characteristics are as described above.
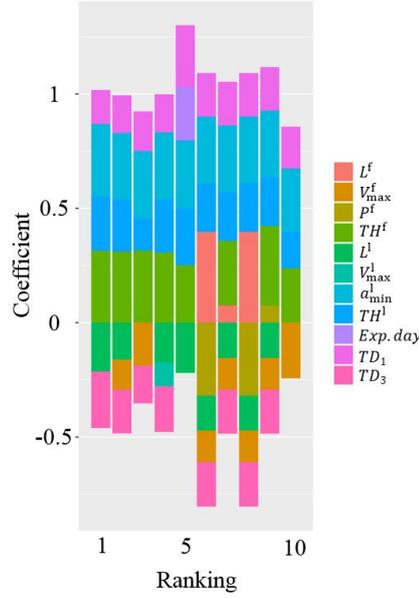
(a) The models for $a_{\min}^{f}$ within $\Delta AIC < 1.5$ from the best model.
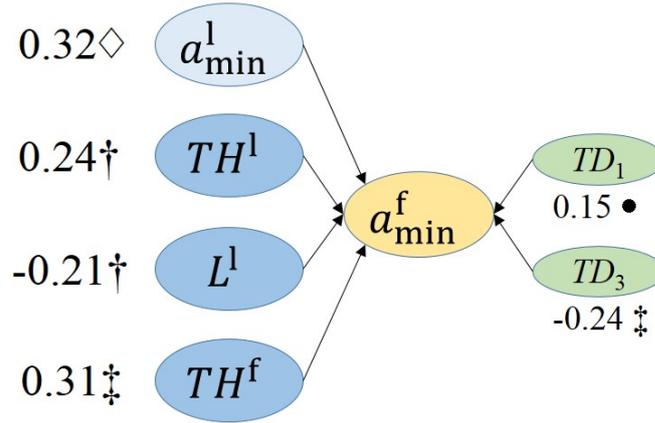
(b) The best multiple regression model for $a_{\min}^{f}$.

Figure 2.16: The best multiple regression model and the coefficients of high-ranking models for $a_{\min}^{f}$ from the AIC perspective.

The non-standardized model of $a_{\min}^{f}$ is as follows.

$$a_{\min}^{f} = 0.37a_{\min}^{l} + 0.14TH^{l} - 0.057L^{l} + 0.19TH^{f} + 0.25TD_{1} - 0.30TD_{3} - 1.1 \quad (2.4)$$

The degree of effect from the drivers became bigger than the cases of the velocity and acceleration. Driver TD1 seemed to have milder deceleration than those of TD6 and TD7 when he drove the motorcycle. Driver TD3 tended to have larger deceleration than TD4 when he drove the truck.

Regarding the distance gap at the starting point, i.e., S_{start}, most of the high-ranking models comprise L^{f} and L^{l} in Figure 2.17a. Although there are some variations of models in same ranking, the alternating factors are "TD_{n}" and "experiment day (*Exp. day*)" which we are not discussing on. As an example of
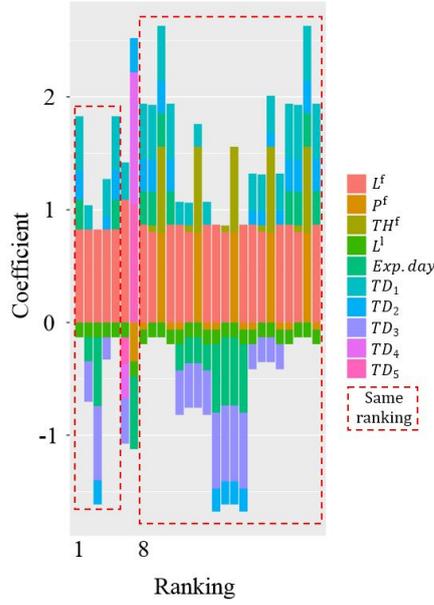
(a) The models for S_{start} within $\Delta AIC < 1.5$ from the best model.
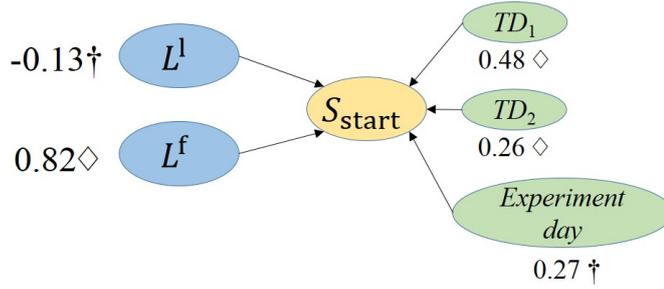
(b) The best multiple regression model for S_{start}.

Figure 2.17: The best multiple regression model and the coefficients of high-ranking models for S_{start} from the AIC perspective.

the best models, we show the leftmost model in Figure 2.17b ($R^2 = 0.57$). We can expect large distance gap with large following vehicles. This would be natural from the safety perspective. However, we will also observe small distance gap with large leading vehicles during stop. This might not be natural but sometimes observed in real traffic. Regarding effects of drivers and experiment environment, the non-standardized model is written as

$$S_{\text{start}} = -0.18L^{\text{l}} + 1.2L^{\text{f}} + 3.9TD_1 + 2.2TD_2 + 1.2Exp.\,day - 0.63. \quad (2.5)$$

The dummy variable "*Exp. day*" takes 0 or 1 to indicate when the experiments were held. We can confirm that the effects of drivers and the environment exceed the order of 1 m.

Regarding the maximum distance gap during a trial, i.e., S_{max}, many high-ranking models comprise positive coefficients for S_{start}, ϵ_{a}, $V_{\text{max}}^{\text{l}}$, and P^{f}. Negative coefficients for TH^{l} and $a_{\text{max}}^{\text{l}}$ are often comprised to these models too in Figure

(a) The models for $S_{\max}$ within $\Delta AIC < 1.5$ from the best model.

(b) The best multiple regression model for $S_{\max}$.
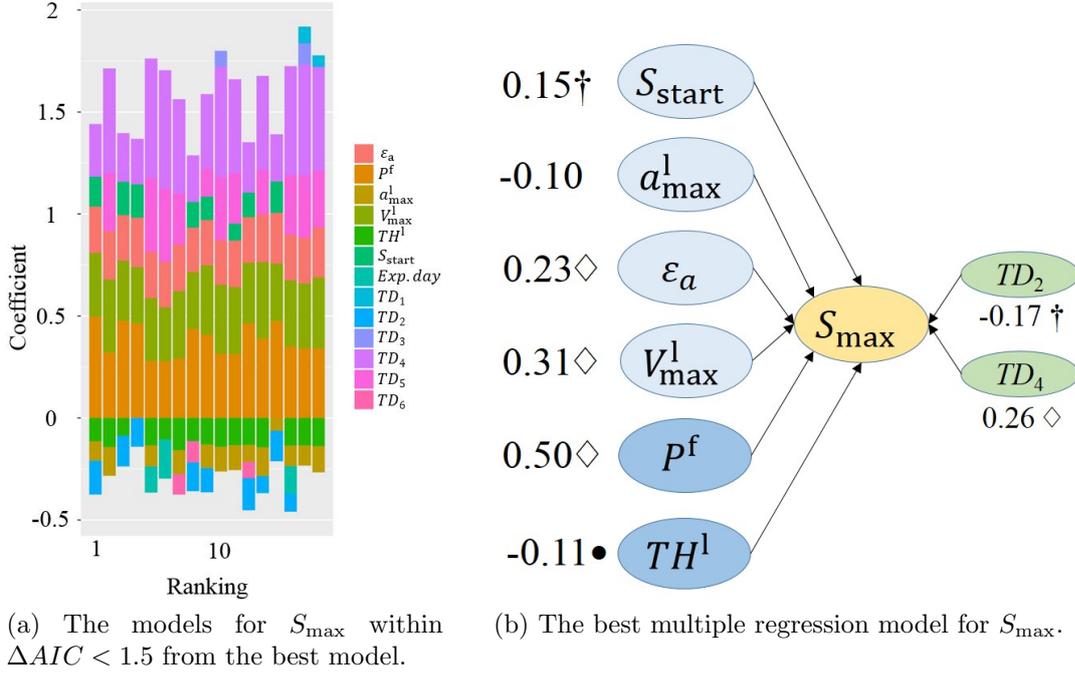
Figure 2.18: The best multiple regression model and the coefficients of high-ranking models for $S_{\max}$ from the AIC perspective.

2.18a. From the best model shown in Figure 2.18b ($R^2 = 0.61$), we can observe the situation that the drivers apt to have short distance gap if the height of leading vehicles becomes high although they tend to keep large distance gap because their vehicle has good engine performance. These behaviors are conflicting from the safety perspective but some drivers would come closer when the leading vehicle is the motorcycle or the truck in the real traffic. The non-standardized model is written as

$$S_{\max} = 1.0S_{\text{start}} - 2.0a_{\max}^{\text{l}} + 1.9\epsilon_a + 3.4V_{\max}^{\text{l}} + 350P^{\text{f}} - 2.3TH^{\text{l}} - 9.7TD_2 + 6.7TD_4 - 33 \tag{2.6}$$

Similar to S_{start}, we confirmed the effects of order of 1 m from drivers. Drivers on the passenger car tended to have variety distance gaps during trials.

Regarding the delay of the maximum acceleration timing (ϵ_{a}), high-ranking models are shown in Figure 2.19a. Three of four best models have the positive coefficient for L^{f}. All of them have the negative coefficient for TH^{l}. The coefficient for P^{f} is ambiguous. The model shown in Figure 2.19b is the leftmost model in Figure 2.19a. The value of $R^2 = 0.21$. We cannot say that the model has sufficient performance, and it seemed to be caused by the randomness of the operation, which is the same reason of the randomness observed in $a_{\min}^{\text{f}}$. From the trends of selected parameters in Figure 2.19a, regarding effects of the vehicle and driving characteristics, it is implied that the acceleration delay depends not on the distance gap at the start point but only vehicle characteristics. Regarding effects of drivers, there were no effects and the non-standardized model is written as

$$\epsilon_a = -0.45TH^{\text{l}} + 0.66L^{\text{f}} - 0.19P^{\text{f}} + 0.96. \tag{2.7}$$
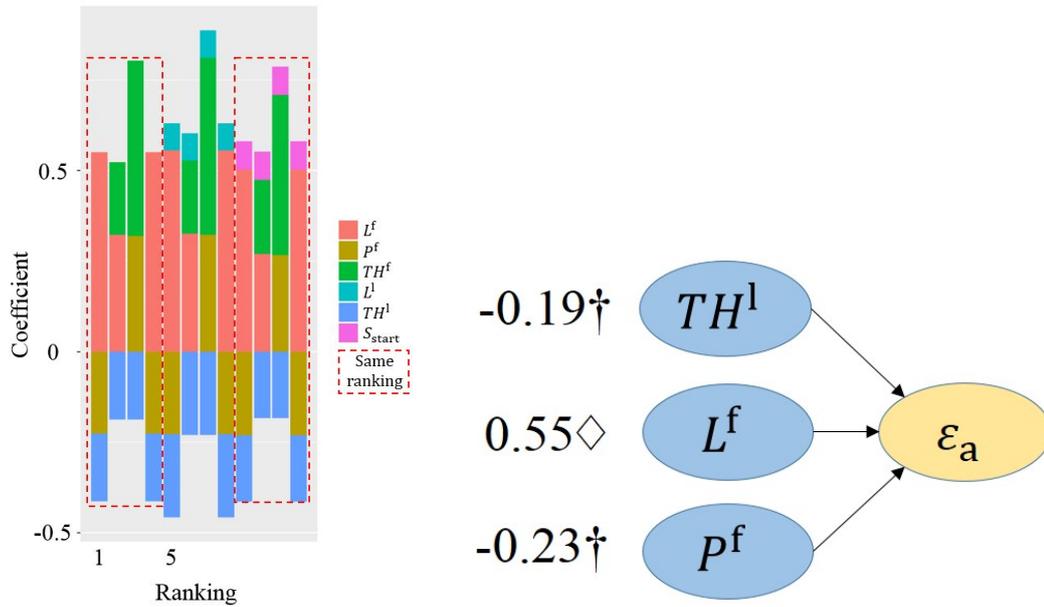
(a) The models for ϵ_a within $\Delta AIC < 1.5$ from the best model.

(b) The best multiple regression model for ϵ_a from the AIC perspective.

Figure 2.19: The best multiple regression model and the coefficients of high-ranking models for ϵ_a from the AIC perspective.

From the discussion above, we found that driving characteristics of leaders were the dominant factor affecting followers' velocity and acceleration. On the other hand, the followers' deceleration, delay of acceleration and distance gaps were also affected by vehicle characteristics of leaders and followers. In particular, although the followers' acceleration is affected mainly by those of leaders, one third of effect comes from vehicle characteristics related vehicle heights. In addition, we could also find that conflicting behaviors from the safety perspective in distance gaps. The drivers apt to have short distance gap when the leading vehicle is large or high. Investigation of this conflicting behavior would belong to the category of psychology of drivers.

Chapter 3

Replicability of Car-Following Models toward Driving Trajectories of Different Following Vehicles

Abstract In this chapter, we investigate the parameters of selected car-following models that can replicate the characteristic trajectories of each following vehicle. We call trajectories of each following vehicle characteristic leaves. Regarding features of characteristic leaves observed in the experiment, we discovered the following:

(A) The distance gap in front of the motorcycles is smaller than that of the cars and trucks when the vehicles travel in the steady velocity, i.e., in a steady phase.

(B) The distance gap in front of the trucks is smaller than that of the cars in the early deceleration phase.

(C) The velocity difference of the motorcycles is more stable than that of the cars and trucks in the steady phase and the deceleration phase.

(D) The magnitude of the velocity difference of the trucks is smaller than that of the cars at the beginning of the acceleration phase and in the latter deceleration phase.

(E) The magnitude of followers' acceleration in the early deceleration phase increases in cases of trucks than in cases of the cars.

(F) The trucks have their acceleration peaks in the latter acceleration phase.

(G) The cars have larger acceleration change in the acceleration phase than the trucks do.

(H) Variation of the velocity difference and followers' acceleration caused by vehicle types were not observed in the initial waiting, steady, and stopping phases.

These features should be replicated by the car-following models.

Regarding the replicability of the models, as long as we investigated, the current car-following models do not have sufficient capability for the replication of the behavioral difference of various types of vehicles. We discovered the following:

(a) In the case of the optimal velocity (OV) model, the vehicle length L^{f} and L^{l} can be used for only the adjustment for the distance gap while we can choose arbitrarily from V_{m}, m, and Δ for the replication of the distance gap and the acceleration. Features of followers' acceleration cannot be replicated perfectly by the OV model without limitation of the parameter range.

(b) In the case of the full velocity difference (FVD) model, it would be preferable to utilize Δs for the replication of the distance gap because it perfectly replicates the features of the distance gap. In the case of followers' acceleration, we have no choice other than β. However, we need to be aware that β will affect the velocity difference simultaneously. For the velocity difference, we need to use V_{m} to keep the fitting independence of each physical value.

(c) In the case of the intelligent driver (ID) model, all the features of the distance gap can be replicated by all the parameters except for L^{f} and L^{l}. In order to assign each parameter to different physical values, it is preferable to assign V_{m} to the acceleration. Assigning s_{min} to the velocity difference is preferable from the perspective of the independence of physical values. Features of the velocity difference cannot be replicated perfectly by the ID model.

(d) In the case of the Gazis-Herman-Rothery (GHR) model, every parameters except for T have the ability to replicate the features of the distance gap without any shift and the unnecessary features. These parameters can also replicate some of the features in the velocity difference. However none of the parameters in the GHR model has the ability to replicate the features in followers' acceleration.

Through the model investigation, we did not find any parameters replicating the differences in all of the distance gap, velocity difference, and followers' acceleration caused by vehicle changes. Furthermore, each parameter caused unnecessary features in some physical values. Although features in the distance gaps can be replicated by some parameters in the models, respective models are weak in different physical values regarding the feature replication.

We first provide an overview of traffic models and select car-following models focused on in Section 3.1. Then, we extract characteristic trajectories from observed time series using decision tree analysis based on the "shapelet" in Section 3.2. The characteristic trajectories in Item (A) to (H) were obtained by this decision tree analysis. In Section 3.3, we simulate follower behaviors with changing parameters of selected models. Simulation setting and rough features in simulated trajectories are introduced. Finally, in Section 3.4, we compare the observed and simulated trajectories if the simulated trajectories can cause variations in the series at where the characteristic trajectories, i.e., shapelets, are observed. The findings mentioned at Item (a) to (d) are explained in Section 3.4.

3.1 Introduction of Traffic Models

3.1.1 Classification of Traffic Models

As we mentioned in Section 1.1, many traffic models have been proposed by different researchers [3, 4]. They had various kinds of motivations for developing the models, including estimating of traffic capacity in a certain road network to contribute to urban design, optimal lane numbers and junction design of free ways, optimization of traffic signal control, development of autonomous driving, detailed replication of traffic phenomena (e.g., traffic jams, etc.), physical investigation of traffic phenomena, extension to mathematical problems, replication of vehicle movement, and psychological analysis of drivers. The examples mentioned above are arranged from macroscopic to microscopic. That is, there are models replicating phenomena observed in road network layers, traffic flow in one road section, inter-vehicle influence, kinetic movement of one vehicle, and even the psychology of a single driver.

As our objective is to properly replicate the phenomena in mixed-traffic flow by models that reflect driving differences between vehicle types, our main focus must not be on models of road network layers, vehicle kinetic movement, or a single driver. There are three types of models that represent traffic flow and inter-vehicle influence: hydrodynamic models, car-following models, and cellular automaton models.

3.1.1.1 Outlines of hydrodynamic models

Figure 3.1a is the schematic representation of hydrodynamic models. Imagine that there are two connected road sections filled with vehicles. For some reasons (e.g., road works), the vehicles cannot proceed with sufficient velocity owing to the high density in the downstream section but they can proceed smoothly until they reach a certain point; hence, there is a boundary of the density and the velocity. As the vehicles cannot go out of the sections, Equation 3.1, i.e., the continuity equation, is established.

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial x} = \frac{\partial \rho}{\partial t} + \frac{\partial Q}{\partial x} = 0 \tag{3.1}$$

where the traffic flow $Q = \rho v$ and x is the position in the road sections.

In particular, Lighthill-Whitham-Richards (LWR) model [47, 48] assumes that the velocity statically corresponds to the density, as shown in Figure 3.1b. This assumption means that Q and ρ are locally equilibrium and models applying this assumption are called models "without dynamic velocity." Using

$$\frac{\partial Q}{\partial x} = \frac{\mathrm{d}Q(\rho)}{\mathrm{d}\rho} \cdot \frac{\partial \rho}{\partial x}, \tag{3.2}$$

we can transform Equation 3.1 as

$$\frac{\partial \rho}{\partial t} + \frac{\mathrm{d}Q(\rho)}{\mathrm{d}\rho} \cdot \frac{\partial \rho}{\partial x} = 0 \tag{3.3}$$

The coefficient for $\partial\rho/\partial x$, i.e., $\mathrm{d}Q/\mathrm{d}\rho$, is the slope in the figure of ρ vs. Q and indicates the velocity of the boundary where the density and velocity are changed. Therefore, as shown in Figure 3.1c, we can calculate the boundary velocity with the flows in different density sections. In the case of Figure 3.1c, the boundary moves from downstream to upstream of the traffic because the slop is negative.
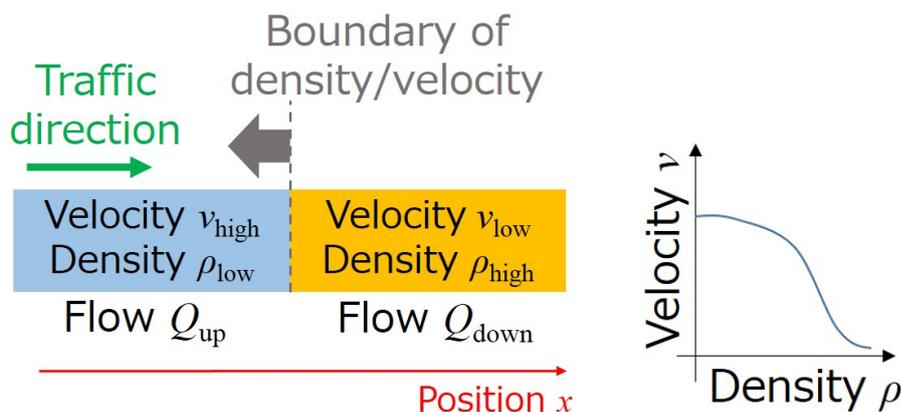
(a) Each section represents road section filled by vehicles with the velocity of v at the density of ρ.

(b) An example of relation between the density and the velocity.
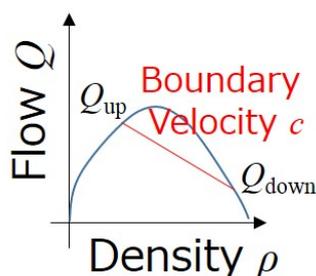
(c) An example of relation between the density and the velocity.

Figure 3.1: Schematics of the hydrodynamic models.

Although the LWR model succeeds in replicating traffic break-down caused by lack of traffic capacity, each vehicle in the traffic flow experiences unrealistic or discontinuous acceleration. In order to replicate traffic wave growth and traffic instability, local acceleration, i.e., the acceleration that each vehicle experiences at a certain time and location should be defined by the local density, velocity, and their gradients. That is

$$\frac{\mathrm{d}V(x,t)}{\mathrm{d}t} = A\left[V(x,t), \rho(x,t)\right] \tag{3.4}$$

where V is the local velocity. Payne and Whitham introduced the relationship

$$\frac{\mathrm{d}V(x,t)}{\mathrm{d}t} = \frac{V_e(\rho) - V}{\tau} + \frac{V_e'(\rho)}{2\rho\tau} \cdot \frac{\partial\rho}{\partial x} \tag{3.5}$$

where V_e is local equilibrium velocity, $V'_e = \partial V_e / \partial x$. By substituting V to v in Equation 3.1, they obtained a continuity equation that considers the speed adaptation by a driver, traffic pressure related to drivers' anticipation, and velocity variance of vehicles[49, 50]. In addition, other models, including diffusion term, etc., have been proposed by other researchers: Kerner-Konhäuser (KK) model [51] and gas-kinetic-based traffic (GKT) model [52].

3.1.1.2 Outlines of car-following models

Target models in the second category consider inter-vehicle influence and are called car-following models. The car-following models define the velocity or acceleration of the vehicle following the leading vehicles using the velocity, velocity difference, distance between vehicles, etc. In Figure 3.2, the accelerations of Vehicles #2 and #3, i.e., a_2 and a_3, respectively, are defined by their velocities v, distances from the leading vehicle s, velocity differences, etc. The acceleration of Vehicle #3 is greater than that of Vehicle #2 as $s_3 > s_2$. Because the car-following models depict movements of respective vehicles, the parameters for each vehicle can be defined separately. If the formulation includes the physical values of multiple leaders, we can even reflect the effects of multiple leaders [53, 54]. The car-following models have the potential to replicate vehicle movements accurately because time and space in the models are not discrete but continuous. Detailed categorization is described in Section 3.1.2.

$$a_2 = f(v_2, s_2, v_2 - v_1, \text{etc.})$$
$$a_3 = f(v_3, s_3, v_3 - v_2, \text{etc.})$$
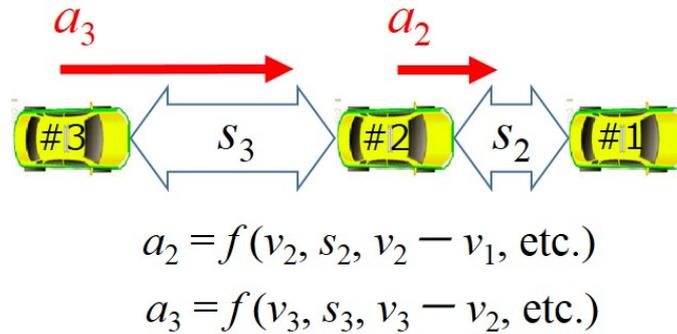
Figure 3.2: Schematics of the car-following models.

3.1.1.3 Outlines of cellular automaton models

The third category consists of cellular automaton (CA) models. As shown in Figure 3.3, time and space are discretized in the CA models. Each vehicle decides whether to go forward or not based on a "rules." For example, "if the leading vehicle occupies the cell right in front of where the vehicle is, the vehicle keeps its current position; otherwise, it moves on to the cell right in front of it." This

rule is called Rule 184 and is formulated as

$$x_\alpha(t+1) = x_\alpha(t) + v_\alpha(t) \tag{3.6}$$

$$v_\alpha(t) = \begin{cases} 1 & (x_{\alpha-1}(t) - x_\alpha(t) > 1) \\ 0 & (otherwise) \end{cases} \tag{3.7}$$

where α is vehicle number. Indeed, detailed replication of traffic phenomena is difficult for CA models. We can observe various traffic phenomena (e.g., occurrence and growth of the traffic jam, propagation of the traffic wave, etc.) based on introduced rules which are easily extended [55]. In addition, owing to simple modeling of particles, CA models can be applied to various kinds of systems, including transfer of objects, information, and states (e.g., a swarm of ants [56], molecular motor proteins [57], stock trading system [58], etc).
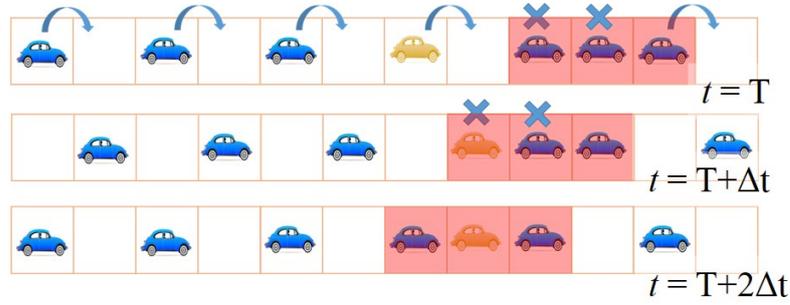
Figure 3.3: Schematics of the cellular automaton models.

In following analysis, we focus on the car-following models. Although the hydrodynamic models and the CA models are representations of traffic flow and inter-vehicle influence, these models have disadvantages when replicating mixed traffic. With the hydrodynamic models, we cannot investigate the internal structure of traffic. With CA models, we cannot replicate detailed behaviors of each vehicle. On the other hand, the car-following models can calculate the behaviors of vehicles in continuous time and space. Vehicle characteristics can be reflected in the parameters of each vehicle. Although the calculation costs of the car-following models are higher compared to other models, the car-following model is the best method to replicate mixed traffic based on behavioral differences between types of vehicles.

3.1.2 Introduction of Car-Following Models

In this section, we introduce some car-following models and choose models to investigate in the following analysis. The car-following models are roughly categorized as follow:

1. $a(t) = f(t)$: the acceleration of a vehicle at a certain time is determined by physical values at that time.

2. $a(t+T) = f(t)$: the acceleration of a vehicle at a certain time is determined by the physical values of time delay T.

3. $v(t+T) = f(t)$: the velocity of a vehicle at a certain time is determined by the physical values of time delay T.

4. Action point model

5. Fuzzy logic-based model

Pipes introduced the first car following model [59] formulated as

$$a(t) = A \cdot \Delta v(t) \tag{3.8}$$

where A is a constant and Δv is the velocity distance, i.e., $v^{\mathrm{l}} - v^{\mathrm{f}}$; the superscripts indicate the leader and follower, respectively. This model is categorized as Item 1. In order to replicate the traffic instability, Chandler *et al.* introduced the time delay T and formulated the acceleration as [60]

$$a(t + T) = A \cdot \Delta v(t). \tag{3.9}$$

This model is categorized as Item 2. Newell *et al.* introduced optimal velocity $V_{\mathrm{opt}}(h)$ and formulated the velocity as [61]

$$v(t + T) = V_{\mathrm{opt}}(h(t)). \tag{3.10}$$

The optimal velocity is a monotonically increasing function and the limit of $V_{\mathrm{opt}}(h)$ as the headway distance h approaches infinity is equal to the maximum velocity. Figure 3.4 shows an example of the optimal velocity function. This model is categorized as Item 3.
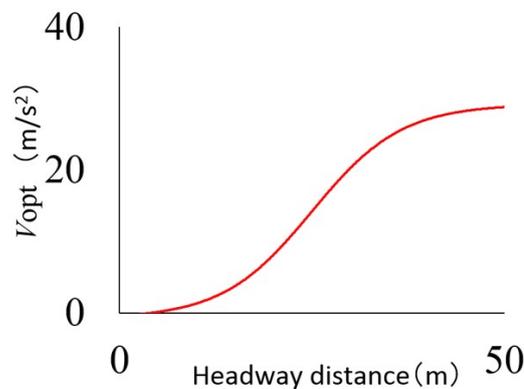
Figure 3.4: An example of the optimal velocity function.

Based on these elementary car-following models, researchers developed various kinds of models. For Item 1, Bando *et al.* proposed an optimal velocity (OV) model for the acceleration formulated as [62]

$$a(t) = A \left\{ V_{\text{opt}}(h(t)) - v(t) \right\}. \tag{3.11}$$

The OV model was derived from Taylor expansion of Newell's model, which is often utilized for mathematical investigations of the traffic models because it does not include time delay T. Note that if we apply Taylor expansion again, we can obtain the OV model formulating not the acceleration but the jerk [63]. Bando *et al.* also presented the OV model by considering the length of a vehicle [64].

Jiang *et al.* extended the OV model as [65]

$$a(t) = \frac{V_{\text{opt}}(h(t)) - v(t)}{\tau} + \gamma \Delta v(t) \tag{3.12}$$

where τ is constant and speed adaptation time of a driver, and γ is a constant. This model is called the full velocity difference (FVD) model. Equation 3.12 takes the velocity difference into account. When the value of Δv is large, the acceleration also becomes large. This model seems more suitable than the OV model to replicate more realistic propagation of traffic waves.

The intelligent driver (ID) model proposed by Treiber *et al.* is another variation of Item 1 [33]. The formulation of the ID model is

$$a(t) = A \left[1 - \left(\frac{v(t)}{V_{\text{m}}} \right)^{\delta} - \left(\frac{s^*(v(t), \Delta v(t))}{s(t)} \right)^{2} \right] \tag{3.13}$$

where V_{m} is the maximum velocity and A is the maximum acceleration. s is the distance gap, i.e., the space distance between the leading vehicle and the following vehicle, which is the headway distance minus the vehicle length. s^* is the desired distance gap determined by the velocity on its own and the velocity difference between the leading vehicle. The desired distance gap s^* is determined as

$$s^*(v(t), \Delta v(t)) = s_{\text{min}} + \max \left(0, v(t)T + \frac{v(t)\Delta v(t)}{2\sqrt{AB}} \right) \tag{3.14}$$

where s_{min} is the minimum distance gap when the vehicle stops, and T is the preferable time gap when the vehicle travels at the velocity of v. B is the comfortable deceleration. When the vehicle is rapidly approaching an obstacle (e.g., a traffic signal), the model produces sufficient deceleration for it to stop. The deceleration approaches B after covering enough distance to stop in front of the obstacle. The ID model can produces more realistic acceleration than the OV model, and its parameters have physical meanings. Therefore, the ID model is often utilized for realistic traffic analysis even for mixed traffic [34], as we mentioned in Section 1.2.

Another example of Item 3 is the Gipps' model [66]. The simplified formulation introduced in [67] is as follows.

$$v(t + T) = \min \left[v(t) + AT, V_{\text{m}}, v_{\text{safe}}(s(t), v^{\text{l}}(t)) \right] \tag{3.15}$$

The first and second terms in the minimization regimes indicate that the vehicle accelerates at the acceleration of A if the vehicle's velocity does not reach the maximum velocity of V_{m}. When the safe speed v_{safe} determined by the distance gap and leader's velocity v is less than the first and second terms, the velocity is adjusted to the value of v_{safe}. The safe speed v_{safe} is the speed to avoid collision with a fixed deceleration parameter B. Because of its simplicity, the Gipps' model is utilized in some commercial simulators. Although both the Gipps' model and the ID model consider the driving strategies of drivers and produce realistic acceleration and deceleration ranges, the transition of acceleration, stable velocity, and deceleration phases are toggled in the Gipps' model. Therefore, the Gipps' model seems to be unsuitable to replicate the detailed following behaviors of various types of vehicles.

For Item 2, there is a large group of models called "GHR" models named after Gazis, Herman and Rothery. The review article presented by Brackstone introduces the history of GHR models [4]. GHR models originated from Chandler's model [60] and [68], in which the sensitivity parameter is a just constant A, as shown in Equation 3.9. The improved models mentioned below have the sensitivity parameter of

$$A(t) = \frac{Cv^m(t)}{h^k(t-T)} \tag{3.16}$$

and the model equation is written as

$$a(t+T) = A(t+T) \cdot \Delta v = \frac{Cv^m(t+T)}{h^k(t)}\Delta v(t) \tag{3.17}$$

where C is a constant, and k and m are exponents for the headway distance and velocity, respectively. First, Herman *et al.* proposed that the sensitivity constant decreased when the headway distance increased: they introduced the model with $k=1$ and $m=0$ [69]. Edie proposed that the necessity of introducing velocity dependence to the sensitivity constant, i.e., $k=1$ and $m=1$ [70]. Gazis *et al.* applied Edie's approach and proposed the general formulation of GHR model of Equation 3.17. They investigated macroscopic relationship between the flow and density in combinations of k and m, and found favorable combinations of k and m from 18 experiment samples [71]. Based on their study, many researchers tried to determine the best combinations of k and m [72, 73, 74, 75, 76, 77]. From the research by Ceder *et al.*, dividing parameters in congested and non-congested phases, or acceleration and deceleration phases, became main stream of parameter fitting. Through various methods of observations (e.g., aerial observation, stationary video recording, etc.), they increased sample numbers to obtain better parameter combinations. Although the parameters in GHR models are difficult to interpret physically, currently GHR models are frequently used in the field of traffic engineering owing to their performance of traffic replication.

The other variation of Item 2 is the Helly model [78], which is generally formulated as

$$a(t+T) = C_1\Delta v(t) + C_2 \left\{ h(t) - D(t+T) \right\} \tag{3.18}$$

where D is desired following distance [4] formulated as

$$D(t + T) = \alpha + \beta v(t) + \gamma a(t). \tag{3.19}$$

Models categorized as this item are also called linear models because the formulation of Helly model is linear. As the formulation is linear, the concept of control engineering can be easily applied to the Helly model.

Although simulated drivers in the models mentioned above react to given physical environment continuously, real drivers cannot perceive small variations of physical values or variations occurring far away from their present location. After recognizing variations of physical values, i.e., after the variation exceeds some threshold, the drivers finally take action. With the introduction of the concept of "action point," many types of the action point models categorized as Item 4 have been developed, and they are even utilized for commercial traffic simulators. Michaels first introduced the concept of the action point [79]. His idea is as follows. The driver can perceive the approach of the leading vehicle according to the change in the visual angle of the leading vehicle. Once the rate of change exceeds a threshold, the driver decelerates until the relative velocity is no longer recognized. Other examples of action points are, the threshold for perception of closing or opening [4], and acceptable maximum and minimum gaps in the car-following regime [67]. It is noted in [4] that the validity of action point models has not been evaluated completely. The reason is that it is difficult to prove the validity of respective thresholds and action points, even when the whole simulation works properly.

The models categorized as Item 5 utilize fuzzy logic. That is, the drivers simulated by these models calculate the degree of "terms." For example, drivers will recognize "slow" when the velocity is 10 km/h. The "membership function" of "slow" becomes 1 and that of "fast" becomes 0. The membership function indicates the degree of truth of the "term" by its value from 0 to 1. When the velocity is 100 km/h, the drivers will recognize "fast" and the membership functions of slow and fast become 0 and 1, respectively. Once the drivers judge which "terms" are most likely to be true, the drivers perform their action based on "fuzzy-logic," e.g., "IF slow AND headway is far, THEN accelerate." The first research that applied this approach was conducted by Kikuchi *et al.* [80]. Regarding this category of models, it was also noted in [4] that the calibration of membership sets has not been sufficient.

From the models mentioned above, we chose Items 1 and 2, and investigated the ability of their parameters to replicate vehicle difference. The important reasons why we chose these models are because of the number of parameters to consider and the clear dependence of physical values on these parameters. Furthermore, in these categories, there are major models for current physical and engineering traffic analysis. We chose the OV model, FVD model, ID model, Helly model, and GHR model. These models are roughly arranged from models frequently used in physics field to ones frequently used in engineering field. By choosing these models, we aim to contribute to both the physical and engineering fields.

3.2 Features in Observed Trajectories of Different Following Vehicles

3.2.1 Idea of Analysis

One of the main objectives of this thesis is finding suitable models to replicate driving differences of each vehicle and finding parameters to replicate these differences in respective models, as described in Section 1.3. We need to compare the relationships between physical values of the models and the observation data with changing model parameters or vehicle characteristics. As the car-following models chosen in Section 3.1.2 define the dependence of the acceleration on the velocity, velocity difference, and distance gap or headway distance, the relationships to be investigated are those between the accelerations and other physical values, i.e., the velocity, velocity difference, etc. Figure 3.5 shows the idea of the analysis conducted in Section 3.2 to 3.4. Imagine that we have one green trajectory describing the relationship between the velocity and acceleration. By changing vehicle type, i.e., vehicle characteristics, we obtain other green trajectories. On the other hand, we obtained red-dotted trajectories from simulations that change one parameter of some models. If the trends of green trajectories obtained from observations and those of red-dotted trajectories obtained from the simulations are consistent, as shown in Figure 3.5, we can say that the parameter can replicate the trajectory difference caused by vehicle types. Furthermore, we can also conclude that the model can perform the replication of vehicle difference. Therefore, we need to extract the characteristic trajectories of each vehicle type and confirm whether the dependence of parameter adjustments on trajectory variations is consistent with that of observations.
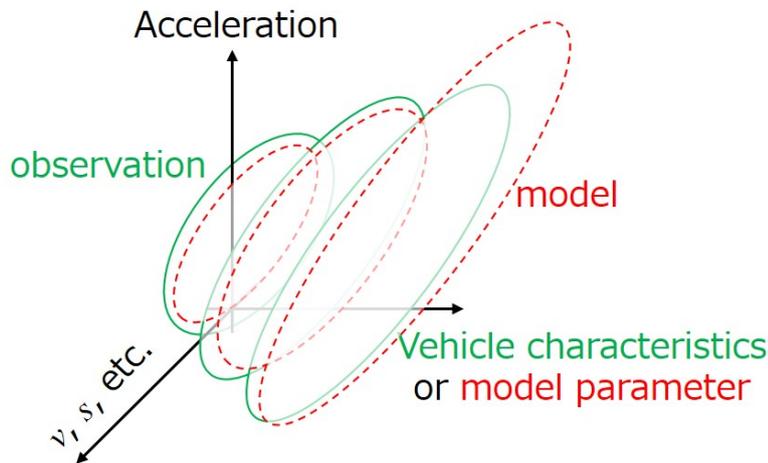
Figure 3.5: A schematic of the analysis. Trends of trajectories when a parameter in a model is changed should be consistent with those of trajectories when a vehicle characteristic is varied.

3.2.2 Data Processing Method

We apply a method of machine learning method to extract the characteristics of the observed trajectories as we can obtain quantitative standards to check if the trajectories are similar or not. If we obtain some standards to distinguish which trajectories were observed from which vehicles, we can utilize the standards as features in the trajectories of each vehicle. As the trajectories are labeled by vehicle type, the problem that we tackle here is the classification with supervised learning. Methods and examples to solve this problem are as follows.

1. Clustering:

   k-nearest neighbor (kNN) [81, 82]

2. Regression:

   linear regression, logistic regression [83], support vector machine [84, 85]

3. Tree:

   decision tree [86, 87, 88], random forest [89]

4. Artificial neural network:

   perceptron [90], convolutional neural network (CNN) [91]

5. Bayes:

   naive Bayes [92, 93, 94]

Clustering by k-nearest neighbor (kNN) first calculates the centroids of each physical value in each category. The classifier of the kNN classifies a data to the category in which the centroid is the nearest to the data.

The classifier of regression methods divides data into two regions with a given regression function. The support vector machine (SVM) also divides data into regions using the maximum-margin hyperplane. The SVM can classify complex data on multidimensional space with kernel functions.

A decision tree categorizes data into each labeled group using multi-layered standards for classification or decision rules. A random forest is a kind of ensemble methods that uses small trees generated from selected data groups. It is widely held that the classification performance of random forests is high but understanding the learning results, i.e., how the random forest classifies the data, is more difficult compared to kNN, regression methods, and decision trees.

An artificial neural network is a classifier inspired by the neural system of the brain. Currently, this method is often utilized for the classification of images or big data, which are not divided linearly [95, 96].

A naive Bayes classifier is based on Bayes' theorem and calculates the posterior probability if certain data belong to a certain category. Assuming the independence of a stochastic variable in the plausibility, the Bayes classifier simplifies plausibility and obtains the posterior probability.

In this thesis, we clarify the features of the trajectories of each vehicle and confirm if some car-following models can replicate these features. The readability

of learning results, i.e., how easily one can interpret the rule based on which the classifier distinguishes data, is an important element when developing a classifier. The decision tree is one of the best methods from the readability perspective. Therefore, we try to develop a decision tree for the observed trajectories in Section 3.2.3.

In Section 3.2.2.1, we introduce the developing method of trees based on characteristic trajectories. In Section 3.2.2.2, we consider some problems of our observed trajectories in order to apply the method introduced in Section 3.2.2.1. To overcome these problems, we apply "dynamic time warping" in Section 3.2.2.3.

3.2.2.1 Decision-tree classification of series data based on distance to shapelets

The basic decision tree learns from data with scalar variable sets and grows itself until the termination condition is satisfied. Ye *et al.* proposed a method to develop a decision tree from series data, e.g., time series or shapes on a two-dimensional plane [97]. This method has been utilized in some researches to extract characteristics from time series [98, 99]. Ye *et al.* used the classification of leaves as an example in their paper [97]. That is, they tried to classify the leaves of two different plants, which included color difference, insect-bite damages, etc., according to their shapes.

Imagine that one classifies leaves shown in Figure 3.6. Humans easily recognize that the shape marked by a red line around a stalk is an essential clue to distinguish leaves. When the leaf shape is deployed on a one-dimensional plane, the red line is plotted as shown in Figure 3.7. The center of gravity of the leaf is indicated as Point G in Figure 3.6. The angle around Point G is θ and the distance from G is r in Figure 3.6 and 3.7.

Leaves of Plant A

Leaves of Plant B
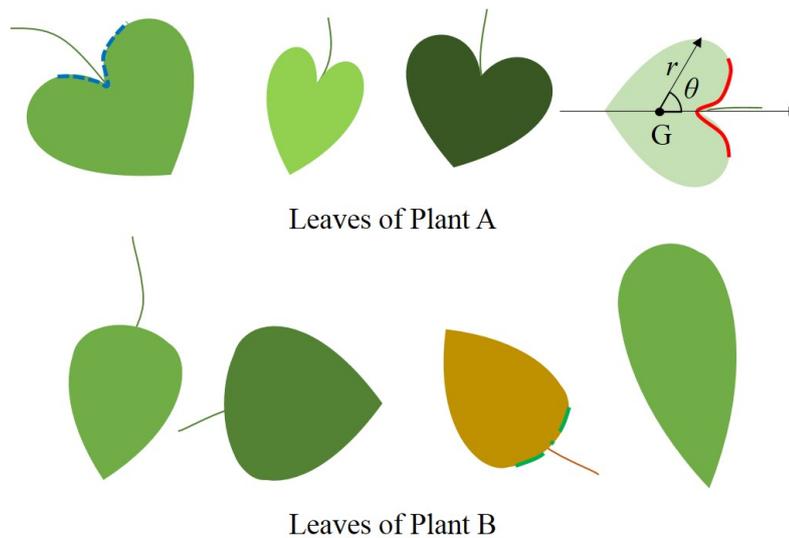
Figure 3.6: Schematics of leaves for classification.
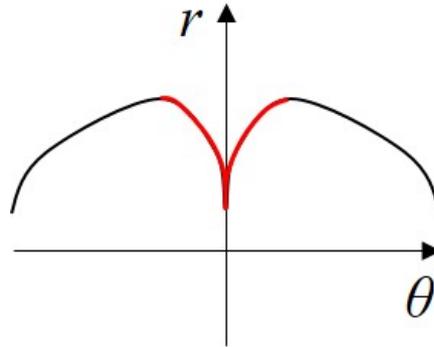
Figure 3.7: Schematics of leaf shape expanded into one-dimensional space.

Then, we choose other leaves from Plant A and Plant B and compare the most similar sub-trajectories. The examples are shown in Figure 3.6 by a blue dashed line and a green chain line. These lines in Figure 3.8 are the deployed shapes on the one-dimensional plane. We can recognize clearly that the error from the red line of the trajectory for Plant A (the blue dashed line) is smaller than that of the trajectory for Plant B (the green chain line). Ye *et al.* named this most significant characteristic red line a "shapelet" and proposed the development of a decision tree based on errors from these shapelets.
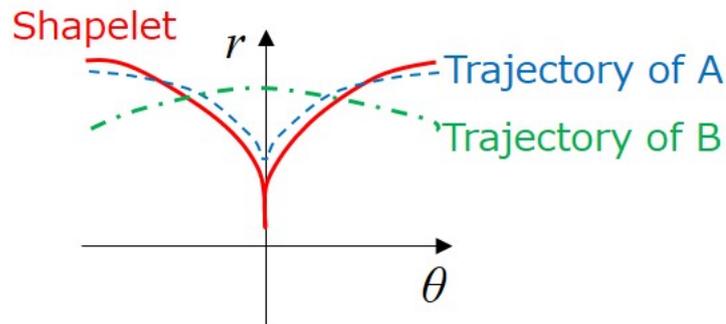
Figure 3.8: Comparisons of the shapelet and trajectories of leaves A and B.

A flowchart of their approach is shown in Figure 3.9. When the algorithm starts, it generates shapelet candidates, i.e., it extracts all the sub-elements from all data series in non-divided original datasets or groups. For these all sub-elements, it calculates the gain, d_{osp}, divided groups, and a margin between the new divided groups. The gain is the information gain calculated from information entropies of the original data group and data groups divided by one node. When the information gain increases, one can say that the node dividing the original group reduces the degree of mixture of types in divided groups. The information gain is calculated as

$$\mathrm{gain} = H(D) - \{f(D, d_1) \cdot H(d_1) + f(D, d_2) \cdot H(d_2)\} \qquad (3.20)$$

where D is the dataset in the original group and, d_1 and d_2 are the datasets in the divided groups. H is the function for the information entropy from datasets and f is the weight function written as

$$H(D) = -\sum_{type} \frac{n_{type}}{N_D} \log \frac{n_{type}}{N_D} \tag{3.21}$$

$$f(D, d) = \frac{N_d}{N_D}, \tag{3.22}$$

where N_D and N_d are the number of data series in D and d, and n_{type} is the number of data series of each labeled *type*. In our case, *type* consists of the motorcycles (MC), cars (C), and trucks (T). The value of d_{osp} is the "distance" threshold where the original dataset is divided into two groups. The concept of "distance" is explained later. The margin of the two divided groups is written as

$$\text{margin} = \delta_2 - \delta_1 \tag{3.23}$$

where δ_1 and δ_2 are the averaged distances of datasets divided into d_1 and d_2, respectively. The trials in Group d_1 are within the threshold of d_{OSP} and those in d_2 are out of it. When the margin becomes large, the average distance of d_2 is far from the shapelet candidate, and the average distance of d_1 is close to the candidate.

The calculation sequence for obtaining gain, etc, consists of the sub-sequence shown on the right side in Figure 3.9. For all data series in the dataset, it calculates the "distance" between the candidate and data series. Although the lengths of the candidate and data series are different, we can find the minimum error value, i.e.,

$$\text{distance} = \min \left\{ \sum_{i=1}^{l_c} |Series_{i+j} - Candidate_i| \right\} \quad (j = 1, 2, \cdots, l_s - l_c) \tag{3.24}$$

where l_s and l_c are the length of the data series and the candidate, respectively. The distance is the minimum error between the data series and candidate, which we compared in Figure 3.8. After sorting the data series based on these distances, the algorithm searches for the optimal split point to divide the original dataset.

In Figure 3.10, an example of a sorted list and division procedure is shown. Imagine that the algorithm chooses a threshold distance dividing the data series as d_{sp}. Then, it can calculate the information gain by Equation 3.20, and simultaneously obtain the margin. By changing the value of d_{sp}, it can find the optimal split point d_{osp} where it obtains the best gain. If there are several split points in which the gains are the same, the algorithm chooses the split point with the largest margin.
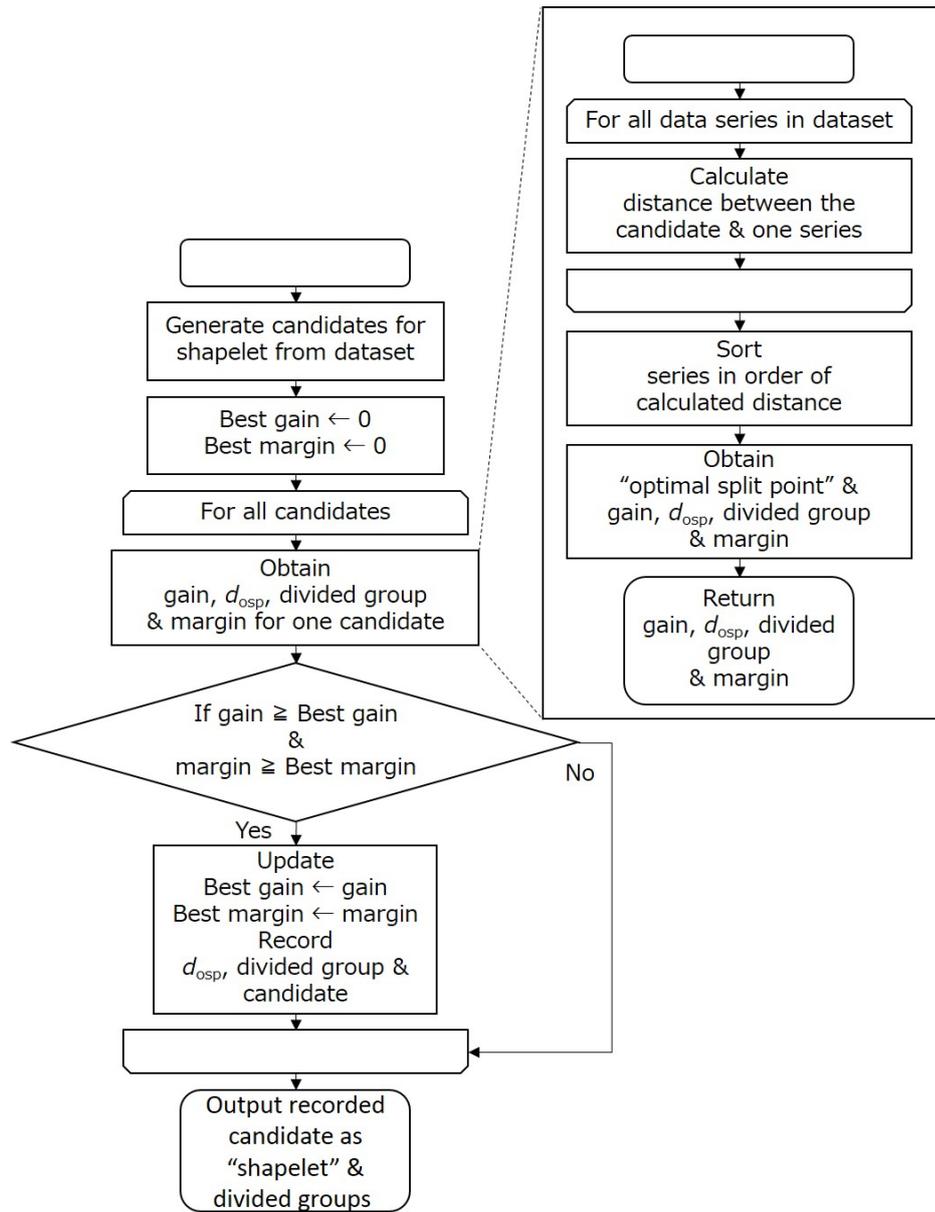
Figure 3.9: The flowchart of decision tree construction with shapelets.
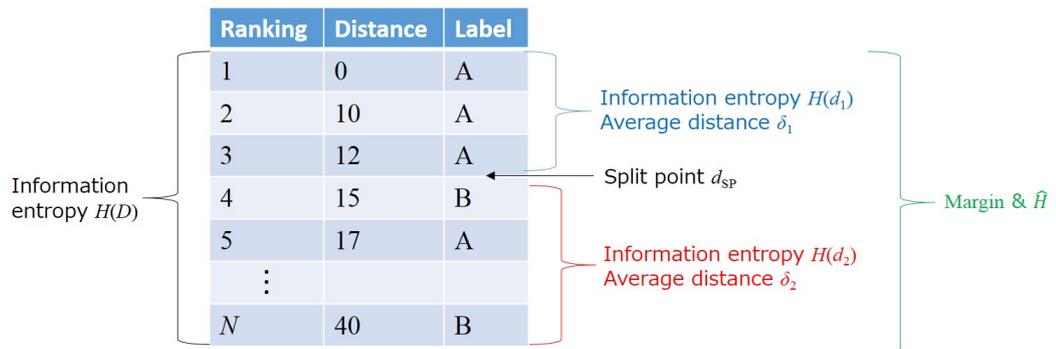
Figure 3.10: Calculated values when the optimal split point is searched.

After recording the best gain, the main sequence checks if the obtained best gain is better than the previous best gain. It also checks whether the obtained margin is greater than the previous largest margin. If both conditions are satisfied, then the main sequence records the shapelet candidate, gain, and margin realized with the candidate, and the optimal split point d_{osp} for division of the group. The algorithm finds the candidate that divides the original group with the best gain and the best margin. This candidate is called the "shapelet" and it provides one node in the decision tree dividing the original data. By applying this sequence continuously, we can develop the decision tree for the data series.

Ye *et al.* [97] considered that the gain was more important compared to the margin in the original algorithm and selected the shapelet with the better gain even when it had the worse margin. However, if we applied this manner of selection, we obtained shapalets with so small margin that we could not distinguish difference between the divided groups. Therefore, we searched for the shapelet with both the best gain and the best margin.

In our analysis, we applied some limitations to the original algorithm. We extracted not all the sub-elements from the dataset but sub-elements with fixed lengths; the index lengths of them were from 22 to 724. The index of the data series will be introduced in following sections. The lengths of 724 and 22 correspond to the half of and 1/64 of the series length, respectively. While doubling the length of the sub-element from 22 to 724, we extracted sub-elements. We also added some conditions to avoid over-fitting. We limited the minimum number of trials included in a node to eight. If the number of trials became less than 16, we did not add deeper decision rules to the node. When one type of vehicle occupies more than 90 % of a node, we also did not add deeper decision rules to the node.

3.2.2.2 Application of the shapelet analysis to the observed trajectories

A decision tree that classifies data based on the series or the two-dimensional shapes can be realized by the method proposed by Ye *et al.* In the case of the two-dimensinal shape, it requires deployment of the shape on one-dimensional space. In the example of the leaves, the angle θ around the gravity point G is utilized for deployment in Section 3.2.2.1. However, it is difficult to apply the angle θ to our observed trajectories. Figure. 3.11 is an example of observed trajectories. If we try to get distances from the center of gravity of each trajectory, the deployed one-dimensional series will have multiple values for the same θ.

One simple approaches to tackle this problem is to deploy trajectories based on observed time. However, we again face a problem: each observed trajectory has different lengths and respective phenomena (i.e., acceleration, deceleration, etc.) occur for different durations in each trajectory. In other words, even when we have the same two trajectories on the distance gap vs. acceleration plane, the series deployed on one-dimensional planes (time vs. distance gap and time vs. acceleration planes) based on time can be different. In order to realize the basic idea introduced in Section 3.2.1, we need a common index for each series. In Section 3.2.2.3, we introduce the method to obtain the common index.
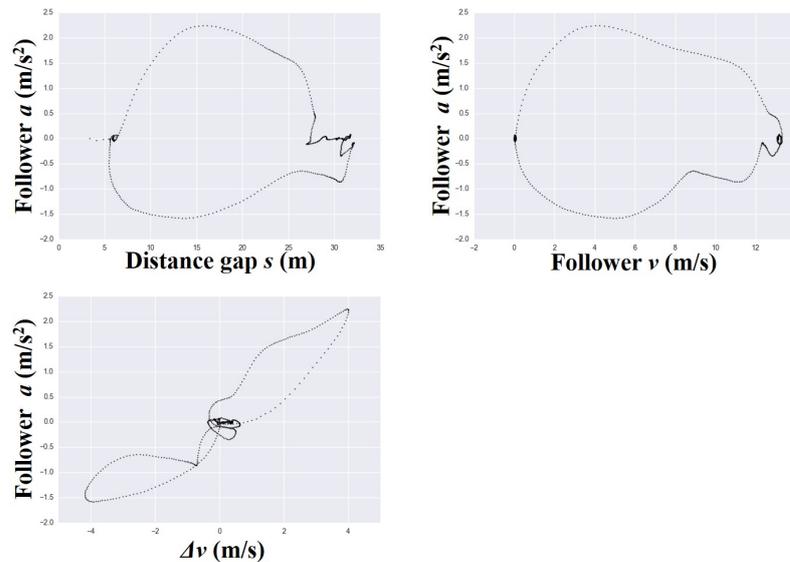
Figure 3.11: An example of observed trajectories. The following vehicle was the car and the leading vehicle was the truck.

3.2.2.3 Dynamic Time Warping

Dynamic time warping (DTW) was first proposed by Bellman *et al.* [100]. Imagine that there are two time series a and b, as shown in Figure 3.12a, and the time index for Series a is counted by i and that of Series b is counted by j. The time intervals for respective indices are the same. Assume that the maximum i and j are 8 and 10, respectively. That is, Series b is longer than Series a. As shown in Figure 3.12a, not only series lengths but also timing of the emergence of similar phenomena and their shapes are somewhat different. We try to assign the common index k by selecting sampling points where similar phenomena are observed in Series a and b. In Figure 3.12b, we locate a purple box $k = 0$ at $(i, j) = (0, 0)$. For the next index, we assign $k = 1$ to $(i, j) = (0, 1)$ because the value has already increased at $i = 1$ in Series a; both series start increasing from $k = 1$. As similar phenomena occur in both series for a while, we can proceed until $(i, j, k) = (4, 5, 5)$. We again have to assign multiple indices for $i = 4$ because the length of the decreasing phase of Series b is longer than that of Series a. Although shapes of the bottom of the concave parts are somewhat different, we assign one index for $(i, j) = (6, 8)$ because we can assume that both parts are roughly in the same phase. Through such manipulation, we can assign the common index to two series and obtain the sets of (i, j, k) as shown in Figure 3.12b, called a "warping path." In the algorithm, the warping path is to be searched greedily to minimize

$$\sum_{k=0}^{K} d_k \quad (d_k = |a_{i_k} - b_{j_k}|) \tag{3.25}$$

where K is the maximum k, and i_κ and j_κ are i and j when $k = \kappa$, respectively.
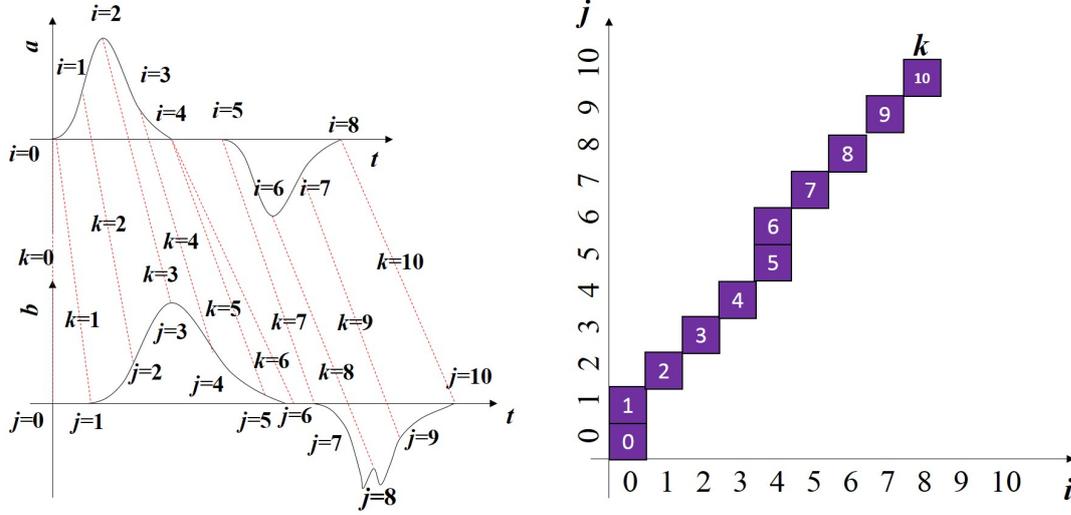
(a) Assigned common index k for each series.

(b) An example of the warping path.

Figure 3.12: Schematics of the DTW.

We can acquire the common index for two series by DTW. However, we need to compare all the observed and simulated trajectories. In other words, the lengths of all the trajectories need to be adjusted to that of a certain reference series, as shown in Figure 3.13. The same length of indices for all series can be obtained with a reference series whose length is sufficiently longer than those of observed and simulated series. As reference signal, we applied a series for the velocity shown in Figure 3.14. It has sufficiently longer length than the observed and simulated series. All the observed and simulated series are to be assigned a new index k based on the comparison between the reference velocity and their velocities. In the range $k = (0, 249)$, the velocity of the reference series V_{ref} is 0. In the range $k = (250, 1349)$, the velocity of the reference series follows

$$V_{\text{ref}} = \frac{V_{\text{max}}}{2} \left[1 - \cos \left\{ 2\pi/T_{\text{ref}}(k - 250) \right\} \right] \tag{3.26}$$

where $T_{\text{ref}} = 1100$ and $V_{\text{max}} = 13.3(\text{m/s})$, which is the averaged maximum velocity of the observed series.

One of the reasons why we applied this series is that when the reference series has a constant area at the top of the velocity, the maximum value affected the assigned index of each series. Another reason is signal smoothness. Of course, smooth velocity series is more similar to the observed series than a series like a trapezoid, which are not differentiable. Furthermore, we can define the early or latter phases of the acceleration and deceleration based on whether the series is convex or concave. Basically, we assume that the concave region in the acceleration phase is the early acceleration phase, the convex region in the acceleration phase is the latter acceleration phase, the convex region in the deceleration phase is the early deceleration phase and the concave region in the deceleration phase is the latter deceleration phase. Because the observed and simulated series have

steady phase while maintaining constant velocity, we assume that the initial range where the velocity is less than 10 % of the maximum value are the waiting or stopping phase, In addition, we also assume that the the range where the velocity is more than 90 % of the maximum value is the steady phase. Specifically, we assume the range of $k = (0, 362)$ as the waiting phase (W), $k = (363, 524)$ as the early acceleration phase (EA), $k = (525, 686)$ as the latter acceleration phase (LA), $k = (687, 912)$ as the steady phase (ST), $k = (913, 1074)$ as the early deceleration phase (ED), $k = (1075, 1237)$ as the latter deceleration phase (LD) and $k = (1238, 1449)$ as the stopping phase (SP).

Series a (i_a)	Value1	Value2	Reference (k)	Series b (i_b)	Value1	Value2
1	X_{a1}	Y_{a1}	1	1	X_{b1}	Y_{b1}
2	X_{a2}	Y_{a2}	2	2	X_{b2}	Y_{b2}
3	X_{a3}	Y_{a3}	3	3	X_{b3}	Y_{b3}
4	X_{a4}	Y_{a4}	4	4	X_{b4}	Y_{b4}
5	X_{a5}	Y_{a5}	5	5	X_{b5}	Y_{b5}
			6	6	X_{b6}	Y_{b6}
			7			
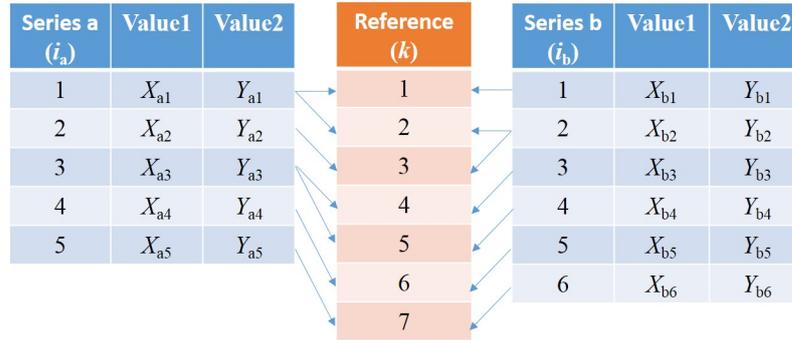
Figure 3.13: Schematics of index assignment to the reference series.
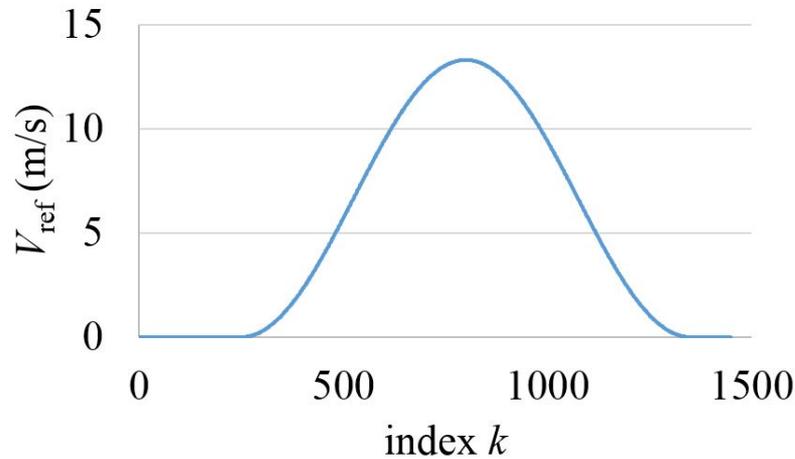
Figure 3.14: The reference series for DTW. The observed and simulated series were assigned new index k based on the velocity comparison with this series.

Even if we prepared sufficient length of the reference series, the DTW algorithm assigned multiple common index k to a few sampling points of the observed and simulated series to minimize Equation 3.25. Therefore, we restricted the algorithm from assigning multiple k to the observed and simulated series; we restricted the horizontal movement on the warping path when assuming j as k in Figure 3.12b. Figure 3.15 is an example of the DTW process toward an observed velocity series. Note that our program was developed based on the program introduced in [101]. We confirmed that all the series were adjusted to the length

of the reference series. Note that, when we develop the decision tree, we did not calculate the distance between the shapelet candidate and the sub-series all over the index k as written in Equation 3.24 but only between the candidate and the sub-series on the same index range. It is because DTW algorithm adjusts the positions of the similar phenomena. We need not to shift the range for the distance calculation to search the similar sub-series.
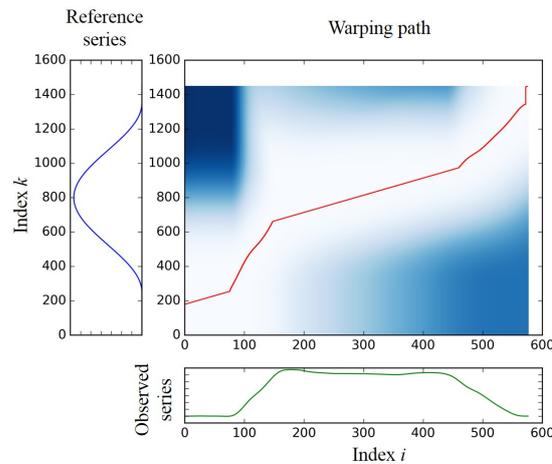
(a) An example of the warping path. The path is depicted by the red line. The depth of blue indicates d_k in Equation 3.25.
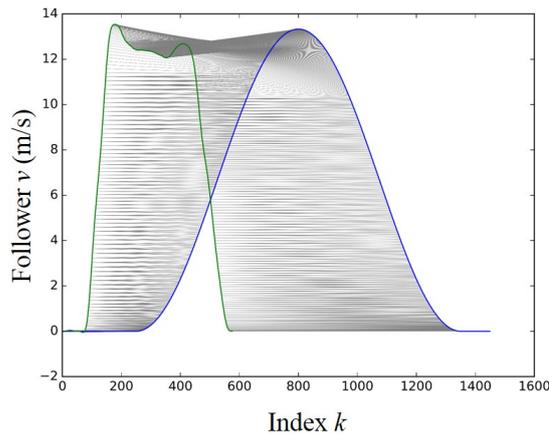
(b) Matched index set between the observed series and the reference series. The green line is the observed series and the blue line is the reference series. Each sampling point of the observed series is bound to the assigned point of the reference series by a gray line.

Figure 3.15: An example of DTW process.

3.2.3 Characteristic Leaves of Different Following Vehicles: Features in Observed Trajectories

In this section, we apply the decision tree analysis based on the shapelet toward the observed series processed by DTW. In following discussion, we call the trajectories shown in Figure 3.11 as characteristic leaves, after the example given by Ye *et al.* We investigate the features of the characteristic leaves for respective following vehicles. In Table 3.1, we listed the trials utilized the analysis. Because we instructed different steady velocity in the experiments held in December, 2015 and September, 2016, we could not treat them as one dataset. In the following analysis, we focus on the data obtained in September, 2016. Some of the observed data were removed because they did not contain sufficient data at the end of the trials when the followers were stopping.

Table 3.1: Utilized trials for the decision tree analysis with the shapelets.

Follower	Leader	Experiment day	Test driver (follower)	Trial number	Trial number of combination
MC	MC	Sep., 2016	TD6	7	25
	C		TD6	13	
	T		TD7	5	
C	MC	Sep., 2016	TD4	1	34
			TD5	6	
	C		TD4	6	
			TD5	7	
	T		TD4	11	
			TD5	3	
T	MC	Sep., 2016	TD4	6	20
	C		TD4	6	
	T		TD4	8	

3.2.3.1 Features in averaged characteristic leaves

Before we proceed with the decision tree analysis, we refer to the "averaged" characteristic leaves in the case of each follower type. Figures 3.16, 3.17, and 3.18 are the averaged characteristic leaves in the case of each follower type. Physical values, i.e., the distance gap s, follower velocity v, velocity difference Δv, and follower acceleration a, at the same indices of all the series processed by DTW were averaged. The averaged characteristic leaves in the case of each follower (the motorcycles, cars, and trucks) are indicated by a blue, orange, and black solid line. The gray solid line indicates the averaged trajectory over all observed series. Dashed lines of each color indicate averaged value $\pm$ standard error for each follower type.

From Figure 3.16, we can confirm that the characteristic leaf of the motorcycles is smaller than those of other followers from the distance gap perspective. This feature will be detected by the decision tree analysis. In Figure 3.17, we cannot find any obvious features for v direction. Regarding the follower a direction, we find small separations in the case of the truck in the acceleration and deceleration phase. In Figure 3.17, we find that the leaf in the case of the motorcycle is smaller than those of others in Δv direction. Furthermore, the shape difference between the cars and trucks is also visible. These features are also expected to be found by the tree analysis.
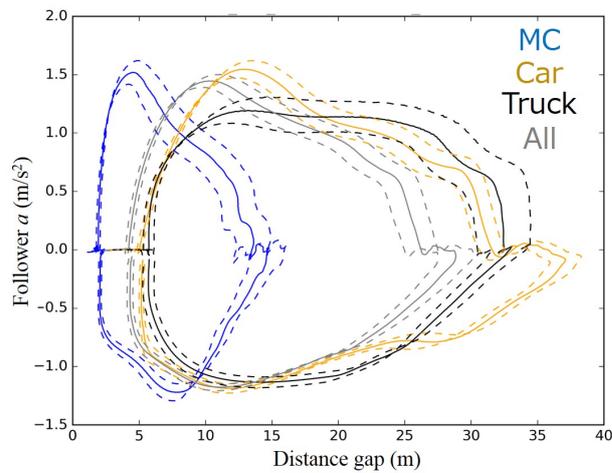
Figure 3.16: Trajectory for each following vehicle on the distance gap and follower acceleration plane. Averaged data over the same indices on DTW series are shown. Dashed lines indicate the standard error of each case.
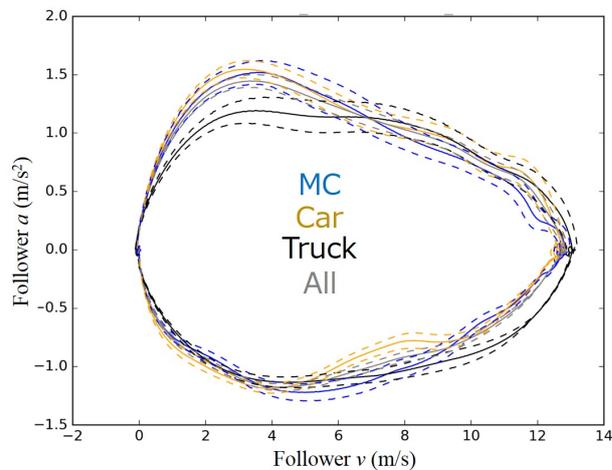
Figure 3.17: Trajectory for each following vehicle on the follower velocity and follower acceleration plane. Averaged data over the same indices on DTW series are shown. Dashed lines indicate the standard error of each case.
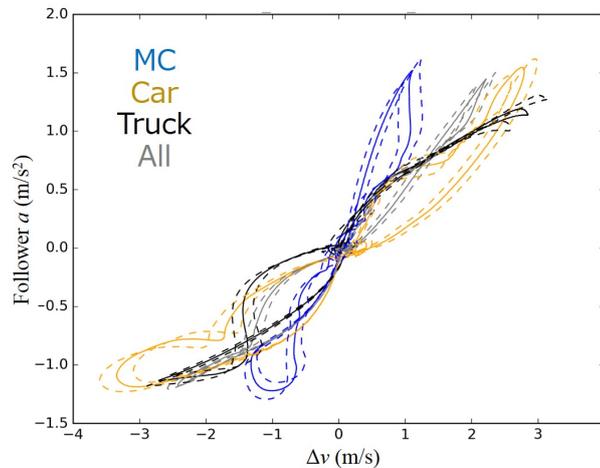
Figure 3.18: Trajectory for each following vehicle on the velocity difference and follower acceleration plane. Averaged data over the same indices on DTW series are shown. Dashed lines indicate the standard error of each case.

From Figure 3.19 to 3.22, we indicate the evolutions of each physical value on the common index k in order to clarify whether the difference of the physical values with the different followers is universal among drivers. Because there were two drivers who were led by all the types of the leaders only for the following cars, we compared the averaged series of TD4 and TD5, who drove the cars, to the averaged series of the motorcycles and trucks. Imagine that a green line in Figure 3.19 for TD4, who drove the cars, is smaller than a blue line for the motorcycles, and a purple line for TD5, who also drove the cars, is greater than the blue line in some index range. If such a case, we cannot conclude that a shapalet located in the range is applicable to all the drivers. For example, in the lower figure in Figure 3.19, the green and the purple lines hold the black line between them roughly in the range $650 \leq k \leq 850$. If we obtain the shapelet in the range to divide trials of the cars and trucks, the shapelet cannot be utilized. In Figure 3.20 for followers' velocity, we cannot confirm clear inversion of the magnitudes. In the lower figure in Figure 3.21 for the velocity difference, we confirmed the inversion of the magnitude at around $k = 450$ and 650. If we obtain the shapelet around here to divide trials of the cars and trucks, it can not be used. In the upper figure in Figure 3.22 for followers acceleration, we observe the inversion in the range $400 \leq k \leq 500$. The shapelet in the range can not be utilized for the division between the motorcycles and cars. On the other hand, in the lower figure in Figure 3.22, we observe the inversion in the range $550 \leq k \leq 650$. The shapelet in this range can not be used to divide trials of the trucks and cars.
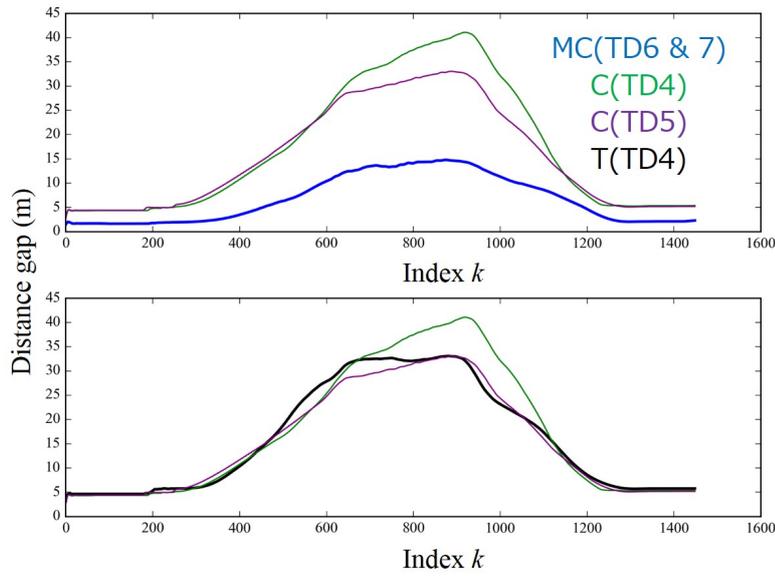
Figure 3.19: Comparison of the distance gap on the common index k in the cases
of the motorcycles and cars, and trucks and cars.
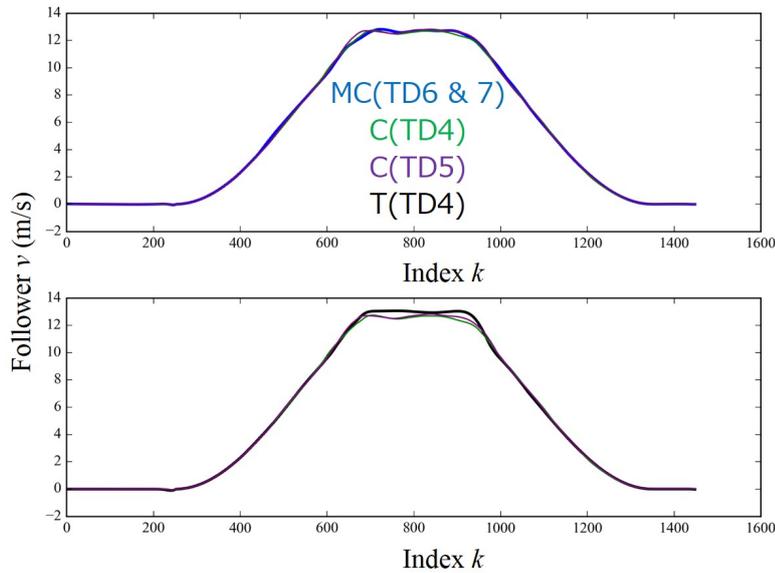
Figure 3.20: Comparison of followers' velocity on the common index k in the
cases of the motorcycles and cars, and trucks and cars.
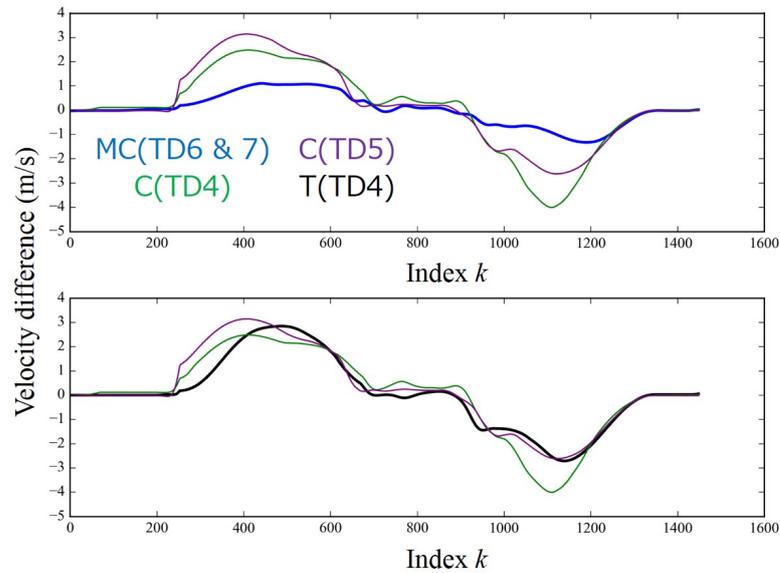
Figure 3.21: Comparison of the velocity difference on the common index k in the cases of the motorcycles and cars, and trucks and cars.
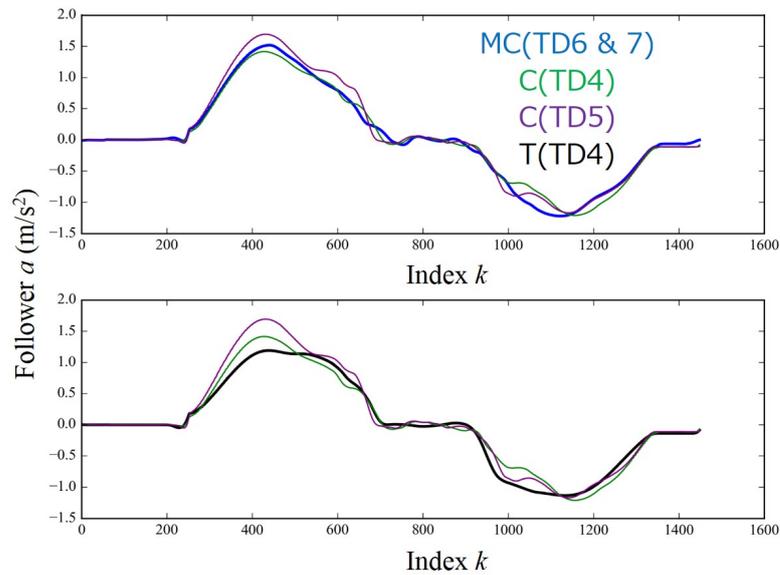
Figure 3.22: Comparison of followers' acceleration on the common index k in the cases of the motorcycles and cars, and trucks and cars.

3.2.3.2 Features of the characteristic leaves from the perspective of the distance gap

In this section, we develop a decision tree for the distance gap using the algorithms introduced in Section 3.2.2 for extracting characteristic features, i.e., the shapelets, observed in the distance gaps of each following vehicle.

As we mentioned in Section 3.2.2, we terminated the tree development and did not add deeper decision rules when one type of the vehicles occupies more than 90 % in a node in order to avoid over-fitting. We also did not add the new nodes when the node has less than 16 trials. That is, if the node is occupied less than 90 % by one type of the vehicles and has as many as or more than 16 trials, we try to add new deeper nodes. On the other hand, when we have a node which cannot be added new nodes because of the number of the trials but is occupied by less than 75 % by one type of the vehicles, we do not utilize the decision rule for the classification because of the lack of the performance.

Figure 3.23 is the decision tree for the distance gap. The tables show the number of vehicles that belong to each node of the tree. The number of vehicles are shown in the column "#." The percentages occupied by each vehicle in each node are shown in the column "% in group." The ratios of the number of vehicles in each node to the total number of each vehicle are shown in the column "% in type." The decision rules based on the extracted shapelets are numbered as "DRn." Each DR divides trials in the upper node into two nodes located on the left and right sides. The trials in the divided node directed by a blue arrow are trials whose trajectories are within of the range determined by the shapelet and d_{OSP}. The remaining trials are divided into the node directed by a red arrow. The obtained gains, d_{OSP}, and margins are shown below the DRs. For example, DR1 extracts 20 trials for the motorcycles, which correspond to 80 % of all motorcycle trials. In addition, "% in group" for the motorcycles becomes 100 % as the divided group comprises only the motorcycle trials. We denote "MC" next to the table because we can regard the node extracted by DR1 as for the motorcycles owing to the high occupancy ratio of the motorcycle.

From Figure 3.24 to 3.29, the shapelets utilized for each DR are shown. The shapelet used for DR1 is shown in Figure 3.24. Figure 3.24a shows the obtained shapelet by a red line. Because d_{OSP} is the averaged distance which the extracted trajectories should be within, the values of the shapelet series $\pm d_{\mathrm{OSP}}$ are the averaged range in which these trajectories are included. This range is indicated by the dashed lines. The distance gaps for the motorcycles, cars, and trucks on the DTW index k are depicted by the blue dots, orange dots, and black dots, respectively. The horizontal dashed lines indicate the respective phases, i.e., from $k = 0$, the waiting phase, early acceleration phase, latter acceleration phase, steady phase, early deceleration phase, latter deceleration phase, and the stopping phase, respectively.

Root node

type	#	% in group	% in type
MC	25	32%	100%
C	34	43%	100%
T	20	25%	100%

DR1 — Gain: 0.81, d_{OSP}: 8.99, Margin: 20.67

MC

type	#	% in group	% in type
MC	20	100%	80%
C	0	0%	0%
T	0	0%	0%

type	#	% in group	% in type
MC	5	8%	20%
C	34	58%	100%
T	20	34%	100%

DR2 — Gain: 0.34, d_{OSP}: 12.06, Margin: 15.42

type	#	% in group	% in type
MC	4	50%	16%
C	3	38%	9%
T	1	13%	5%

type	#	% in group	% in type
MC	1	2%	4%
C	31	61%	91%
T	19	37%	95%

DR3 — Gain: 0.41, d_{OSP}: 4.43, Margin: 11.50

T

type	#	% in group	% in type
MC	0	0%	0%
C	2	22%	6%
T	7	78%	35%

type	#	% in group	% in type
MC	1	2%	4%
C	29	69%	85%
T	12	29%	60%

DR4 — Gain: 0.27, d_{OSP}: 24.87, Margin: 13.28

type	#	% in group	% in type
MC	0	0%	0%
C	5	63%	15%
T	3	38%	15%

type	#	% in group	% in type
MC	1	3%	4%
C	24	71%	71%
T	9	26%	45%

DR5 — Gain: 0.46, d_{OSP}: 7.34, Margin: 9.36

C

type	#	% in group	% in type
MC	0	0%	0%
C	8	100%	24%
T	0	0%	0%

type	#	% in group	% in type
MC	1	4%	4%
C	16	62%	47%
T	9	31%	45%

DR6 — Gain: 0.44, d_{OSP}: 6.95, Margin: 6.94

C

Type	#	% in group	% in type
MC	1	9%	4%
C	10	91%	29%
T	0	0%	0%

type	#	% in group	% in type
MC	0	0%	0%
C	6	40%	18%
T	9	60%	45%

Figure 3.23: The decision tree for the distance gap based on the shapelet analysis.

Figure 3.24b is the averaged characteristic leaves on the plane of the distance gap and followers' acceleration. The data points corresponding to the range of the shapelet index are emphasized by the bold lines for the reference. We conclude that DR1 is the decision rule to divide the motorcycles using the distance gap during the steady phase ($k = 748$ to 769) from Figure 3.24a. We can confirm the obvious difference in the distance gap between the motorcycles and other vehicles in Figure 3.24b.

On the other hand, DR2 in Figure 3.25 extracts the remaining motorcycles but the extracted node also includes some cars and trucks. We conclude that DR2 does not have the ability to divide trials into specific following vehicle cases.

From DR3, the problem almost becomes the classification of trials for the cars and trucks. DR3 utilizes the shapelet located in the early deceleration phase in Figure 3.26a and extracts the truck trials. DR3 claims that the distance gap of trucks is smaller than that of cars in the early deceleration phase. We can confirm this trend in Figure 3.26b.

Owing to the degree of mixture, we cannot utilize DR4 in extracting the cars trials.

DR5 and DR6 are located in the early deceleration phase, as shown in Figure 3.28a and 3.29a, and they claim that the distance gap in the early deceleration phase becomes larger in the case of cars than that of trucks.

Through the discussion above, we determined the following logical formulas:

$$P_{\mathrm{MC}} \subseteq \mathrm{DR1}\left(P_{\mathrm{all}}\right) \tag{3.27}$$

$$P_{\mathrm{C}} \subseteq \left\{\mathrm{DR5}\left(\overline{\mathrm{DR4}}\left(\overline{\mathrm{DR3}}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\left(P_{\mathrm{all}}\right)\right)\right)\right)\right)\right\}$$
$$\vee \left\{\mathrm{DR6}\left(\overline{\mathrm{DR5}}\left(\overline{\mathrm{DR4}}\left(\overline{\mathrm{DR3}}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\left(P_{\mathrm{all}}\right)\right)\right)\right)\right)\right)\right\} \tag{3.28}$$

$$P_{\mathrm{T}} \subseteq \mathrm{DR3}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\left(P_{\mathrm{all}}\right)\right)\right) \tag{3.29}$$

where P_{all} is all the trial populations and P_{MC}, P_{C}, and P_{T} are the trial populations for the motorcycles, cars, and trucks, respectively. The logical expression $\overline{\mathrm{DR}}(P)$ refers to the remaining population that the decision rule DR extracted from the original population. Based on these logical formulas, the shapelet combinations essential for the classification are plotted in Figure 3.30. We call this diagram a logic diagram of shapelets. The blue areas filled with diagonal lines, the orange areas, and the black areas filled with horizontal lines show the shapelets for the motorcycles, passenger cars, and trucks, respectively. This figure shows that, for example, the series of distance gaps for the cars which are extracted by the decision tree, pass all areas on different indices filled in orange. If there are two areas on the same index, the series extracted by the decision tree passes either of them. There are two logic diagrams since there are two logical formulas for the cars. In both cases, we can insist that distance gap of the motorcycles is smaller than those of the others at around $k = 750$ (the steady phase), and distance gaps of the cars and trucks differ by around $k = 1000$ (the early deceleration phase). These features need to be replicated by the car-following models for the replication of mixed traffic.
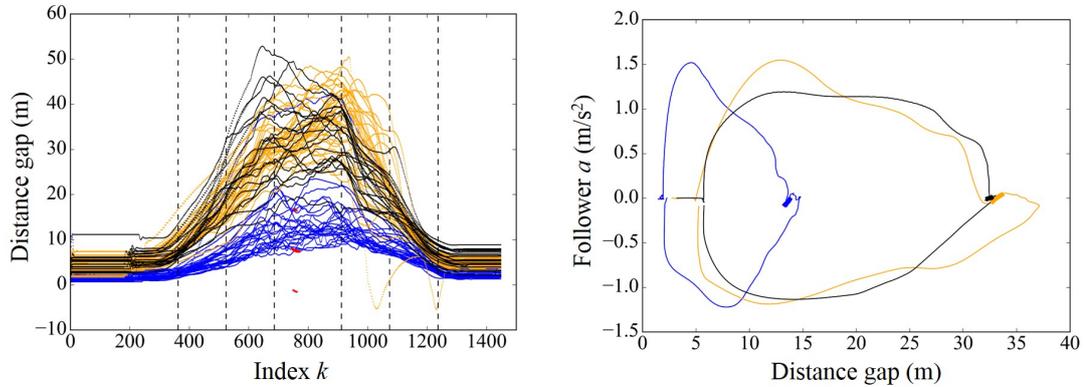
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.24: The shapelet for the distance gap assigned to DR1. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.25: The shapelet for the distance gap assigned to DR2. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.26: The shapelet for the distance gap assigned to decision rule DR3. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.27: The shapelet for the distance gap assigned to decision rule DR4. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
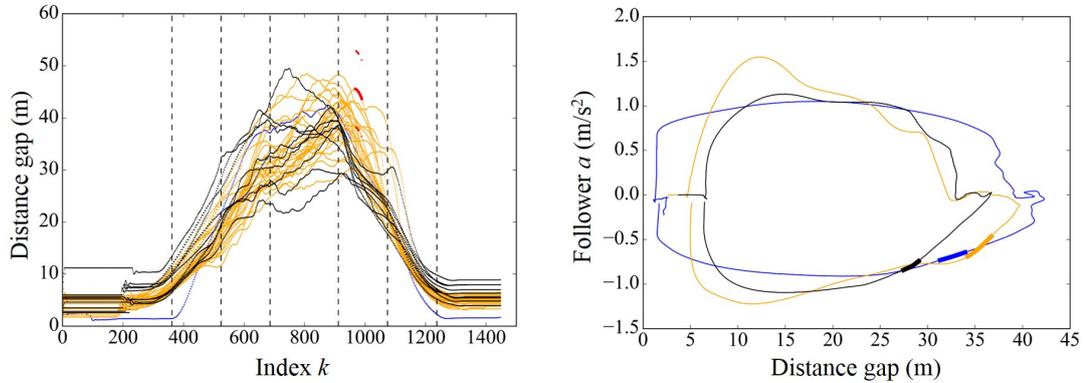
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.28: The shapelet for the distance gap assigned to decision rule DR5. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
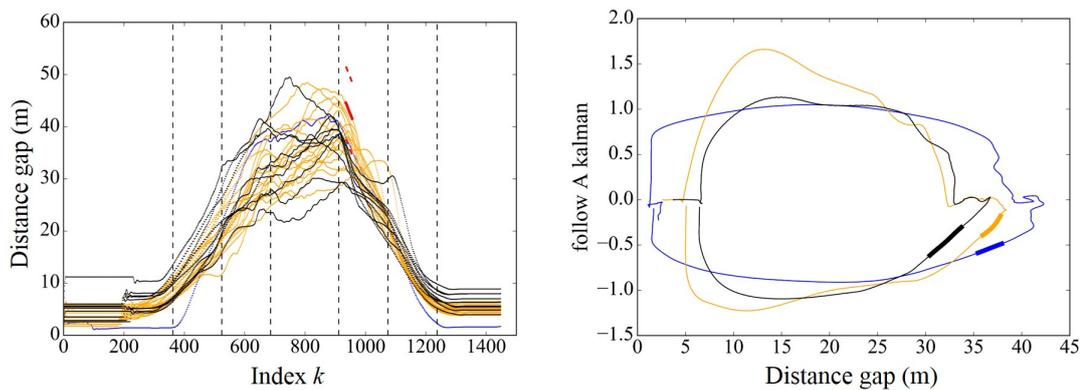
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.29: The shapelet for the distance gap assigned to decision rule DR6. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
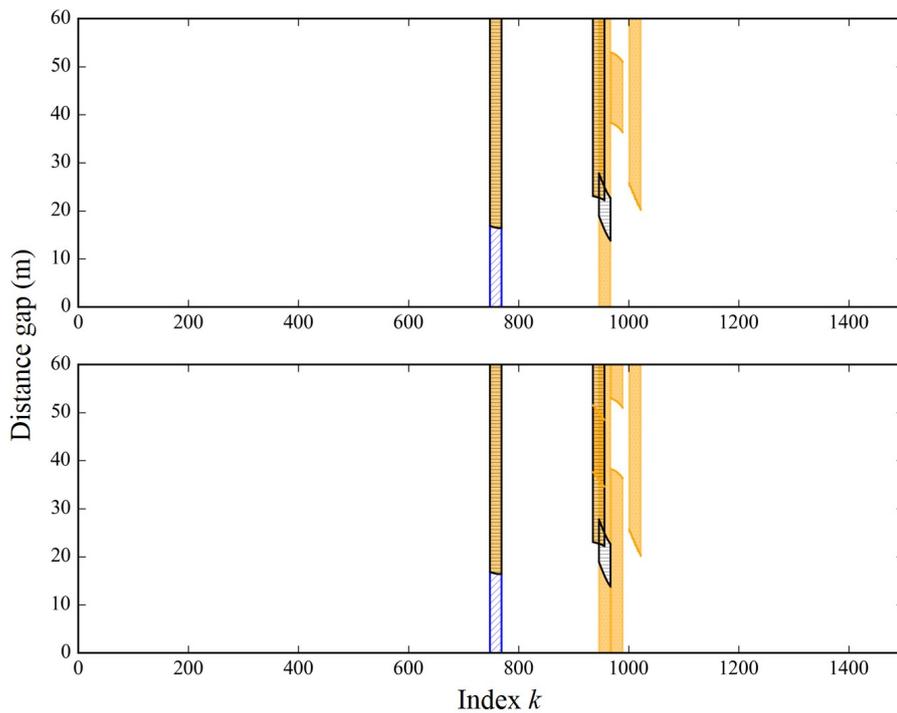
Figure 3.30: The logic diagram of shapelets for the distance gap. The blue areas filled by diagonal lines show the shapelets for the motorcycles. The orange areas show ones for the passenger cars. The black areas filled by horizontal lines show ones for the trucks.

3.2.3.3 Features in the characteristic leaf from the perspective of the velocity difference

We obtained the decision tree for the velocity difference as shown in Figure 3.31

DR1 extracts most of the motorcycle trials. It utilizes the shapelet covering both the early and latter deceleration phases, as shown in Figure 3.32a. The decision rule claims that the motorcycles tend to have smaller velocity difference in the deceleration phases compared to other vehicles. This trends can be recognized in Figure 3.32b.

Owing to the degree of mixture, DR2 can not be utilized as a decision rule.

DR3 extracted the car trials perfectly. It detects the high velocity difference when the vehicle starts accelerating, as shown in Figure 3.34a. The cars tended to have larger velocity difference at the beginning of the acceleration phase. In other words, the drivers of the cars did not start accelerating until the velocity difference became larger than when they were in the trucks.

The decision rule DR4 focuses on the latter deceleration phase, as shown in Figure 3.35a. It extracts the truck trials which have the smaller magnitudes of the velocity difference than those of cars, i.e., the trucks had the larger velocity difference, including the sign that cars make in the latter deceleration phase.
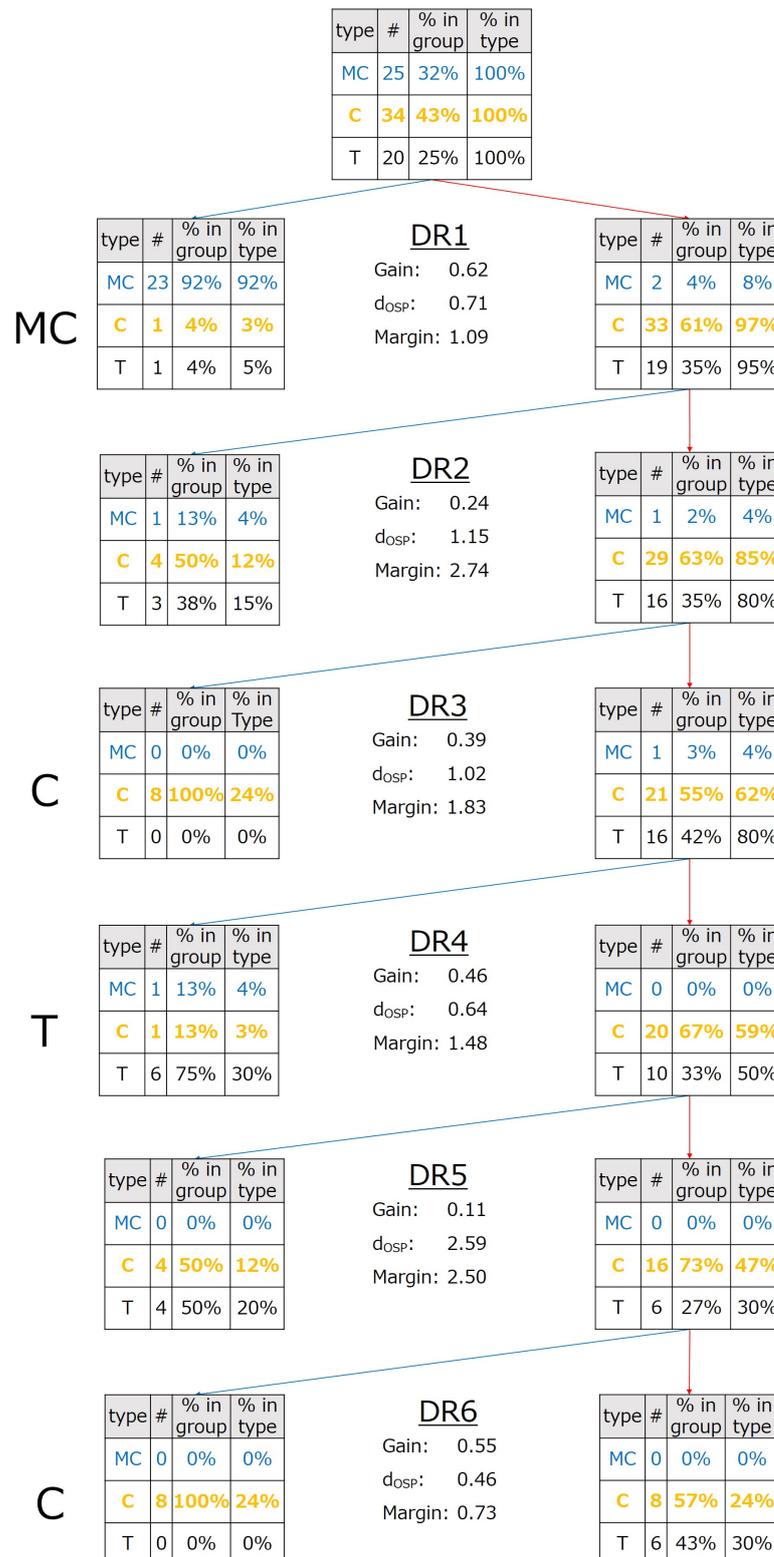
Figure 3.31: The decision tree for the velocity difference based on the shapelet analysis.

DR5 cannot be used for classification because of poor performance. The shapelet is indicate in Figure 3.36a.

The deepest decision rule DR6 detects the car trials perfectly by focusing on the velocity difference at the beginning of the acceleration as DR3. The shapelet is shown in Figure 3.37a The trend which DR6 claims is the same as DR3, i.e., the passenger cars tend to have larger velocity difference than trucks at the beginning of acceleration.

From Figure 3.31, the logical formulas for populations of respective vehicles are written as follows.

$$P_{\mathrm{MC}} \subseteq \mathrm{DR1}\,(P_{\mathrm{all}}) \tag{3.30}$$

$$P_{\mathrm{C}} \subseteq \mathrm{DR3}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\,(P_{\mathrm{all}})\right)\right)$$
$$\vee \left\{\mathrm{DR6}\left(\overline{\mathrm{DR5}}\left(\overline{\mathrm{DR4}}\left(\overline{\mathrm{DR3}}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\,(P_{\mathrm{all}})\right)\right)\right)\right)\right)\right\} \tag{3.31}$$

$$P_{\mathrm{T}} \subseteq \mathrm{DR4}\left(\overline{\mathrm{DR3}}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\,(P_{\mathrm{all}})\right)\right)\right) \tag{3.32}$$

The logic diagram of shapelets for the velocity difference is shown in Figure 3.38. The most prominent feature is the stable evolution of the velocity difference by the motorcycles in the range $k = 905$ to 1266.

In the classification of cars and trucks, we need to focus on the beginning of the acceleration, i.e., $k = 300$ to 400 to detect cars. Some of cars tend to have the larger acceleration than the trucks do. When we focus on the range $k = 1100$, i.e., the latter deceleration phase, we can pick up the truck trials; the trucks tend to have smaller magnitude of velocity difference than cars. When we consider the replication performance of the car-following models, we need to confirm the behaviors when $k = 300$ to 400, $k = 1100$ and stability at $k = 900$ to 1300.
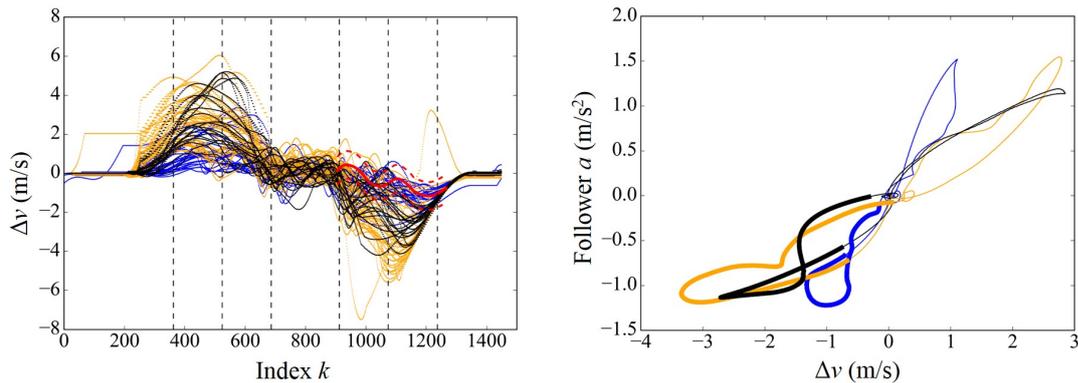
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.32: The shapelet for the velocity difference assigned to DR1. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
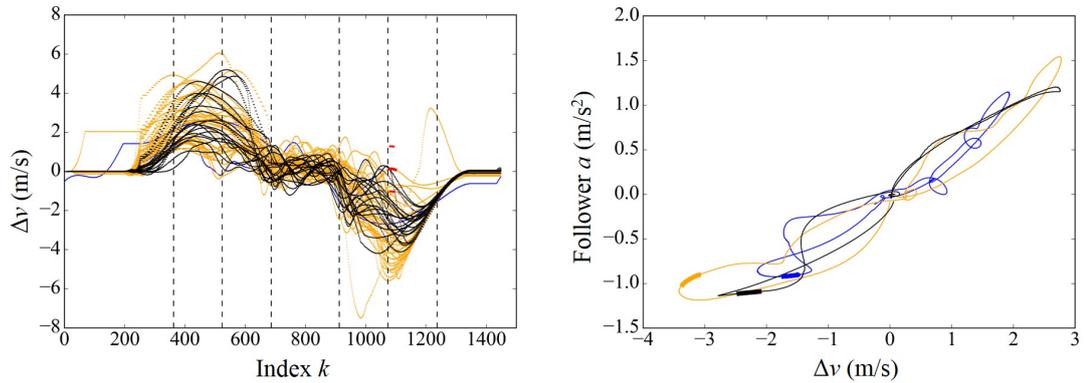
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.33: The shapelet for the velocity difference assigned to DR2. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.34: The shapelet for the velocity difference assigned to DR3. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.35: The shapelet for the velocity difference assigned to DR4. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.36: The shapelet for the velocity difference assigned to DR5. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
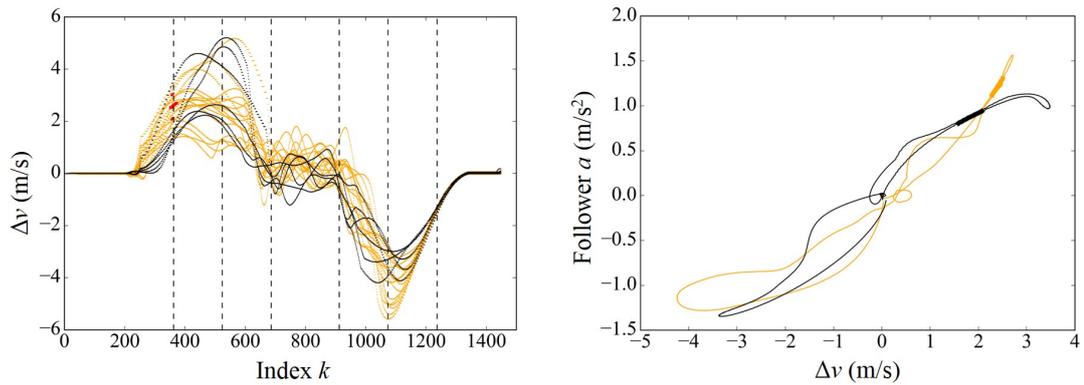
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.37: The shapelet for the velocity difference assigned to DR6. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
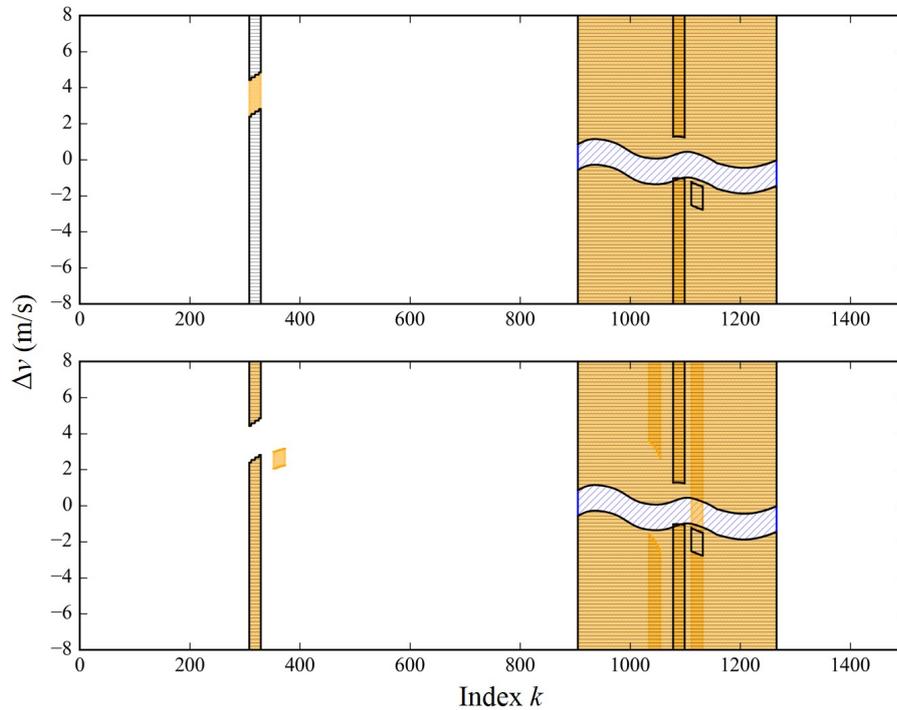
Figure 3.38: The logic diagram of shapelets for the velocity difference. The blue areas filled by diagonal lines show the shapelets for the motorcycles. The orange areas show ones for the passenger cars. The black areas filled by horizontal lines show ones for the trucks.

3.2.3.4 Features of the characteristic leaf from the perspective of the followers' velocity

We obtained the decision tree for followers' velocity difference as shown in Figure 3.39

DR1 extracts part of truck trials. From Figure 3.40a DR1 seems to detect the rapid deceleration of the trucks in the early deceleration phase.

DR2 extracts trials with larger velocity at the end of the steady phase, as shown in Figure 3.41. Neither the extracted node nor the remaining nodes are the terminated nodes supplying obvious classification, and deeper decision rules are required.

DR3 divides the trials which increases the velocity from the end of the acceleration phase to the beginning of the steady phase. It removes the car trials by detecting increase in velocity at the end of the latter acceleration phase and the beginning of the steady phase.

DR4 also removes the motorcycle trials. It focuses on the velocity in the early acceleration phase, as shown in Figure 3.43. The rule achieved perfect extraction for the motorcycles, and the obtained margin was small, i.e., 0.31 m/s. Because of this smallness, we decided not to use DR4 for the classification. A criterion was 5 % of the maximum value of the averaged series of all the series. This criterion is the same to a criterion which to find the location where the series are not varied. The reasons will be explained in detail in Section 3.2.4.

DR5 detects the truck trials by removing trials with the oscillation of the velocity in the steady phase. It detects trials which have relatively large and stable velocities, as shown in Figure 3.44.

DR6 extracts trials in which the velocity in the early deceleration phase passes in a certain range, as shown in Figure 3.44a. From Figure 3.44b, we concluded that DR6 detects the large variations in velocity from $k = 957$ to 978 in the case of cars.

Although the deepest rule DR7 focuses on stable velocity at the steady phase, its classification performance was not sufficient. Therefore, we decided not to use DR7 for the classification.

The logic formulas for followers' velocity can be written as

$$P_C \subseteq \overline{\text{DR3}} \left(\text{DR2} \left(\overline{\text{DR1}} \left(P_{\text{all}} \right) \right) \right)$$
$$\vee \, \text{DR6} \left(\text{DR4} \left(\overline{\text{DR2}} \left(\overline{\text{DR1}} \left(P_{\text{all}} \right) \right) \right) \right) \tag{3.33}$$
$$P_T \subseteq \text{DR1} \left(P_{\text{all}} \right)$$
$$\vee \, \overline{\text{DR5}} \left(\text{DR3} \left(\text{DR2} \left(\overline{\text{DR1}} \left(P_{\text{all}} \right) \right) \right) \right) . \tag{3.34}$$

We could neither obtain the logic formula for the motorcycles nor display the shapelets for the motorcycle on the logic diagram of the shapelets shown in Figure 3.47. This is because there was no decision rule to extract the motorcycle trials obviously in the tree shown in Figure 3.39. Although the diagrams are complex, we concluded that the passenger cars tend to have smooth transfer from the latter acceleration phase to the steady phase, while the some trucks increase their velocity from the latter acceleration phase to the steady phase. On the other

hand, from the end of the steady state to the beginning of the early deceleration phase, the cars tend to maintain high velocity while some trucks decrease the velocity quickly. We need to focus on the index k range from 630 to 810 and from 855 to 978 if the simulated series varied their rates of change. Furthermore, these features are expected to be observed in the feature analysis of followers' acceleration conducted in Section 3.2.3.5.

Root node

type	#	% in group	% in type
MC	25	32%	100%
C	34	43%	100%
T	20	25%	100%

DR1 — Gain: 0.25, d_{OSP}: 0.34, Margin: 0.68

T branch:
type	#	% in group	% in type
MC	0	0%	0%
C	1	13%	3%
T	7	88%	35%

type	#	% in group	% in type
MC	25	35%	100%
C	33	46%	97%
T	13	18%	65%

DR2 — Gain: 0.20, d_{OSP}: 0.51, Margin: 0.56

type	#	% in group	% in type
MC	10	27%	40%
C	14	37%	41%
T	13	35%	65%

type	#	% in group	% in type
MC	15	44%	60%
C	19	56%	56%
T	0	0%	0%

DR3 — Gain: 0.36, d_{OSP}: 1.14, Margin: 0.56

type	#	% in group	% in type
MC	10	40%	40%
C	4	16%	12%
T	11	44%	55%

C branch:
type	#	% in group	% in type
MC	0	0%	0%
C	10	83%	29%
T	2	17%	10%

DR4 — Gain: 0.47, d_{OSP}: 0.07, Margin: 0.31

type	#	% in group	% in type
MC	5	21%	20%
C	19	79%	56%
T	0	0%	0%

MC branch:
type	#	% in Group	% in type
MC	10	100%	40%
C	0	0%	0%
T	0	0%	0%

DR5 — Gain: 0.38, d_{OSP}: 0.35, Margin: 0.52

type	#	% in group	% in type
MC	9	56%	36%
C	4	25%	12%
T	3	19%	15%

T branch:
type	#	% in group	% in type
MC	1	11%	4%
C	0	0%	0%
T	8	89%	40%

DR6 — Gain: 0.42, d_{OSP}: 0.70, Margin: 0.70

C branch:
type	#	% in group	% in type
MC	0	0%	0%
C	14	100%	41%
T	0	0%	0%

type	#	% in group	% in type
MC	5	50%	20%
C	5	50%	15%
T	0	0%	0%

DR7 — Gain: 0.44, d_{OSP}: 0.37, Margin: 0.31

type	#	% in group	% in type
MC	5	63%	20%
C	0	0%	0%
T	3	38%	15%

type	#	% in group	% in type
MC	4	50%	16%
C	4	50%	12%
T	0	0%	0%

Figure 3.39: The decision tree for followers' velocity based on the shapelet analysis.
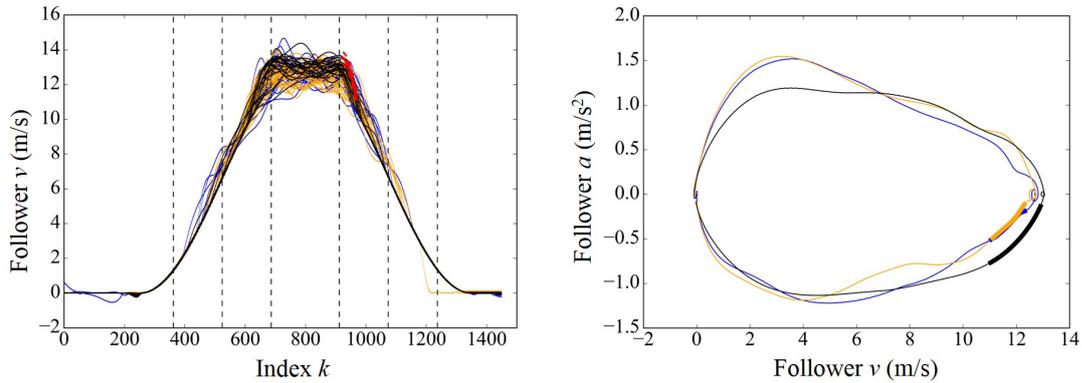
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.40: The shapelet for followers' velocity assigned to DR1. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.41: The shapelet for followers' velocity assigned to DR2. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
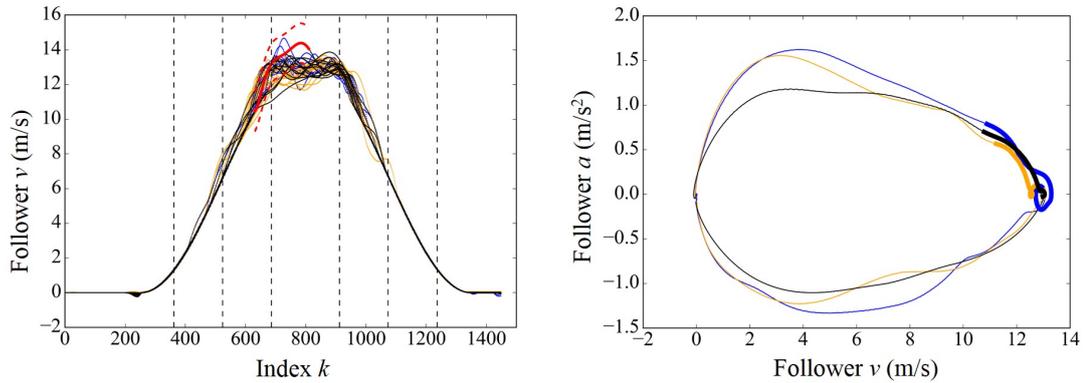
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.42: The shapelet for followers' velocity assigned to DR3. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.43: The shapelet for followers' velocity assigned to DR4. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.44: The shapelet for followers' velocity assigned to DR5. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.45: The shapelet for followers' velocity assigned to DR6. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
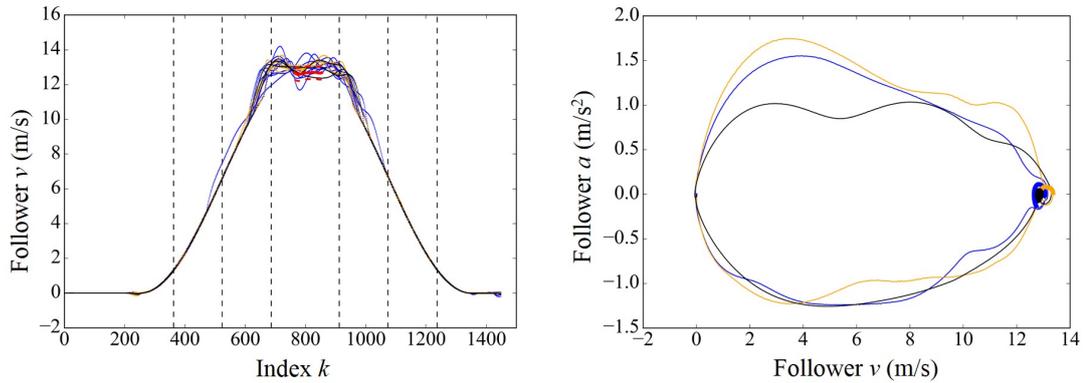
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.46: The shapelet for followers' velocity assigned to DR7. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.
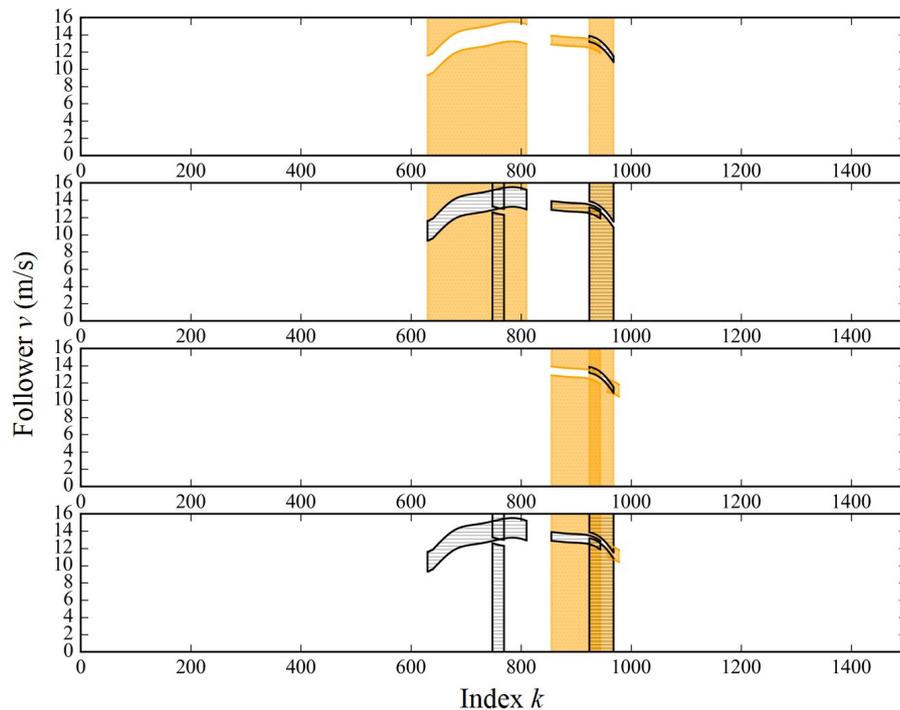
Figure 3.47: The logic diagram of shapelets for followers' velocity. The blue areas filled by diagonal lines show the shapelets for the motorcycles. The orange areas show ones for the passenger cars. The black areas filled by horizontal lines show ones for the trucks.

3.2.3.5 Features of the characteristic leaf from the perspective of the followers' acceleration

We obtained the decision tree for followers' acceleration as shown in Figure 3.48

DR1 detects the trials which have large acceleration in the early acceleration phase as shown in Figure 3.49. The extracted node has more trials of cars than of the motorcycles; it was decided that this decision rule will not be utilized for the classification from the perspective of the classification performance.

DR2 also focuses on large acceleration in the early acceleration phase, as shown in Figure 3.50 and extracts rather more trials of cars than of the other vehicles. However, for the same reason as DR1, it was decided that this decision rule will be utilized as only the intermediate decision rule for the deeper ones.

From DR3 shown in Figure 3.51, we can find that the car trials tend to have smaller acceleration during the latter acceleration phase than those of the others. However, it utilizes the series around $k = 600$ where we implied that the shapelet here would not be universal among drivers from Figure 3.22. Therefore, we decided to remove this decision rule from the classification.

On the other hand, DR4 can be regarded as the rule for classification of the trucks. As shown in Figure 3.52, it covers the early acceleration phase to the early deceleration phase and claims that the truck trials tend to have both increased acceleration in the latter acceleration phase and quick rise of the deceleration. This shapelet grasps the features mentioned in Section 3.2.3.4.

The terminated nodes by DR5 and DR6 shown in Figure 3.53 and 3.54, respectively, which are focusing on the acceleration phase, cannot be utilized for the classification because of the performance in extraction. However, they can be followed by new nodes as they have sufficient number of trials in the other nodes.

Following DR7, which is shown in Figure 3.55a and the deepest decision rule, has the perfect performance for the classification. It claims that the cars trials have the acceleration remaining in the latter acceleration phase. However, it utilizes the acceleration around $k = 600$ where the tendency is not universal among drivers from Figure 3.22. Because the left node divided by DR6 has sufficient performance for the classification, we decided to use not DR7 but DR6 for extraction of the car trials. DR6 claims that the cars tend to have larger acceleration change from the early acceleration phase to the end of the latter acceleration phase than the trucks. This tendency is universal among drivers in Figure 3.22.

Through the discussions above, the logical formulas can be written as

$$P_{\mathrm{C}} \subseteq \overline{\mathrm{DR6}}\left(\mathrm{DR5}\left(\overline{\mathrm{DR4}}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\left(P_{\mathrm{all}}\right)\right)\right)\right)\right) \tag{3.35}$$

$$P_{\mathrm{T}} \subseteq \mathrm{DR4}\left(\overline{\mathrm{DR2}}\left(\overline{\mathrm{DR1}}\left(P_{\mathrm{all}}\right)\right)\right). \tag{3.36}$$

The logic diagram is depicted in Figure 3.56 from these formulas. First, the shapelet covering both the acceleration phase and deceleration phase can be utilized for classification of the trucks. On the other hand, the car trials tend to have larger acceleration change from the early acceleration phase to the end of the latter acceleration phase than those of the others.
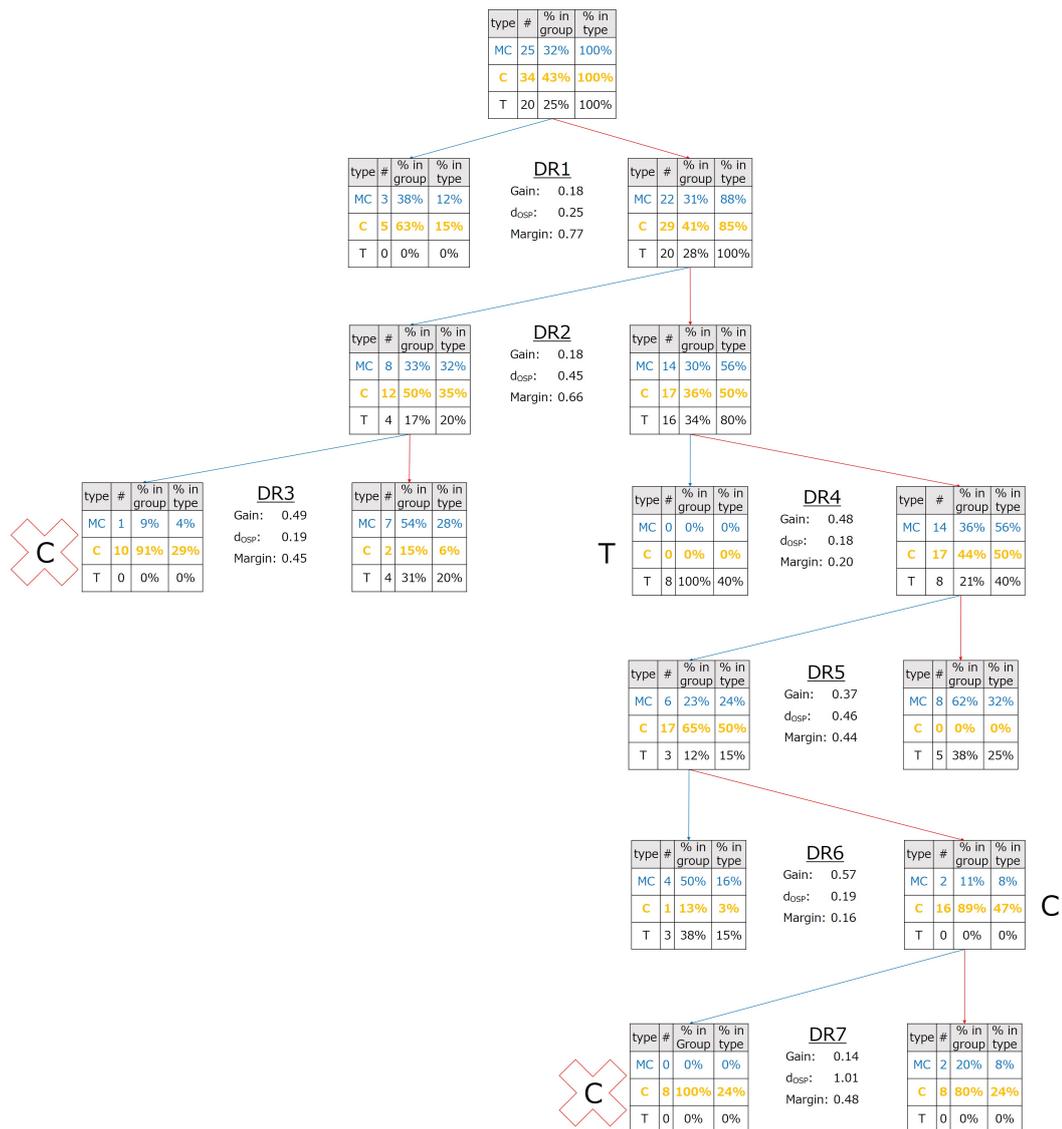
Figure 3.48: The decision tree for followers' acceleration based on the shapelet analysis.

From the discussion above, the deceleration in the early deceleration phase in the range $k = 1000$ to 1100 should be varied by the parameters in the car-following models. The acceleration peak should be located at the latter acceleration phase in the case of the trucks. In addition, the difference of the acceleration change covering whole the acceleration phase needs to be replicated. In Figure 3.57, we show two cases of characteristic accelerations of the cars and trucks, and their difference. When the car trials have their acceleration peak in the early acceleration phase, the difference of these accelerations will have sign alternation in the acceleration phase, as shown in Figure 3.57a. When the car trials have their peak of the acceleration in the latter acceleration phase as the trucks, the difference of the accelerations will not have sign alternation. In both cases, the sign in the early deceleration phase and major sign in the acceleration phase need to be the same.

Because the features extracted from the acceleration have the same meanings as the features of the velocity discussed in Section 3.2.3.4, we decided not to discuss the replicability of the velocity but to discuss replicability of the acceleration in Section 3.4.
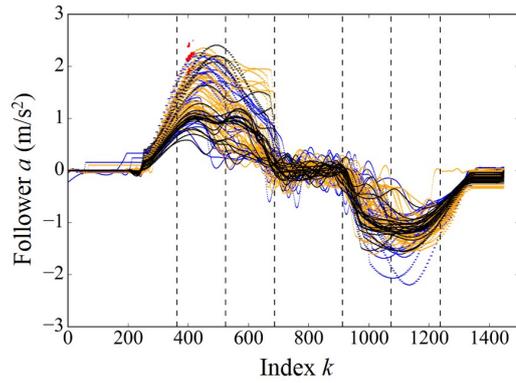
(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.49: The shapelet for followers' velocity assigned to DR1. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.50: The shapelet for followers' velocity assigned to DR2. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.51: The shapelet for followers' velocity assigned to DR3. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.52: The shapelet for followers' velocity assigned to DR4. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.53: The shapelet for followers' velocity assigned to DR5. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.54: The shapelet for followers' velocity assigned to DR6. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

(a) The obtained shapelet on the index k. A red solid line indicates the shapelet. Red dashed lines indicate the range for the group selection.

(b) The averaged characteristic leaves on the gap distance and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(c) The averaged characteristic leaves on the velocity difference and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

(d) The averaged characteristic leaves on followers' velocity and the acceleration plane for each following vehicle. Bold area indicates where the shapelet is.

Figure 3.55: The shapelet for followers' velocity assigned to DR7. Blue, orange, and black lines show the trajectories when the followers were the motorcycles, cars, and trucks, respectively.

Figure 3.56: The logic diagram of shapelets for followers' acceleration. The blue areas filled by diagonal lines show the shapelets for the motorcycles. The orange areas show ones for the passenger cars. The black areas filled by horizontal lines show ones for the trucks.

(a) When the trial of cars has acceleration peak at the early acceleration phase.

(b) When the trial of cars has acceleration peak at the latter acceleration phase.

Figure 3.57: Schematics of the accelerations of the cars and trucks, and their difference. The abbreviations of EA, LA, ST, ED indicate the phases, i.e., the early acceleration phase, the latter acceleration phase, the steady phase and the early deceleration phase, respectively.

3.2.4 Acceptable Variability and Fixed Ranges in Trajectories for Different Following Vehicle

In Section 3.2.3, we investigated features of the characteristic leaves by the shapelets. The simulated series calculated by different parameters need to be varied basically in the index ranges of the shapelets. In contrast, we can presume that there are index ranges where the series should not be varied by parameter change. We call this index range as a "fixed range." Furthermore, even if the variation is acceptable in a certain range of the indices, it is preferable that the acceptable magnitude of variation is recognized. The acceptable magnitude of a certain index range will be determined by comparison of the degree of variation in the range under discussion and other index ranges.

In this section, we first introduce the "local maximum margin" indicating the acceptable degree of variation. We also introduce the "local maximum gain." We can search the fixed ranges using both the local maximum margin and local maximum gain. The acceptable degrees of variation and the fixed ranges for the distance gap, the velocity difference, and the acceleration will be introduced at the end of this section.

In the development of the decision tree based on the shapelets, we divided the data series into two nodes based on the best gain and margin obtained with a certain shapelet. The divided series can be regarded as two series replicated by two different values of a certain parameter of the car-following model. The difference in a certain index range between series replicated by different parameters corresponds to the best margin because the best margin is the difference between the averaged distances of two divided nodes from the shapelet. Therefore, the trend of the magnitude of variation for each index needs to follow that of the best margin for each index.

To obtain the best margin for each index, we introduce the "local maximum margin," which is the best margin obtained by the shapelets with minute length. In other words, for every index, we try to divide all observed series by small shapelets on these indices and search the best margins for every indices. Here, we determined the length of minute shapelets as 22, which is the minimum length of shapelets that we applied in the decision tree analysis in Section 3.2.3.

Furthermore, we can imagine that the variations of replicated series in a certain index range are not acceptable if the observed series are not apparently divided into two nodes or if the margin between two nodes is too small even if the series are clearly divided. As we have already introduced the local maximum margin indicating the acceptable magnitude of variation on every index; we also introduced the local maximum gain by these minute shapelets in the same manner. That is, the local maximum gain is the best gain obtained by the minute shapelets on every index with which we try to divide all series into two nodes.

Figure 3.58 to 3.60 indicates the local maximum margins and the local maximum gains for the distance gap, the velocity difference, and followers' acceleration, respectively. For the fixed ranges, we determined the margin threshold of 5 % of the maximum value of the average of all the series. We also determined the gain threshold of 0.09. These thresholds are depicted by red dashed lines in

each figure. Note that this gain threshold corresponds to situation that six trials for the motorcycles are extracted from the original dataset to a new divided node, while six trials for the cars are extracted to the other divided node. The remaining trials are equally distributed to both the nodes.

Regarding the distance gap, we could not find any point less than the thresholds, i.e., there are no fixed ranges in Figure 3.58. We can conclude that the magnitude of variation of the waiting and stopping phases should be smaller than the other phases from the local maximum margin in Figure 3.58a.

From Figure 3.59a, we can claim that the variation of the velocity difference in the steady phase should be smaller than in the acceleration and the deceleration phases. In addition, there are some ranges below the thresholds in Figure 3.59, i.e., $k \leq 230$ and $1254 \leq k$ for the margin, and $748 \leq k \leq 769$ in the gain. Variation of the series is not acceptable in these ranges.

Regarding the acceleration, the magnitude of variation in the waiting, stopping, and steady phases need to be suppressed when they are compared with the acceleration and the deceleration phases from Figure 3.60a. Plus, the ranges $k \leq 263$, $1309 \leq k \leq 1341$ in the margin, and $825 \leq k \leq 835$ in the gain need to be fixed ranges from Figure 3.60 as the margin and the gain are below the thresholds.

(a) The local maximum margin and the margin threshold.

(b) The local maximum gain and the gain threshold.
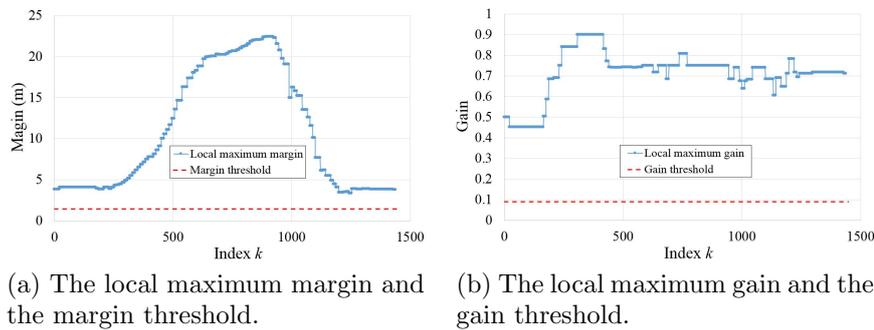
Figure 3.58: The local maximum margin and gain for the gap distance.

(a) The local maximum margin and the margin threshold.

(b) The local maximum gain and the gain threshold.
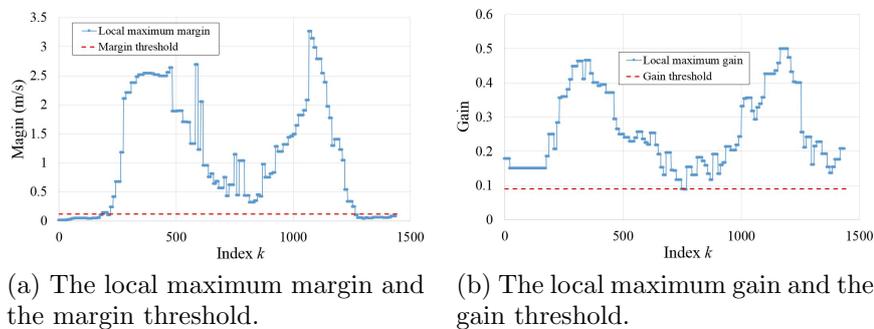
Figure 3.59: The local maximum margin and gain for the velocity difference.

(a) The local maximum margin and the margin threshold.

(b) The local maximum gain and the gain threshold.
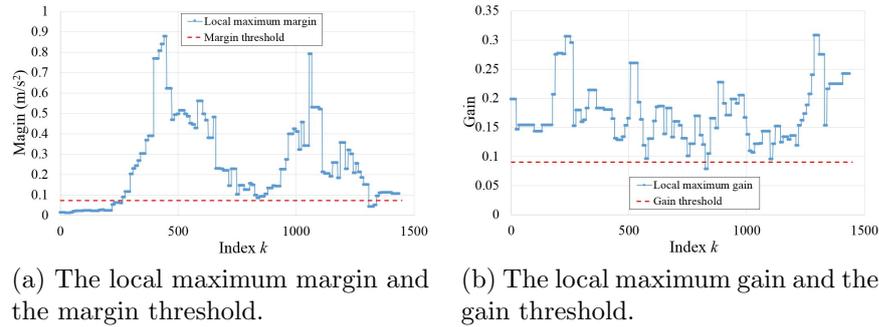
Figure 3.60: The local maximum margin and gain for followers' acceleration.

From the discussions in Section 3.2.3 and 3.2.4, we obtained the features in observed characteristic leaves, as listed in Table 3.2 and 3.3. The abbreviations of MC, C, and T stand for the motorcycles, cars, and trucks, respectively. The abbreviations of W, A, EA, LA, ST, D, ED, LD, SP indicate phases, i.e., the waiting phase, whole acceleration phase, early acceleration phase, latter acceleration phase, steady phase, whole deceleration phase, early deceleration phase, latter deceleration phase and stopping phase, respectively. The abbreviations of acc. and dec. stand for the acceleration and deceleration. We investigate whether the respective car-following models can replicate these features by changing parameters of them in following sections.

Table 3.2: Extracted features of the characteristic leaves from the shapelets. The abbreviations of MC, C, and T stand for the motorcycles, cars, and trucks, respectively. The abbreviations of W, A, EA, LA, ST, D, ED, LD, SP indicate the respective phases. The abbreviations of acc. and dec. stand for the acceleration and deceleration, respectively.

	MC	C & T
Distance gap	Smaller than the others @ ST	T<C @ ED
Velocity difference	Small variation @ ST & D	T<C @ beginning A, C<T @ LD
Acceleration	None	C>T@ ED, peak of T @ LA, C>T (acc. change) @ A

Table 3.3: Located positions of the fixed ranges in respective physical values.

	Position of fixed range
Distance gap	None
Velocity difference	W, ST, SP
Acceleration	W, ST, SP

3.3 Simulated Car-Following Behaviors accompanied by Variable Change

In this section, we simulate followers' behaviors, i.e., the characteristic leaves while varying parameters of the respective car-following models introduced in Section 3.1.2.

3.3.1 Simulation Setting

The platoon simulated in this section also comprises two vehicles, i.e., the leader and the follower, as in the experiment. While giving a fixed velocity configuration of the leader, we obtained various car-following behaviors under parameter change. The velocity configuration given to the leader is shown in Figure 3.61. The leader accelerates at $a^{\mathrm{l}}_{\mathrm{simmax}} = 1$ m/s^2 until it reaches the maximum velocity of $V^{\mathrm{l}}_{\mathrm{simmax}} = 15$ m/s. The leader maintains a steady velocity of $V^{\mathrm{l}}_{\mathrm{simmax}}$ for $T_{\mathrm{const}} = 30$ s, and then starts deceleration of $a^{\mathrm{l}}_{\mathrm{simmin}} = -1$ m/s^2. When the leader stops, the simulation continues for $T_{\mathrm{wait}} = 30$ s.
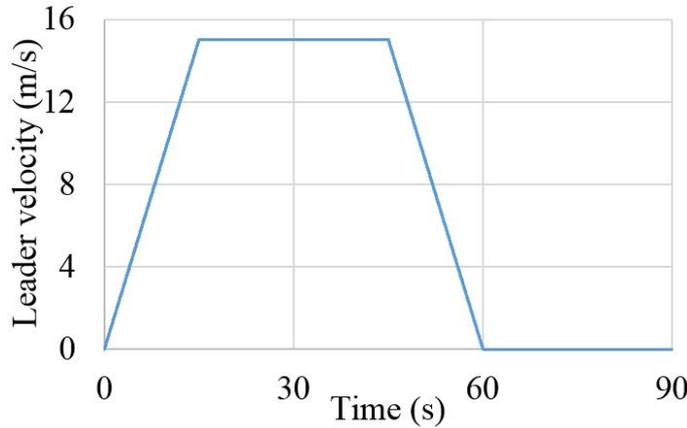
Figure 3.61: Velocity configuration of the leading vehicle in the simulation.

Regarding the OV model, although Bando *et al.* did not take the vehicle lengths into account in OV model formulation [62], Bando *et al.* presented the OV model considering the vehicle lengths afterwards [64]. Assuming h_{min} as the minimum headway distance, the model can be written as

$$a(t) = A\left\{V_{\mathrm{opt}}(h(t)) - v(t)\right\} \tag{3.37}$$

$$V_{\mathrm{opt}}(h) = V_{\mathrm{m}}\frac{\tanh\left\{m(h - \Delta)\right\} - \tanh\left\{m(h_{\mathrm{min}} - \Delta)\right\}}{1 - \tanh\left\{m(h_{\mathrm{min}} - \Delta)\right\}}. \tag{3.38}$$

Note that we added a term to the denominator of function V_{opt} to treat V_{m} as the realizable velocity as in other models. Bando *et al.* provided fitted values of parameters based on the car-following experiment on the Chuo motorway. The parameters which we applied are listed in Table 3.4. Note that V_{m}, s_{min}, L^{f} and L^{l} are equalized to those of other models for comparison. The value of h_{min} is

calculated as $s_{\min} + (L^{\mathrm{l}} + L^{\mathrm{l}})/2$; the lengths of both leaders and followers are considered in the headway distance. By varying these parameters, we obtained trajectories responding parameter variations in Section 3.3.2. When we investigated the response of the characteristic leaves to a certain parameter, we varied the focusing parameter from 50 % to 150 % of the "center value when varied" in Table 3.4, at the step of 10 %. On the other hand, we fixed the other parameter values as "standard value" in Table 3.4.

Table 3.4: Parameters of the OV model.

	Standard value	Center value when varied
V_{m}	15.0 m/s	15.0 m/s
m	0.086	0.086
$s_{\min}$	0.0 m	10.0 m
$L^{\mathrm{f}}(L^{\mathrm{l}})$	4.0 m	4.0 m
Δ	25.0 m	25.0 m
A	2.0	2.0

The parameters of the FVD model were obtained from [67] and formulations are written as

$$a(t) = \frac{V_{\mathrm{opt}}(s(t)) - v(t)}{\tau} + \gamma \Delta v(t) \tag{3.39}$$

$$V_{\mathrm{opt}}(s) = V_{\mathrm{m}} \frac{\tanh\left(s/\Delta s - \beta\right) + \tanh\beta}{1 + \tanh\beta}. \tag{3.40}$$

Table 3.5 lists the parameter sets. As $s_{\min}$ is excluded from the model formulation, we basically set $s_{\min} = 0$ and located leaders and followers, where the distance gap $s = 0$ at the start of the simulation trials. This is the reason why we basically fixed $s_{\min} = 0$ in the simulations for the OV model.

Table 3.5: Parameters of the FVD model.

	Standard value	Center value when varied
V_{m}	15.0 m/s	15.0 m/s
Δs	8.0 m	8.0 m
β	1.5	1.5
$s_{\min}$	0.0 m	10.0 m
$L^{\mathrm{f}}(L^{\mathrm{l}})$	4.0 m	4.0 m
τ	0.65 s	0.65 s
γ	0.6 /s	0.6 /s

Regarding the ID model, the formulation is written as Equation 3.13 and 3.14. The parameters of the ID model were obtained from [67] and listed in Table 3.6. The reason why we applied the small value of 0.01 m instead of zero for the standard value of $s_{\min}$ is to avoid the negative infinite acceleration. The ID model contains s as a denominator in Equation 3.13.

Table 3.6: Parameters of the ID model.

	Standard value	Center value when varied
V_{m}	15.0 m/s	15.0 m/s
δ	4	4
T	1.0 s	1.0 s
$s_{\min}$	0.01 m	10.0 m
$L^{\mathrm{f}}(L^{\mathrm{l}})$	4.0 m	4.0 m
A	1.0 m/s^2	1.0 m/s^2
B	1.5 m/s^2	1.5 m/s^2

Although the original formulation of the Helly model is written as Equation 3.18 and 3.19, We utilized the model presented by Xing in 1995 [102]. The formulation was based on a simplified GHR model [4] and is integrated into the Helly model as

$$a(t) = \alpha \Delta v(t - T_1) + \beta \left\{ h(t - T_2) - D\left(v(t - T_2)\right) \right\} \tag{3.41}$$
$$D\left(v(t)\right) = \gamma + \delta v(t). \tag{3.42}$$

The parameters of this model are listed in Table 3.7.

Table 3.7: Parameters of the Helly model.

	Standard value	Center value when varied
α	0.5 /s	0.5 /s
β	0.05 /s^2	0.05 /s^2
γ	7.0 m	7.0 m
δ	0.5 s	0.5 s
T_1	0.83 s	0.83 s
T_2	3.43 s	3.43 s
$s_{\min}$	0.0 m	10.0 m
$L^{\mathrm{f}}(L^{\mathrm{l}})$	4.0 m	4.0 m

The parameters of GHR model are listed in Table 3.8, which were introduced in the review paper [4] as one of reliable fitting results. These values correspond to those of the model presented by Herman *et al.* [69]. In order to keep the model formulation similar to the model which Herman *et al.* presented, we fixed $k = 1$ and $m = 0$.

Table 3.8: Parameters of the GHR model.

	Standard value	Center value when varied
C	6.035 m/s	6.035 m/s
m	0	Fix
k	1	Fix
T	1.2 s	1.2 s
$s_{\min}$	0.0 m	10.0 m
$L^{\mathrm{f}}(L^{\mathrm{l}})$	4.0 m	4.0 m

3.3.2 Characteristic Leaves of Respective Car-Following Models

In this section, we introduce the simulated characteristic leaves for each parameter value. The leaves for respective values are represented by different colors. From 50 % to 150 % of the center value, the line colors are gradually changed from blue to red. The leaf by 100 % of the center value is represented by white lines. All simulations were calculated by the Euler method at intervals of 0.01 s because some of the models included time delay. Although the OV model, Helly model and GHR model determine the relationship between the headway distance and the acceleration, we investigated the distance gap in these models because the features of the characteristic leaves of the distance gap were extracted for the comparison. Note that, in this thesis, the relation between the distance gap s and the headway distance h was determined as $h = s - (L^{\mathrm{f}} + L^{\mathrm{l}})/2$. The distance gap is the distance between the rear bumper of the leading vehicle and the front bumper of the following vehicle. The headway distance is the distance between the centers of the leading and following vehicles.

3.3.2.1 Optimal Velocity Model

Figure 3.62 to 3.67 are the characteristic leaves on the plane of followers' velocity vs. followers' acceleration, and the distance gap and followers' acceleration when the parameters V_{m}, m, $s_{\min}$, L^{f} or L^{l}, Δ, and A are varied, respectively. In Figure 3.62, one characteristic leaf cannot go around completely owing to the lack of the velocity. We can observe the velocity variations in Figure 3.62a when V_{m} is less than the maximum velocity of the leaders. We also confirm the variation of the distance gap at the steady phase in Figure 3.62b.
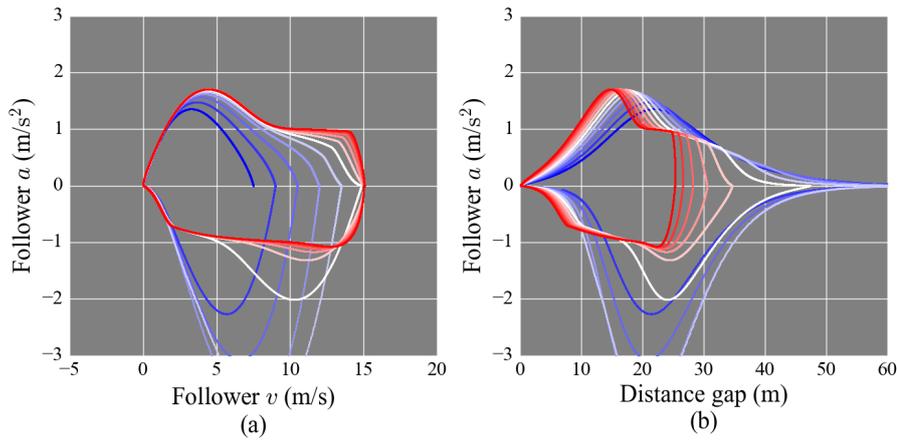
Figure 3.62: Response of the characteristic leaf to the variation of V_{m}.
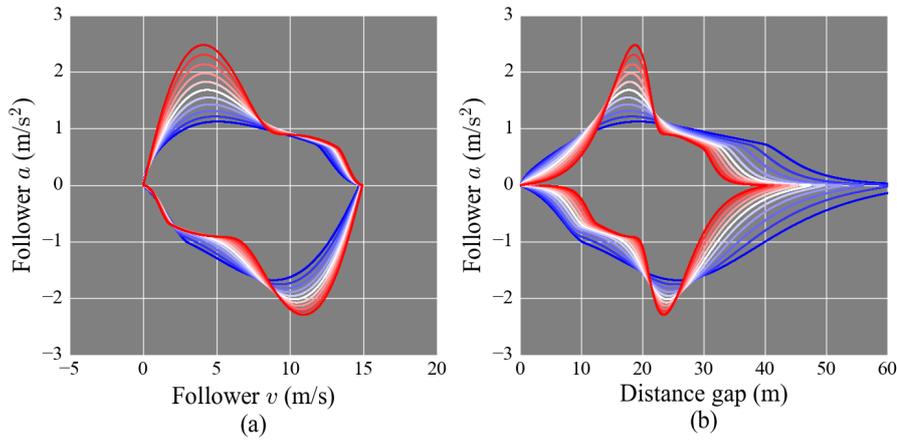
Figure 3.63: Response of the characteristic leaf to the variation of m.

In Figure 3.64b, there seem to be no variation of the distance gap at the steady phase though the initial distance gaps are varied by the change of s_{min}.
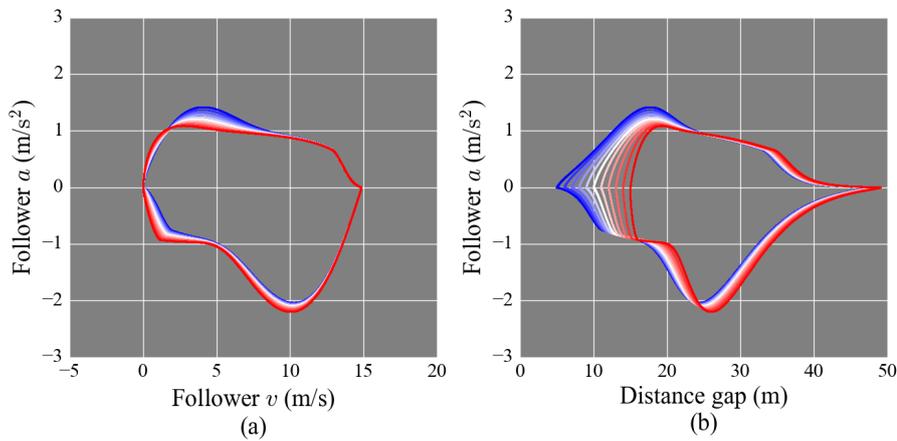
Figure 3.64: Response of the characteristic leaf to the variation of s_{min}.

From Figure 3.65, the variation of L^{f} and L^{l} in the OV model does not seem to significantly affect the characteristic leaf.
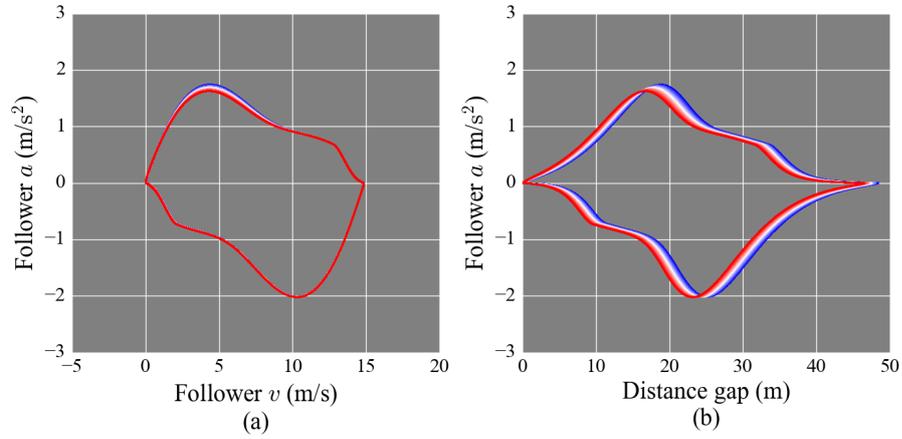
Figure 3.65: Response of the characteristic leaf to the variation of $L^{\mathrm{f}}(L^{\mathrm{l}})$.

In Figure 3.66, we can observe the unnecessary acceleration variation with parameter variation.
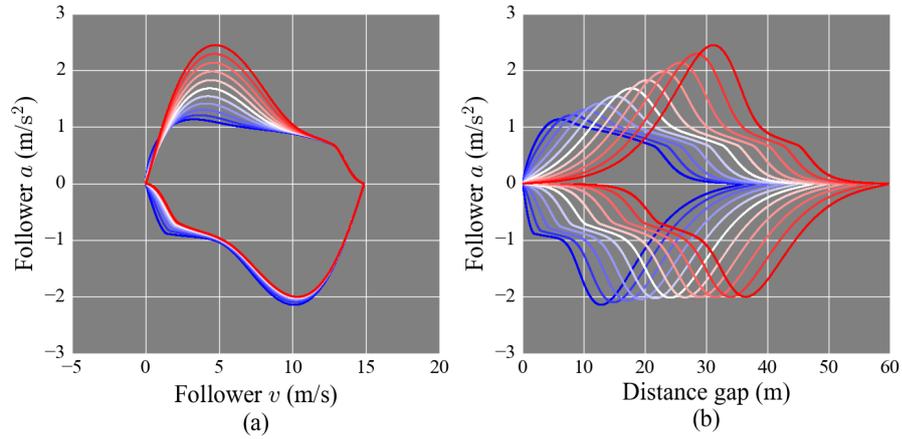
Figure 3.66: Response of the characteristic leaf to the variation of Δ.

As parameter A is response sensitivity, it does not affect values in the steady phase, as shown in Figure 3.67.
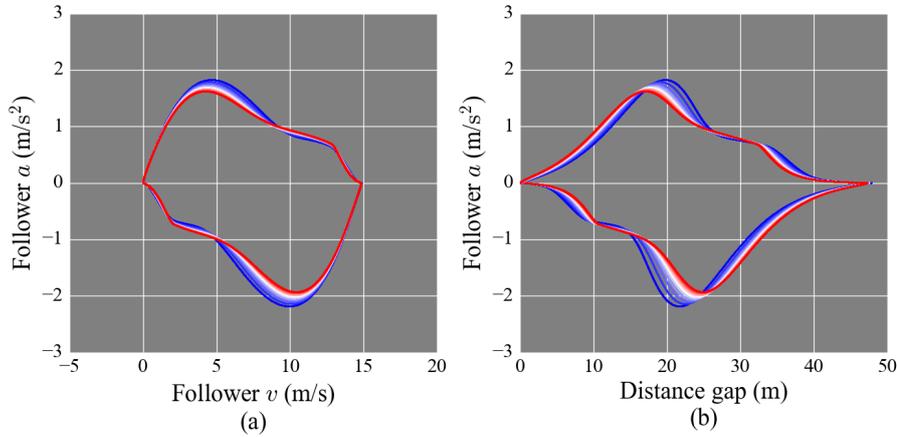
Figure 3.67: Response of the characteristic leaf to the variation of A.

3.3.2.2 Full Velocity Difference Model

From Figure 3.68 to 3.74, we show the response of characteristic leaves to the parameter variation in the FVD model. When we compared them to those of the OV model, the leaves had complex shapes. It seemed to be the effect of the term of $\gamma \Delta v$ added to the OV model.

Figure 3.68: Response of the characteristic leaf to the variation of V_{m}.
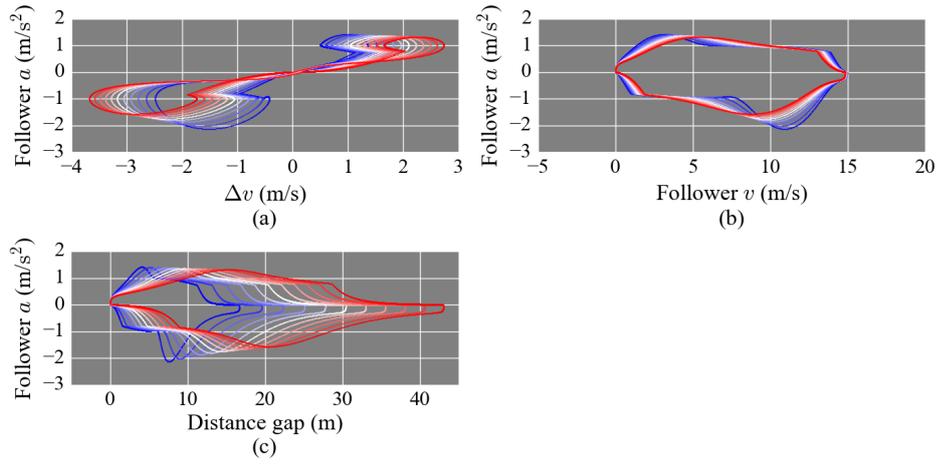
Figure 3.69: Response of the characteristic leaf to the variation of Δs.
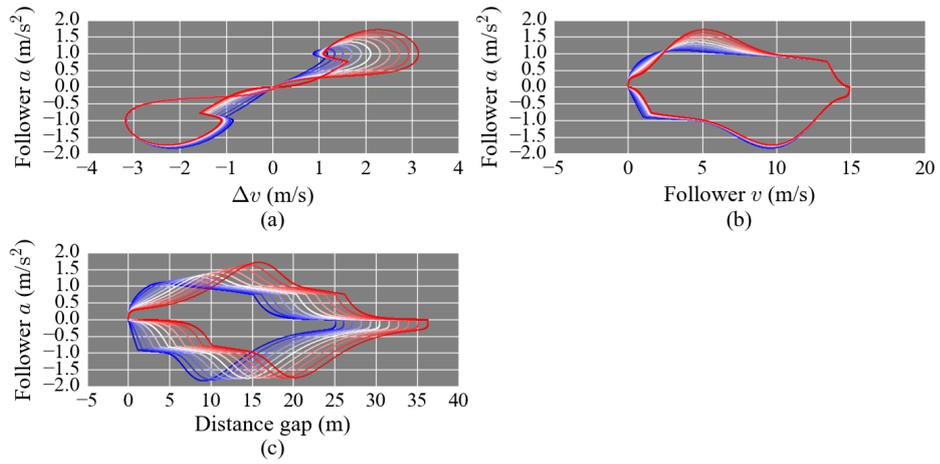
Figure 3.70: Response of the characteristic leaf to the variation of β.
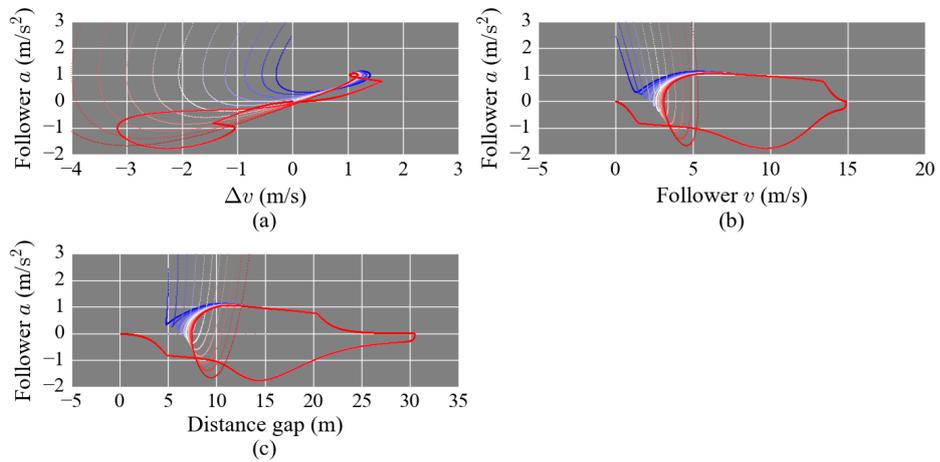
Figure 3.71: Response of the characteristic leaf to the variation of $s_{\min}$.
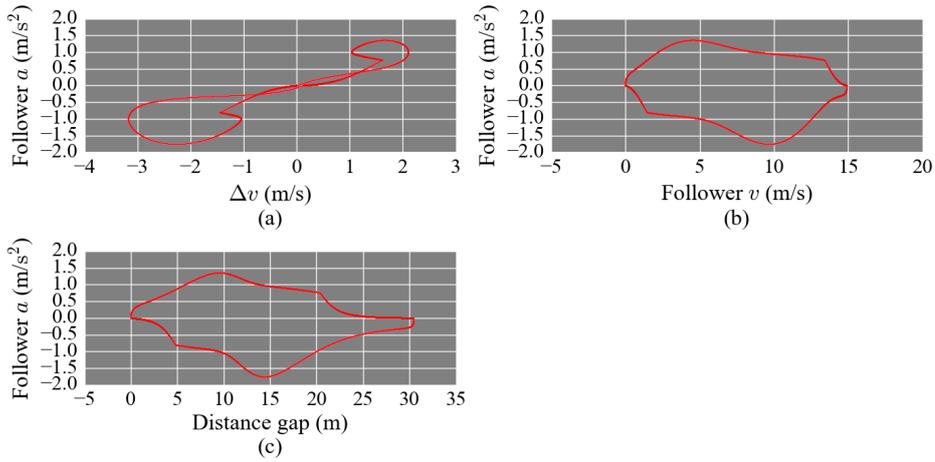
Figure 3.72: Response of the characteristic leaf to the variation of $L^{\mathrm{f}}(L^{\mathrm{l}})$.
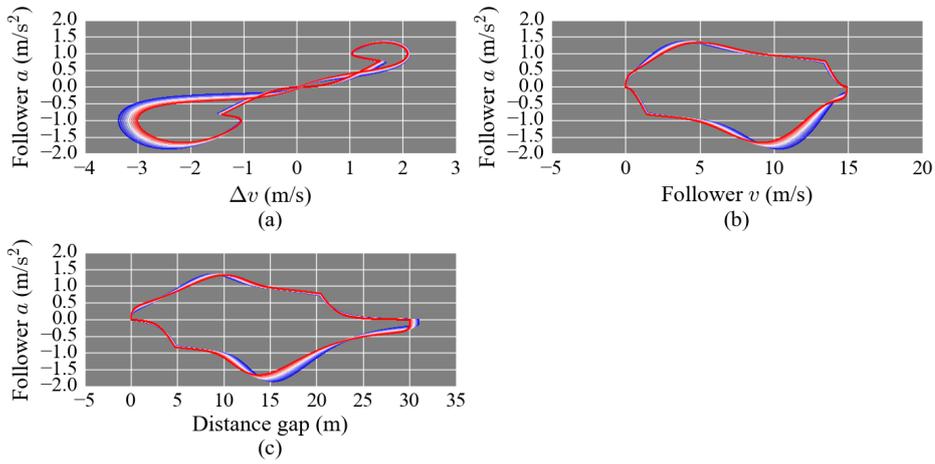
Figure 3.73: Response of the characteristic leaf to the variation of τ.

Figure 3.74: Response of the characteristic leaf to the variation of γ.

3.3.2.3   Intelligent Driver Model

In the simulation of the ID model, we assumed the velocity in the range $-1 \times 10^{-5} < v < 0$ m/s as zero. This is because, in the program language Python, the negative computation error disabled the power of a decimal, which is included in the term of $(v/V_{\mathrm{m}})^\delta$. We confirmed that the calculation had ignorable differences in cases with and without this condition when the exponents were the natural numbers. In addition, some simulation results diverged when the followers ended their deceleration owing to the overtaking accompanied by small fluctuations in the acceleration. We did not utilize for the analysis data after the divergence if the overtaking occurred.
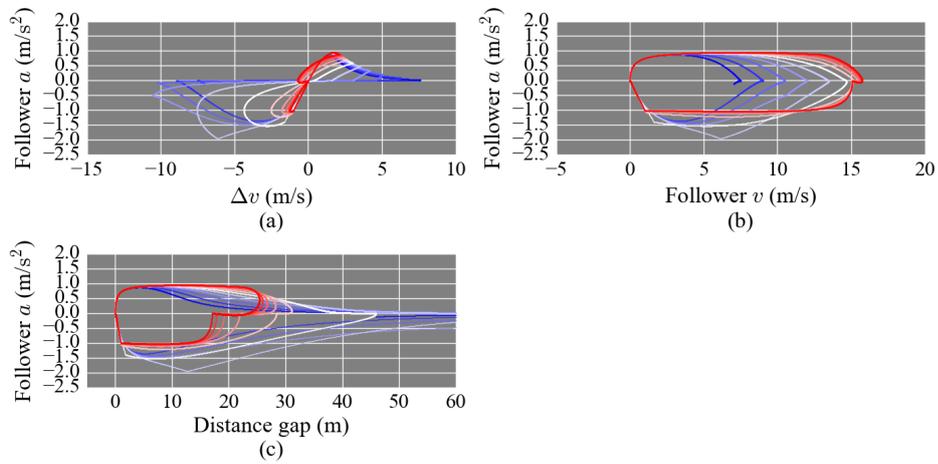
Figure 3.75: Response of the characteristic leaf to the variation of V_{m}.
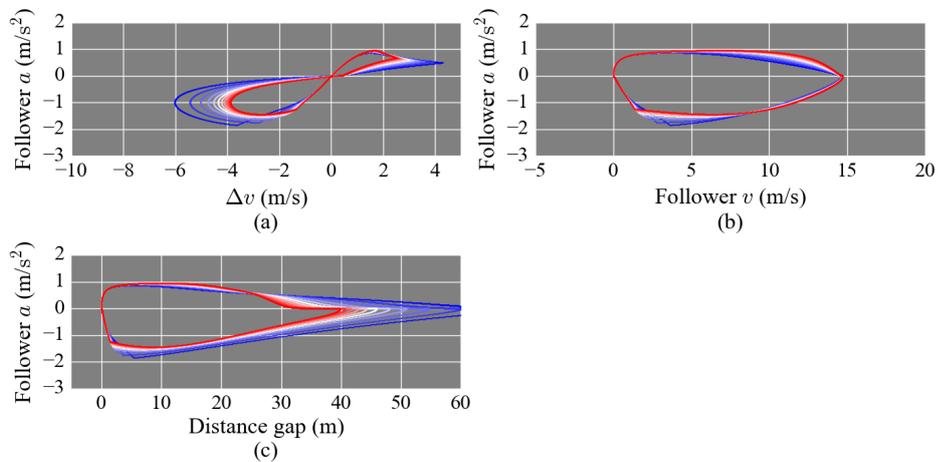
Figure 3.76: Response of the characteristic leaf to the variation of δ.
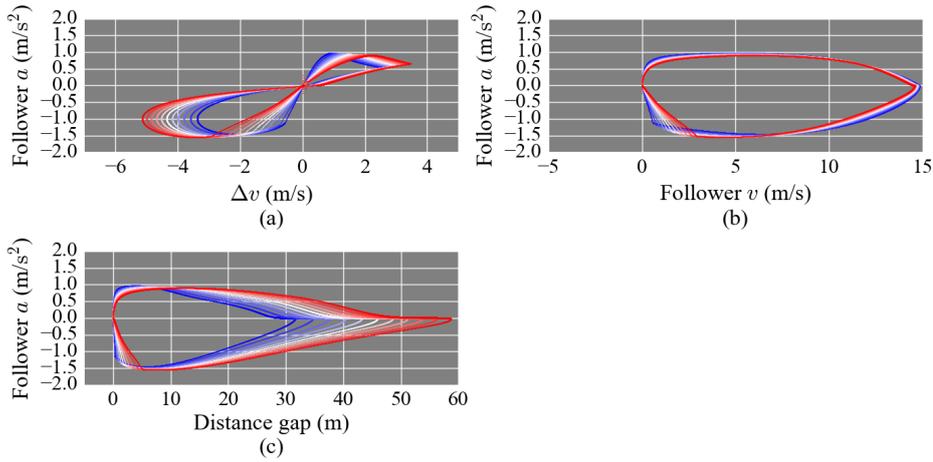
Figure 3.77: Response of the characteristic leaf to the variation of T.
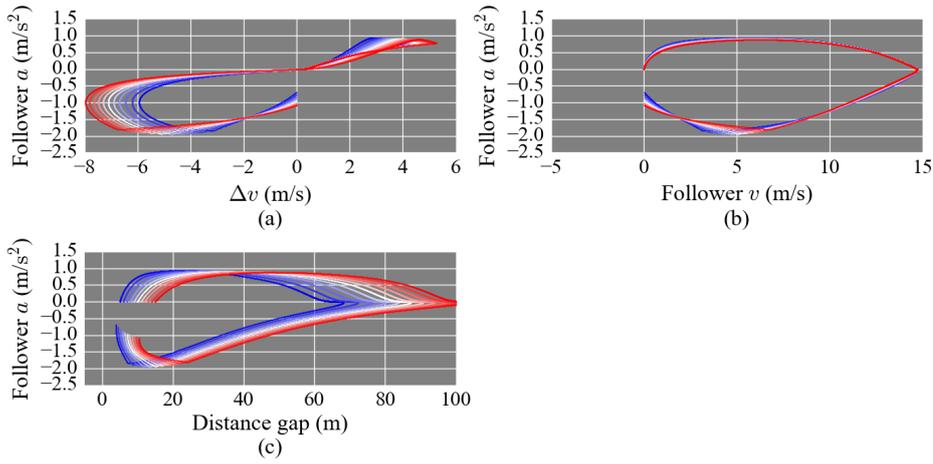
Figure 3.78: Response of the characteristic leaf to the variation of $s_{\min}$.
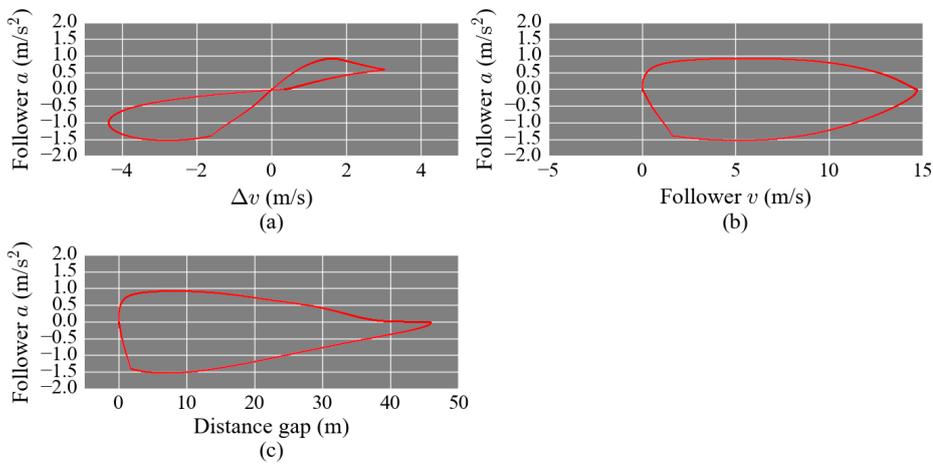
Figure 3.79: Response of the characteristic leaf to the variation of $L^{\mathrm{f}}(L^{\mathrm{l}})$.
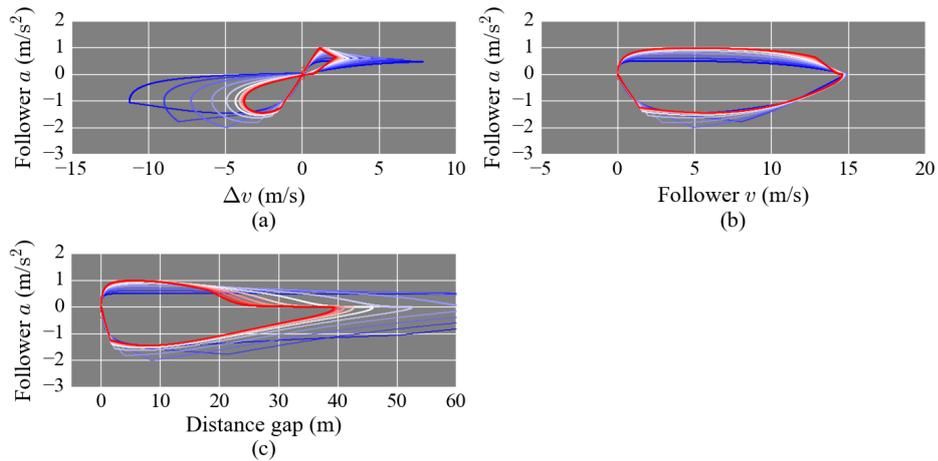
Figure 3.80: Response of the characteristic leaf to the variation of A.
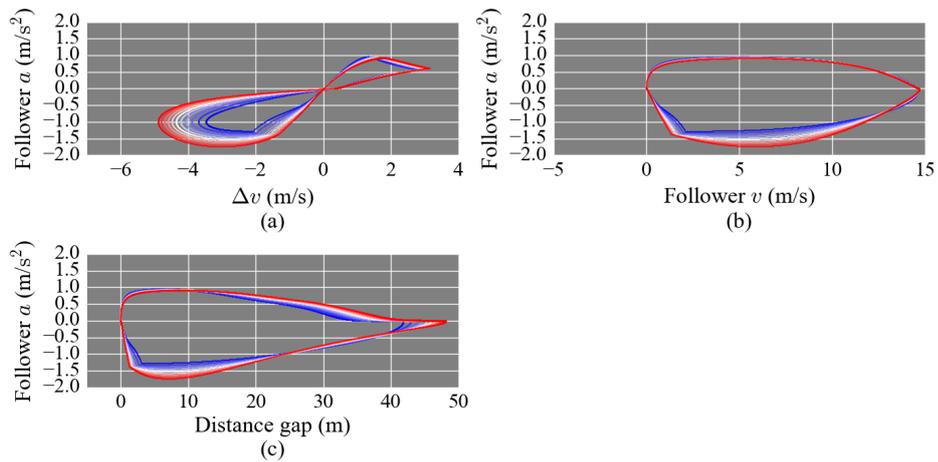
Figure 3.81: Response of the characteristic leaf to the variation of B.

3.3.2.4 Helly Model

From Figure 3.82 to 3.89, we show the response of the characteristic leaves to the variation of respective parameters in the Helly model. The shapes of the characteristic leaves are totally different from the observed ones shown in Figure 3.16 to 3.18. Concerning the velocity difference Δv vs. followers' acceleration, the acceleration starts decreasing after Δv becomes less than zero. Because of this behavior, we observe the overshooting of the distance gap before the follower starts traveling in the steady velocity.

In Figure 3.90, we plotted the values of each term in the Helly model and leaders' velocity v^l. The blue and green dashed lines show the values of each term in the Helly model. The red line indicates follower's acceleration and the black one indicates the velocity of the leader. We can observe that the peaks of the acceleration and deceleration are maintained steady values until the slopes of leader's velocity, i.e., leader's acceleration is changed. Actual drivers would relax

the acceleration gradually when the velocity becomes high. In fact, such behavior
can be seen in the characteristic leaves obtained from the experiments. Although
we can confirm that the steady acceleration and deceleration are caused by the
conflict of the two terms in the Helly model, it is easily assumed that changing
the ratio of α and β will cause other problems because it would be different to
which term the drivers are sensitive in various situations. In other words, the
linearity of the Helly model is not suitable to simulate the stop-and-go pattern,
which we are discussing in this thesis. It seems to be preferable to apply the Helly
model with limitation of the velocity range, etc. From this reason, we remove the
Helly model from the comparison of the simulated and the observed characteristic
leaves.

Figure 3.82: Response of the characteristic leaf to the variation of α.
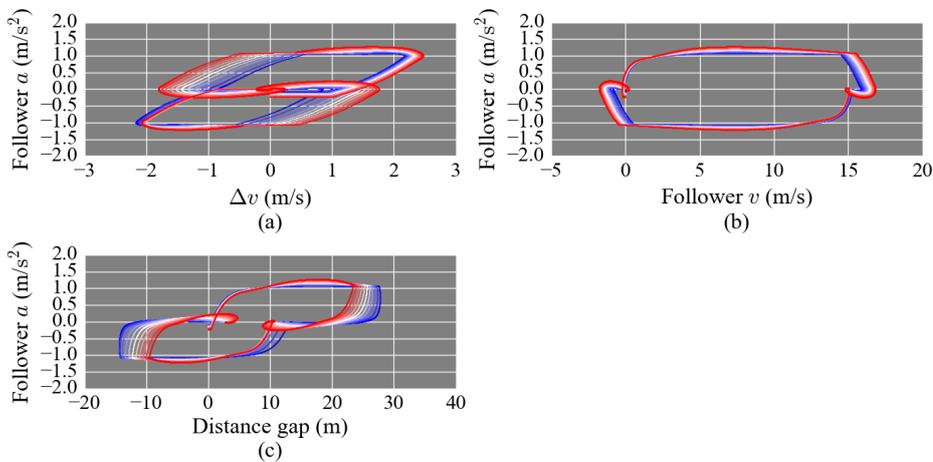
Figure 3.83: Response of the characteristic leaf to the variation of β.
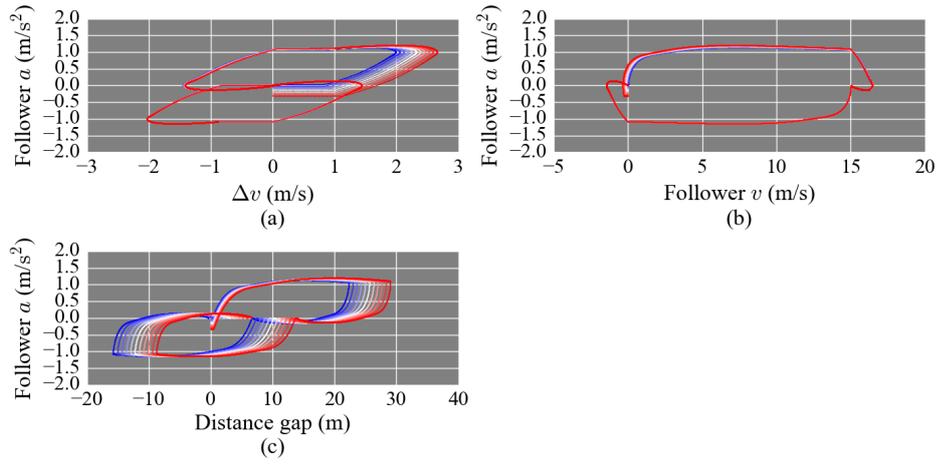
Figure 3.84: Response of the characteristic leaf to the variation of γ.
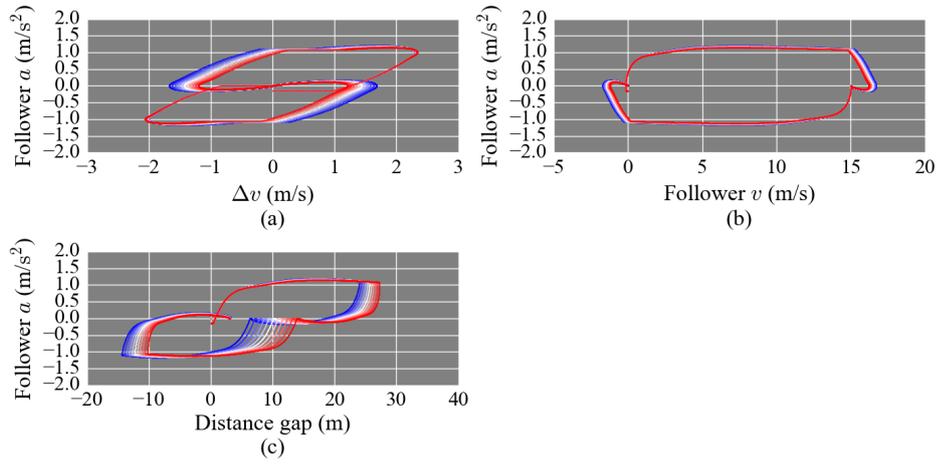
Figure 3.85: Response of the characteristic leaf to the variation of δ.
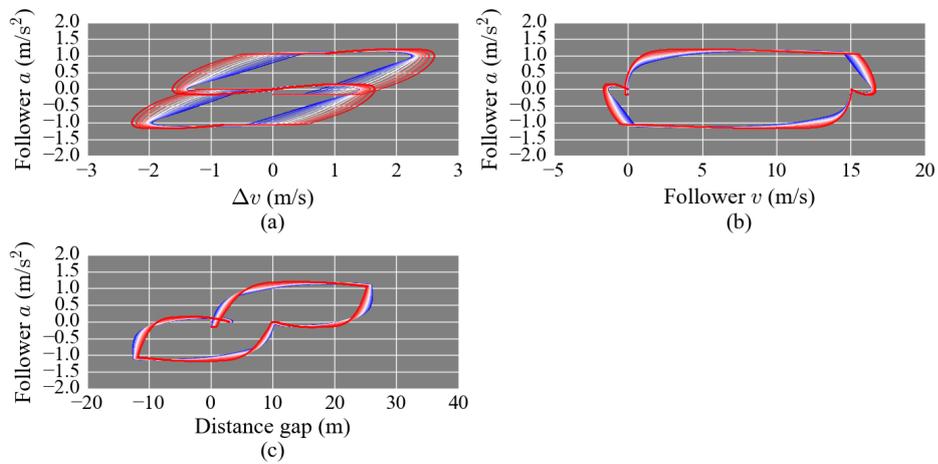
Figure 3.86: Response of the characteristic leaf to the variation of T_1.
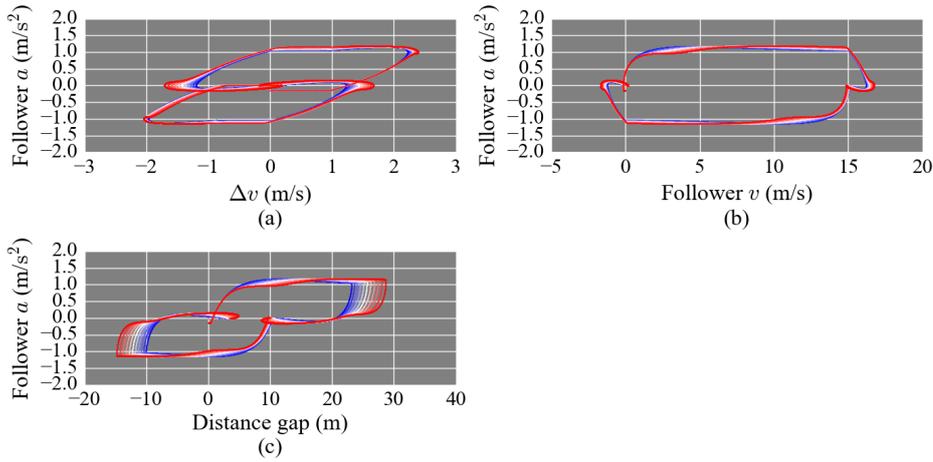
Figure 3.87: Response of the characteristic leaf to the variation of T_2.

Figure 3.88: Response of the characteristic leaf to the variation of $s_{\min}$.

Figure 3.89: Response of the characteristic leaf to the variation of $L^{\mathrm{f}}(L^{\mathrm{l}})$.
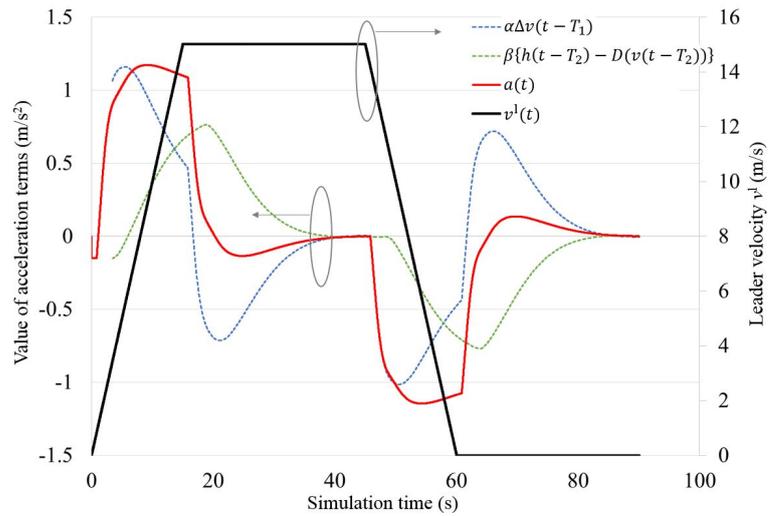
Figure 3.90: Evolution of the respective terms in the Helly model and the velocity of the leader.

3.3.2.5 Gazis-Herman-Rothery Model

Figure 3.91 to 3.94 indicate the response of the characteristic leaves to the variation of the parameters in GHR model. Note that we terminated the simulation when the follower had the negative velocity when they tried to stop in front of the leader. Because GHR model refers not to the distance gap but to the headway distance at their denominator, the distance gap can be negative value.

Except for the variation of $s_{\min}$ shown in Figure 3.93, we observed oscillation of the acceleration in the acceleration phase in some parameter range.
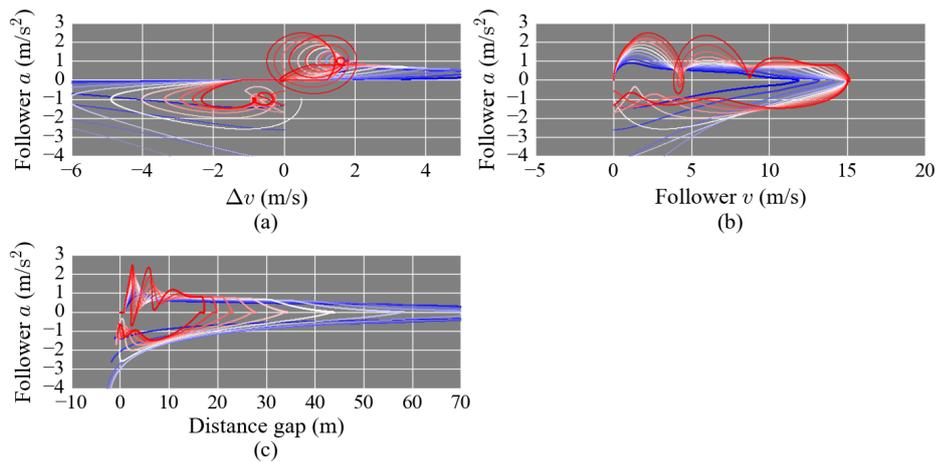
Figure 3.91: Response of the characteristic leaf to the variation of C.
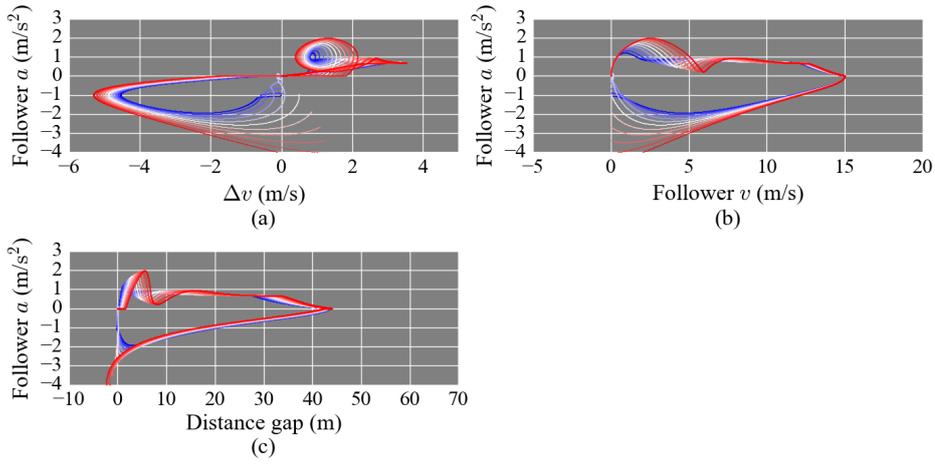
Figure 3.92: Response of the characteristic leaf to the variation of T.
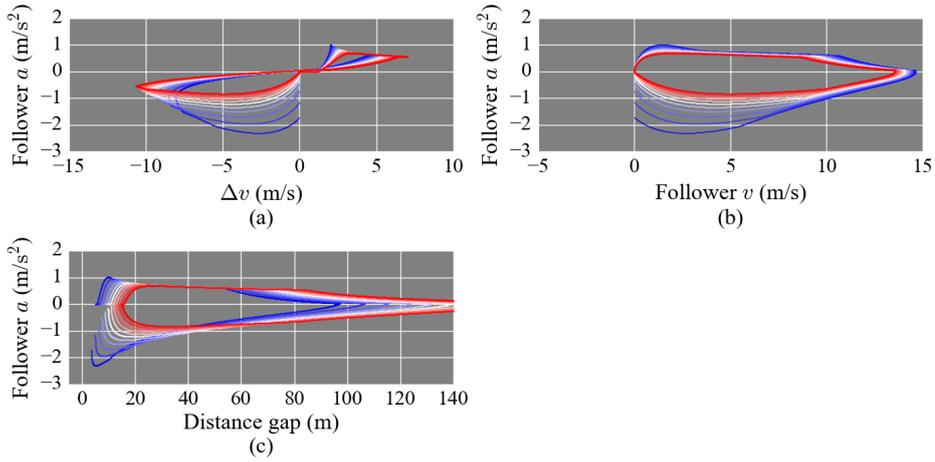
Figure 3.93: Response of the characteristic leaf to the variation of $s_{\min}$.
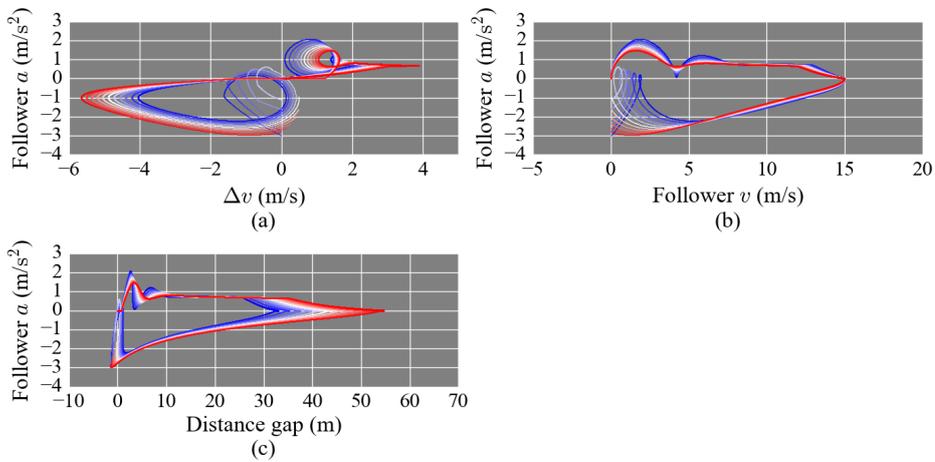
Figure 3.94: Response of the characteristic leaf to the variation of $L^{\mathrm{f}}(L^1)$.

3.4 Comparison of Observed and Simulated Characteristic Leaves

In this section, we compare the observed and simulated characteristic leaves from the perspective of each physical value, i.e., the distance gap, the velocity difference and followers' acceleration. We have already clarified the preferable index ranges of simulated series to be varied based on the positions of shapelets in Section 3.2.3. On the other hand, the fixed ranges and trends of the series variation have also been discussed in Section 3.2.4. With processing of simulated series by the DTW, we can compare the observed and simulated series on the same index k. In following discussion, we investigate the simulated series processed by the DTW. Furthermore, because we focus on the variation of simulated series with the parameter change, we compare the differences between simulated series of some sets of two parameters, the position of the shapelets, fixed ranges and the trends of the local maximum margin on the common index k.

Note that we applied the reference signal shown in Figure 3.14 shifted to 150 steps earlier for DTW process because the leaders start acceleration as soon as the simulation starts. After getting shifted index, we corrected the observed index, i.e., just added the initial state for 150 steps to the DTW series to make them comparable to the observed leaves.

3.4.1 Optimal Velocity Model

3.4.1.1 For variation of parameter V_{m}

In this section, we discuss on the replicability performance of the parameter V_{m} in the OV model. We show the logic diagrams of the shapelets, the trends of the local maximum marginm and differences between a certain set of two series calculated by different V_{m} in Figure 3.95. We call this type of figure as trend diagrams. Figure 3.95a shows the distance gap, and Figure 3.95b shows followers' acceleration. Although we simulated the parameter range of 50 % to 150 % of the center values, we picked the series differences between 90 % and 50 %, 120 % and 80 %, and 150 % and 110 % as representatives. These differences are depicted by a blue, green, and red line, respectively. For example, the blue line at index $k = 1000$ in Figure 3.95a represents the gap distance simulated by 90 % of the center value of V_{m} minus the one simulated by 50 % of the center V_{m} at the index $k = 1000$. The values of differences are indicated by the second axis of the figure. The local maximum margin is indicated by the black line in each figure and the first axis of the figure. Note that the values of the margin are multiplied to read easily. The manner of display of the logic diagram of the shapelet is the same to that of in Section 3.2.3.

From Figure 3.95a, we can find that the variation of V_{m} was able to replicate the variation of the distance gap at $k = 750$ and 1000 as the green and blue lines apparently have non-zero values on these indices. Furthermore, the trends of the local maximum margin and these series differences are similar, especially in the case of 120 % and 80 %, although their signs are opposite. Therefore, we

conclude that the parameter V_{m} in the specific range can replicate the variation of the distance gap depending on the type of following vehicle. The reason why V_{m} can replicate the variation of the distance gap is that it limits the maximum velocity of the follower. If we apply $V_{\mathrm{m}} < V_{\max}^{\mathrm{l}}$, the distance gap can be lengthen in the steady phase.

For Figure 3.95b, we need to check the fixed ranges. We could not find non-zero values in the series differences at the suggested fixed ranges, i.e., $k \leq 263$, $825 \leq k \leq 835$, and $1309 \leq k \leq 1341$. Then, we conducted the comparison between shapelet positions and series differences. Regarding the characteristic shapelet for the trucks at $k = 1000$ to 1100, the combinations shown by the green and blue lines seem to have the potential to replicate. On the other hand, the difference covering whole the acceleration phase is not observed, although inverted sign difference in the latter acceleration phase ($k = 600$) is observed. Parameter V_{m} will replicate only the acceleration peak of the trucks in the latter acceleration phase but will not replicate the large acceleration change of the cars.

From the discussion above, we concluded that the parameter V_{m} can partly replicate the variation of followers' acceleration. One problem is that there is a limited range of parameters. We need to select V_{m} less than the maximum velocity of the leader. In addition, we cannot replicate the large acceleration variation observed in the cases of cars.

3.4.1.2 For variation of parameter m

Figure 3.96 shows the trend diagrams for the variation of parameter m. From Figure 3.96a, we confirmed that the replicability around shapelets is guaranteed. However, the series differences reverse their signs at $k = 400$ and 1200. This means that we need to compensate for the inverse trends of the distance gap at the early acceleration phase and latter deceleration phase when we adjust the distance gap around the shapelets by m. Regarding acceleration, there seems to be no problem with the fixed ranges from Figure 3.96b. Focusing on the series differences, we can confirm that the rough trends are consistent with the local maximum margin and trends shown in Figure 3.57 but there is large inversion of the sign at $k = 1000$. We conclude that the parameter m has the ability to replicate rough trends of the acceleration and is partly appropriate for the replication of the distance gap. The distance gap in the early acceleration phase and the latter deceleration phase will have opposite tendencies. Besides, there will be some unnecessary features in the acceleration.
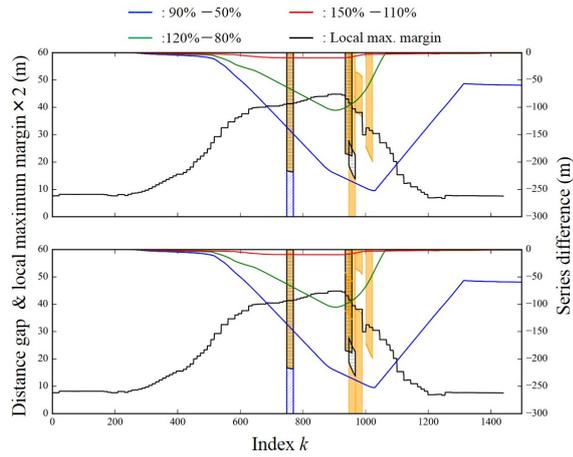
(a) Evolution on index k vs. the distance gap plane.
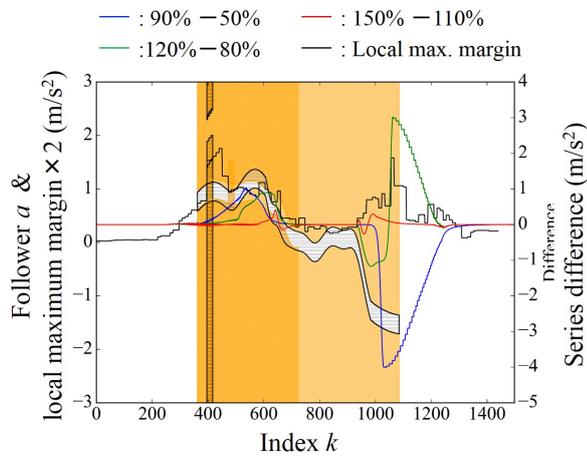
(b) Evolution on index k vs. the followers' accelera-
tion plane.

Figure 3.95: Evolution of the shapelets, the local maximum margin and difference
between simulated series with the variation of parameter V_{m}.
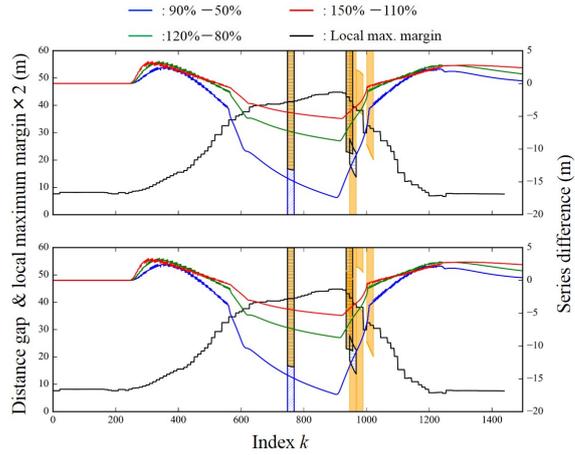
(a) Evolution on index k vs. the distance gap plane.
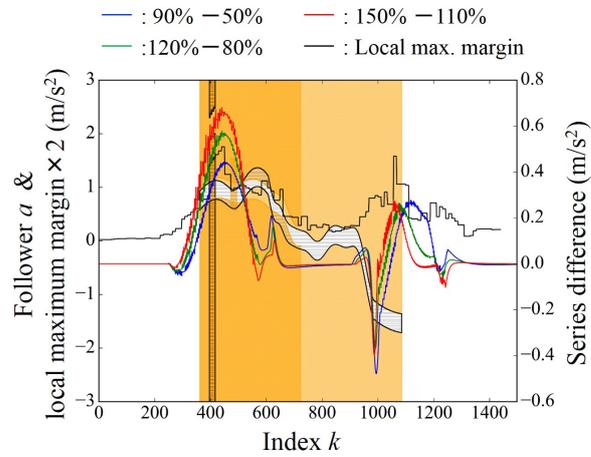
(b) Evolution on index k vs. the followers' acceleration plane.

Figure 3.96: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter m.

3.4.1.3 For variation of parameter $s_{\min}$

Figure 3.97 shows the trend diagrams for the variation of parameter $s_{\min}$. Parameter $s_{\min}$ is not appropriate to replicate the trend of the distance gap because the series differences in Figure 3.97a increase in the both the waiting and stopping phases, i.e., at $k = 0$ and 1499 than at the place of the shapelets. We should use $s_{\min}$ only to adjust the initial distance gap before the acceleration phase and after the stopping phase. In Figure 3.97b, we observed non-zero values of series differences at the fixed ranges of $1309 \leq k \leq 1341$. We can conclude that $s_{\min}$ causes unnecessary acceleration variation at the beginning of the stopping phase. Furthermore, because the sign alternation in the acceleration phase is too early, the acceleration peak will occur not in the latter acceleration phase but earlier phase.

3.4.1.4 For variation of parameter L^{f} or L^{l}

Figure 3.98 shows the trend diagrams for the variation of parameter L^{f} or L^{l}. We can conclude that L^{f} and L^{l} can replicate the trends of the distance gap from Figure 3.98a but there are some difference of increasing and decreasing timings. We observed the small variations of the acceleration differences at the fixed ranges in Figure 3.98b. Furthermore, the series differences similar to $k = 1050$ were observed at $k = 1250$. The sign alternation at $k = 300$ is also inconsistent with the required difference. Adjustment of L^{f} or L^{l} will cause unnecessary variations of the acceleration and the deceleration.

3.4.1.5 For variation of parameter Δ

Figure 3.99 shows the trend diagrams for the variation of parameter Δ. We can claim that Δ has the replication performance around the shapelets but the series differences maintain the non-zero values in both the waiting and stopping phases. Although the series differences becomes zero at the beginning and end of the series in Figure 3.66, the differences remain longer than the observed margin. The series differences for the acceleration shown in Figure 3.99b change their values at the fixed ranges at the beginning of the stopping phase. Except for it, the trends of the series difference follow the required difference when we choose $150 \% - 110 \%$, i.e., the red line. We conclude that Δ is somewhat proper parameter to replicate the trends of the acceleration.
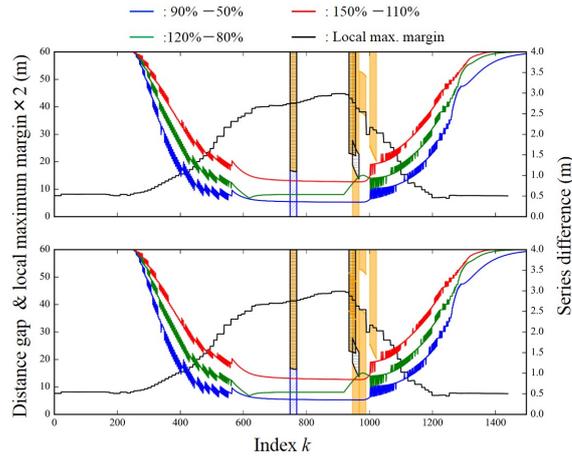
(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the followers' acceleration plane.

Figure 3.97: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter $s_{\min}$.

(a) Evolution on index k vs. the distance gap plane.
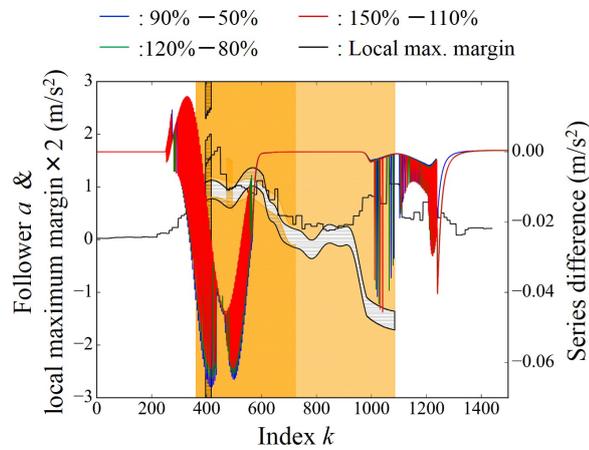
(b) Evolution on index k vs. the followers' acceleration plane.

Figure 3.98: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter L^{f} or L^{l}.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the followers' acceleration plane.

Figure 3.99: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter Δ.

3.4.1.6 For variation of parameter A

Figure 3.100 shows the trend diagrams for the variation of parameter A. From Figure 3.100a, parameter A seems to cause unnecessary variation of the distance gap at the acceleration phases and did not replicate necessary variation at $k = 750$. Indeed, there are chances to replicate the trend at $k = 1000$, but the trends of the series differences and the local maximum margin are totally different. Therefore, we conclude that A is not a proper parameter for replication of the distance gap. Regarding the acceleration shown in Figure 3.100b, the major signs in the acceleration phase and the deceleration phase are opposite; the series differences are not consistent with the required tendency. Therefore, we conclude that A is not a proper parameter to replicate the change of the distance gap and followers' acceleration accompanied by vehicle change.
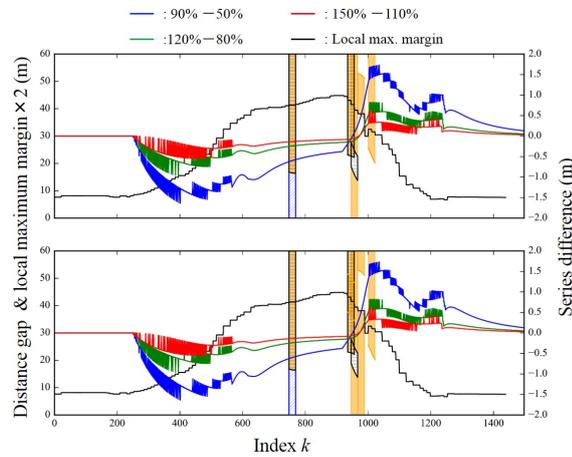
(a) Evolution on index k vs. the distance gap plane.
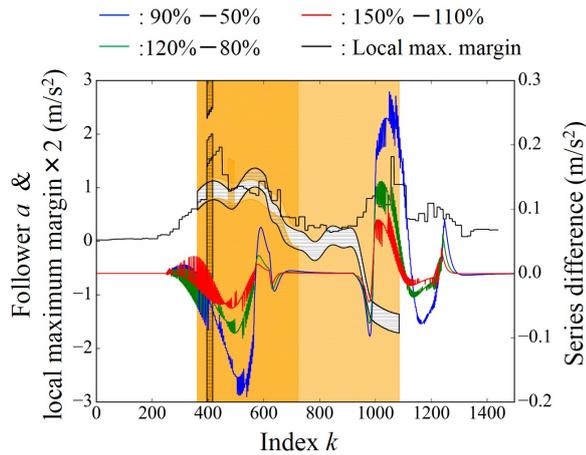
(b) Evolution on index k vs. the followers' acceleration plane.

Figure 3.100: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter A.

3.4.1.7 Summary: The replication performance of the OV model

In summary, we obtained the comparison results as listed in Table 3.9 and 3.10. By bold characters, we noted the evaluation of each parameter. The evaluation was conducted from four perspectives.

1. Did we need to select proper parameter range for the replication?: Yes or No.

2. Were all the observed features replicated?: All features, part of the features, or none of the features.

3. Were the replicated features located at the required index?: Yes or No, or they were shifted.

4. Were any unnecessary features caused?: Yes or No.

For the respective evaluation items, we prepared corresponding sub-sentences and connected them.

1. For all parameter (param.) range, or For some param. range,

2. fully replicable, partly replicable or no necessary variation

3. with (w/) or without (w/o) shift

4. with (w/) or without (w/o) unnecessary features (UF).

From Table 3.9 and 3.10, we found that there are four parameters, i.e., V_{m}, m, $L^{\mathrm{f}}(L^{\mathrm{l}})$, and Δ, which can replicate the features of the distance gap. Three parameters replicating the features of followers' acceleration are V_{m}, m, and Δ. Note that each parameter replicates the features with at least one of the shift and the unnecessary features or has limitation on parameter selection. That is, we need to conclude that the OV model can not perfectly replicate the features of the characteristic leaves of the respective vehicles. We added problems to be caused with the adjustment of each parameter in Table 3.9 and 3.10. If we try to replicate the features of each physical value using different parameters, L^{f} and L^{l} can be used only for the distance gap. One can arbitrarily choose the dedicated parameters from V_{m}, m, Δ.

Table 3.9: Replicablity of respective parameters in the OV model for the distance gap of various following vehicles.

	Distance gap
V_{m}	**For some param. range, fully replicable w/o shift w/o UF.**
m	Replicable around shapelets. Inversed trends observed at $k = 400, 1200$ (EA & LD) **For all param. range, fully replicable w/o shift w/ UF.**
s_{min}	Large difference occur at W & SP. **For all param. range, no necessary variation, w/UF.**
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**For all param. range, fully replicable w/ shift w/o UF.**
Δ	Non-zero values remain at W & SP. **For all param. range, fully replicable w/ shift w/ UF.**
A	Unnecessary variation at A. Not replicate variation at k=750. Trend totally different. **For all param. range, no necessary variation, w/ UF.**

Table 3.10: Replicablity of respective parameters in the OV model for followers' acceleration of various following vehicles.

	Followers' acceleration
V_{m}	**For some param. range, partly replicable w/o shift w/o UF.** Need to choose less than $V_{\mathrm{max}}^{\mathrm{l}}$. Unable to replicate large acc change of cars. Only replicate the acc. peak of truck in LA.
m	Sign alternation observeded in ED. **For all param. range, partly replicable w/ shift w/UF.**
s_{min}	Non-zero value at fixed range. Unnecessary acc. at beginning of SP. Peak position shifted. **For all param. range, no neccesary variation, w/ UF.**
$L^{\mathrm{f}}(L^{\mathrm{l}})$	Small variation at fixed ranges. Unnecessary features in A and D. **For all param. range, no necessary variation, w/ UF.**
Δ	Variation at fixed range at SP. Rough trend replicated. **For some param. range, partly replicable w/o shift w/ UF.**
A	Major trends are opposite in acc. and dec. **For all param. range, no necessary variation, w/ UF.**

3.4.2 Full Velocity Difference Model

3.4.2.1 For variation of parameter V_{m}

Figure 3.101 shows the trend diagrams for the variation of parameter V_{m}. In Figure 3.101a, we confirm that the variations of the distance gap are roughly located where the shapelets are, as well as that the trends of the series differences follow that of the local maximum margin in the cases of the blue and green lines. However, in the case of the blue line, the series difference in the stopping phase has non-zero values. We need to choose a certain range of V_{m} for replication of the variation of the distance gap. Regarding the velocity difference shown in Figure 3.101b, the green and blue lines have non-zero values at the fixed ranges in the stopping phase and steady phase. If we choose the green line, the large velocity difference of the trucks at $k = 1100$ will be replicated. Parameter V_{m} in a certain range can replicate part of the features with unnecessary fluctuations at the fixed ranges. Regarding the acceleration, the trends of the series differences are different from the required ones, although the fluctuations are almost zero at the fixed ranges. We conclude that V_{m} cannot replicate the variations in followers' acceleration.

3.4.2.2 For variation of parameter Δs

Figure 3.102 shows the trend diagrams for the variation of parameter Δs. From Figure 3.102a, Δs seems to replicate the variation of the distance gap at the shapelet positions. In addition, the trends of the series differences follow those of the local maximum margin. We conclude that Δs is an appropriate parameter for the replication of the trends of the distance gap. When we focus on the velocity difference shown in Figure 3.102b, the series differences seem to be varied at the index shifted from where the shapelets are. Furthermore, the series differences have non-zero values at the fixed range in the stopping phase. If we adjust Δs to fit the velocity difference, we need to ignore these shifts and fluctuations. Regarding followers' acceleration shown in Figure 3.102c, we observed the totally different trends of the series differences from the local maximum margin and shapelets. The variation of Δs will cause a problem in the evolution of followers' acceleration.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.101: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter V_{m}.
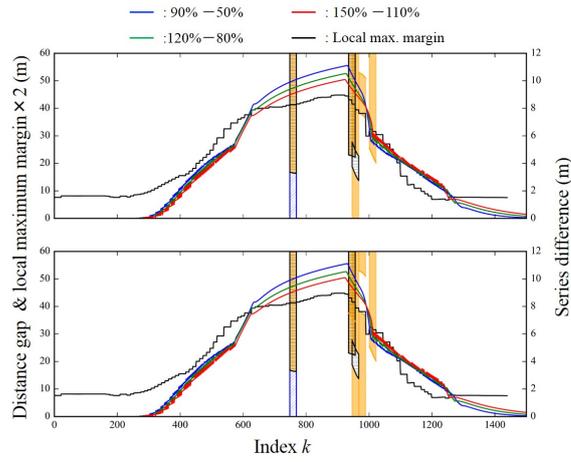
(a) Evolution on index k vs. the distance gap plane.
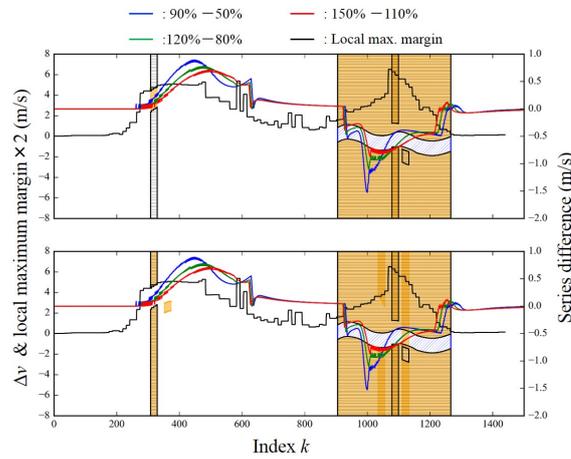
(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.102: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter Δs.

3.4.2.3 For variation of parameter β

Figure 3.103 shows the trend diagrams for the variation of parameter β. From Figure 3.103a, β seems to replicate the variation of the distance gap at the shapelets. However, unnecessary variation of the distance gap will also occur at the beginning of the acceleration phase and the end of the deceleration phase. In Figure 3.103b, we observed a shifted peak of the series differences at $k = 450$ and 1230. In addition, we also observed small perturbations at $k = 1300$ where the fixed range is. Although β replicates the rough trends of the observed series on the k vs. Δv plane, there are some different points which we need to be aware of, i.e., there are shifted peaks and unnecessary variaton at the fixed range in stopping phase. For followers' acceleration shown in Figure 3.103c, we observed the large peak of the series differences at the fixed ranges at $k = 1300$. Furthermore, we also observed the large oscillation at around $k = 400$ which is not insistent to the trend of the local maximum margin. Although parameter β has ability to replicate followers' acceleration difference between the vehicle types with the timing shift, it seriously causes unnecessary features.

3.4.2.4 For variation of parameter $s_{\min}$

Figure 3.104 shows the trend diagrams for the variation of parameter $s_{\min}$. In any of the figures in Figure 3.104, we could not find any capability of $s_{\min}$ to replicate the variation trends of the distance gap, the velocity difference and followers' acceleration. It can be said that $s_{\min}$ affects only the initial behavior of the vehicle from Figure 3.71.

3.4.2.5 For variation of parameter L^{f} or L^{l}

Figure 3.105 shows the trend diagrams for the variation of parameter L^{f} or L^{l}. Note that the magnitudes of the secondary axes are in the order of 10^{-12} to 10^{-10}, which we can not recognize on the characteristic leaves. It is obvious that L^{f} and L^{l} do not affect the behavior of the vehicle in the FVD model. This phenomenon can also be confirmed in Figure 3.72.

3.4.2.6 For variation of parameter τ

Figure 3.106 shows the trend diagrams for the variation of parameter τ. Regarding the distance gap shown in Figure 3.106a, the series differences alter their signs several times. In addition, the signs of the series differences where the shapelets exist are inconsistent with the required ones. The trends of the series differences are not consistent with those of the local maximum margin and the shapelets. The adjustment of τ will have a negative effect on the desired variation of the distance gap. In the case of the velocity difference in Figure 3.106b, we cannot conclude that the series differences are consistent with the trend diagram. These contain the alternation of their signs in the acceleration and deceleration phases. For the followers' acceleration shown in Figure 3.106c, τ seems not to have capability for replication of the variation in followers' acceleration similar to the distance gap and the velocity difference. The series differences contain unnecessary sign alternations and peaks.
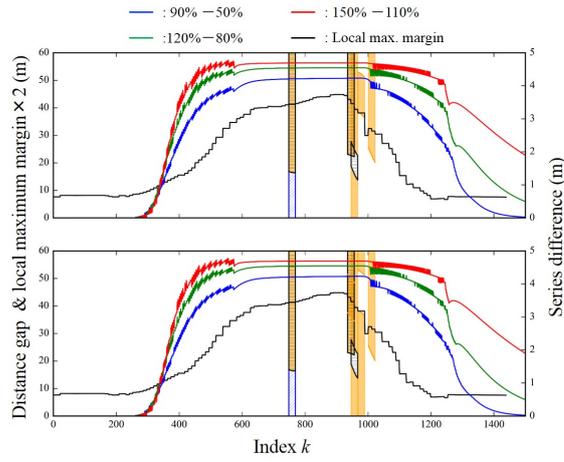
(a) Evolution on index k vs. the distance gap plane.
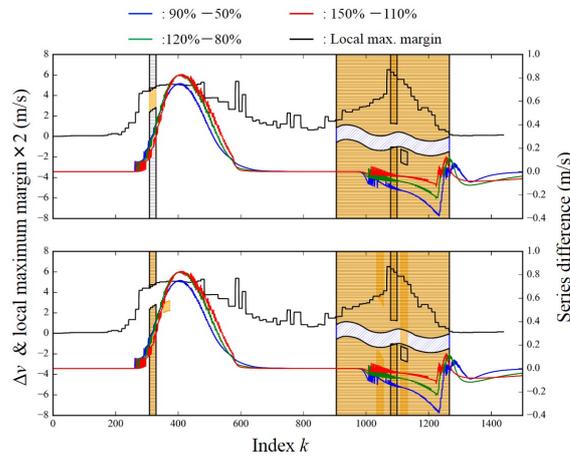
(b) Evolution on index k vs. the velocity difference plane.
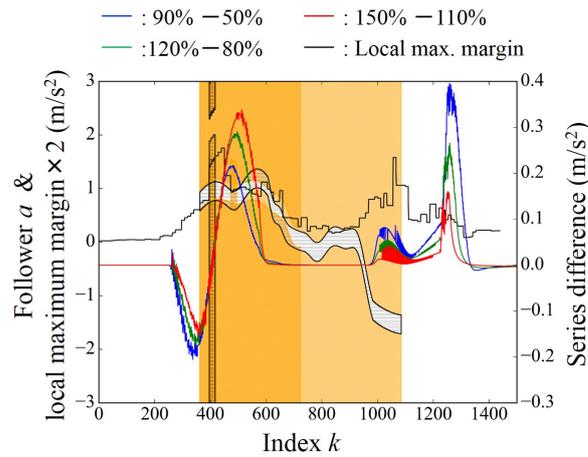
(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.103: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter β.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.104: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter $s_{\min}$.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.105: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter L^{f} or L^{l}.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.106: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter τ.

3.4.2.7 For variation of parameter γ

Figure 3.107 shows the trend diagrams for the variation of parameter γ. The trends of the series differences in any trend diagrams are similar to those for parameter τ shown in Figure 3.106. Same as τ, γ has no capability to replicate the difference in the distance gap and the velocity difference caused by the vehicle types, while it can realize only the large deceleration of the trucks. Regarding the acceleration, similar to τ, we will have multiple alternations of the trends if we try to replicate the variation in followers' acceleration with γ.

3.4.2.8 Summary: The replication performance of the FVD model

We summarize the results of the comparison from Table 3.11 to 3.13. Let us consider the parameter selection for the fitting to keep the independence of physical values replicated by the selected parameters. In the other words, we would like to assign specific parameters for the replication of the features in a specific physical value. For the distance gap, it would be preferable to utilize Δs because it replicates the features of the distance gap without the shift or unnecessary features. In the case of followers' acceleration, we have no choice other than β. However, we need to be aware that β will affect the velocity difference simultaneously. For the velocity difference, we need to use V_{m} to keep the fitting independence of each physical value.

Table 3.11: Replicablity of respective parameters in the FVD model for the distance gap of various following vehicles.

	Distance gap
V_{m}	**For some param range, fully replicable w/o shift w/o UF.** Need to choose a certain range of V_{m}
Δs	**For all param. range, fully replicable w/o shift w/o UF.**
β	**For all param range, fully replicable w/ shift w/ UF.** Unnecessary variation at beginning of A and end of SP.
s_{min}	**For all param. range, no necessary variation w/ UF.** Affect only W and EA. Different trend.
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**No effect.**
τ	Sign alternation at shapelet positions. Different trend. **For all param. range, no necessary variation w/ UF.**
γ	**For all param. range, no necessary variation w/ UF** similar to τ.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.107: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter γ.

Table 3.12: Replicablity of respective parameters in the FVD model for the velocity difference of various following vehicles.

	Velocity difference
V_m	Fluctuation at fixed range & different trend to margin are observed. **For some param. range, partly replicable w/o shift w/ UF.**
Δs	Replicable to trend with shift. Variation at SP. **For all param. range, fully replicable w/ shift w/ UF.**
β	Unnecessary variation at fixed range in SP. Shifted peaks. **For all param. range, fully replicable w/ shift w/ UF.**
s_min	**For all param. range, no necessary variation w/ UF.** Affect only W and EA. Different trend.
$L^\mathrm{f}(L^\mathrm{l})$	**No effect.**
τ	Sign alternation in A and D. **For all param. range, no necessary variation w/ UF.**
γ	**For all param. range, no necessary variation w/ UF** similar to τ.

Table 3.13: Replicablity of respective parameters in the FVD model for followers' acceleration of various following vehicles.

	Followers' acceleration
V_m	Totally different trend. **For all param range, no necessary variation w/ UF.**
Δs	Totally different trend. **For all param. range, no necessary variation, w/ UF.**
β	Fluctuation at fixed ranges. Serious large peak at around $k = 400$ and 1300 (EA & SP). **For all param. range, fully replicable w/ shift w/ UF.**
s_min	**For all param. range, no necessary variation w/ UF.** Affect only W and EA. Different trend.
$L^\mathrm{f}(L^\mathrm{l})$	**No effect.**
τ	Unnecessary sign alternations and peaks in acc. or dec. **For all param. range, partly replicable w/o shift w/ UF.**
γ	Unnecessary feature in acc. or dec. **For all param. range, no necessary variation w/ UF** similar to τ.

3.4.3 Intelligent Driver Model

As we mentioned in Section 3.3.2.3, some simulations diverged at the end of the simulation. In the following comparisons, we utilized only the data before the divergence.

3.4.3.1 For variation of parameter V_{m}

Figure 3.108 shows the trend diagrams for the variation of parameter V_{m}. As shown in Figure 3.108a, V_{m} seems to have sufficient ability to replicate the difference in the distance gap around the shapelets and follow the trends of the local maximum margin. However, we need to note that it is required to choose V_{m} less than the maximum velocity of the leader. Regarding the velocity difference shown in Figure 3.108b, V_{m} causes unnecessary deviation at the fixed ranges of $748 \leq k \leq 769$, i.e., in the steady phase, and cannot replicate the variation in the early acceleration phase. For the followers' acceleration, if we choose the proper combinations of the parameters, i.e., $120\% - 80\%$, V_{m} can replicate the large deceleration of the trucks around the shapelet. However, the large acceleration of the cars cannot be replicated. Besides, it again requires the selection of the smaller V_{m} than the maximum velocity of the leader.

3.4.3.2 For variation of parameter δ

Figure 3.109 shows the trend diagrams for the variation of parameter δ. From Figure 3.109a, we can conclude that δ replicate the variation of the distance gap accompanied by the change in vehicle type, although series difference decreases earlier than the local maximum margin. When we focus on the trend diagram for the velocity difference in Figure 3.109b, we find that the variation in the deceleration phase can be replicated by δ. However, in the range $400 \leq k \leq 800$, we observe unnecessary variation in the velocity difference. In the trend diagram for the acceleration shown in Figure 3.109c, we observed unnecessary variation at $k = 1200$. Although the major signs of the series differences are the same in the acceleration phase and the deceleration phase, we confirm shifts of the series differences from the local maximum margin.

3.4.3.3 For variation of parameter T

Figure 3.110 shows the trend diagrams for the variation of parameter T. Parameter T also has the ability to replicate the variation of the distance gap as shown in Figure 3.110a. In Figure 3.110b, we can confirm that none-zero values of series differences still remain at the fixed range of $1254 \leq k$ and the trend of the series difference varies from that of the local maximum margin in the acceleration phase. Although the basic trends in the acceleration phase and deceleration phase are consistent with the experiments, the timings of features are shifted. From Figure 3.110c, we can observe the totally different trend of the series difference, especially at the beginning of the stopping phase.
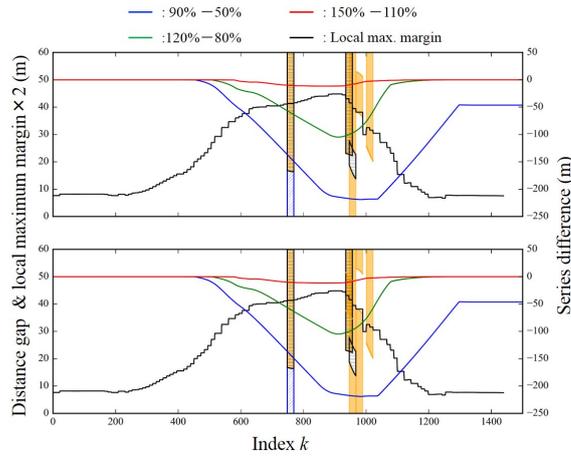
(a) Evolution on index k vs. the distance gap plane.
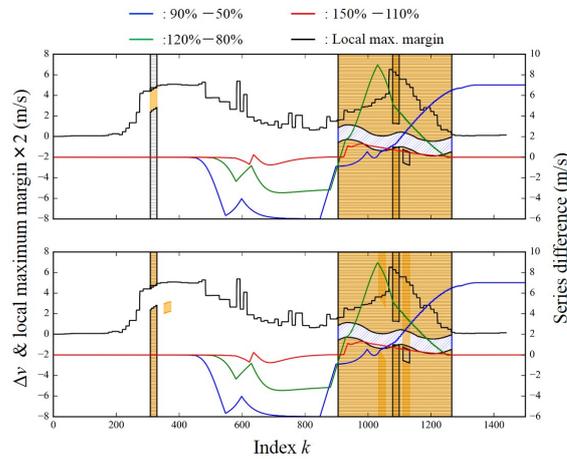
(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.108: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter V_{m}.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.109: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter δ.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.110: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter T.

3.4.3.4   For variation of parameter $s_{\min}$

Figure 3.111 shows the trend diagrams for the variation of parameter $s_{\min}$. Regarding the distance gap, $s_{\min}$ achieved the significant difference at the steady phase compared to the waiting and stopping phases in Figure 3.111a, although the decreasing timing is early. For the velocity difference shown in Figure 3.111b, although small perturbation occurred at the fixed ranges in the stopping phases, rough trends of the local maximum margin were captured by the series differences. As long as we understand, the largest difference in the acceleration phase occurs not in the early acceleration phase but latter acceleration phase. On the other hand, when confirming the trend diagram for the followers' acceleration, we face the series differences totally different from trends observed in the experiment. In particular, it significantly affects the acceleration from the latter deceleration phase.

3.4.3.5   For variation of parameter L^{f} or L^{l}

Figure 3.112 shows the trend diagrams for the variation of parameter L^{f} or L^{l}. As shown in Figure 3.79, the parameters L^{f} and L^{l} do not affect the characteristic leaves. This is because the ID model refers to the distance gap from which the effects of the vehicle length are removed. Although the series differences in Figure 3.112 seem to have non-zero values, their values are small.

3.4.3.6   For variation of parameter A

Figure 3.113 shows the trend diagrams for the variation of parameter A. From Figure 3.113a, we can confirm that A replicates the characteristic variation of the distance gap with the small shift of the variation timing. Regarding the velocity difference, although there is small perturbation at the fixed ranges at $k = 750$ and the peak timings of the series differences are shifted from those of the local maximum margin, it is implied that A can replicate the rough features observed in the velocity difference. On the other hand, for the followers' acceleration shown in Figure 3.113c, respective series differences showed different trends. In the cases of the green and the red lines, their trends follow the required ones including the sign inversion in the latter acceleration phase, which corresponds to the acceleration peak of the trucks. Note that the magnitude of the series differences are larger in the latter deceleration phase than the magnitude in the early deceleration phase. Although we need to be aware of some shifts and the parameter range, parameter A has the ability to replicate the differences.
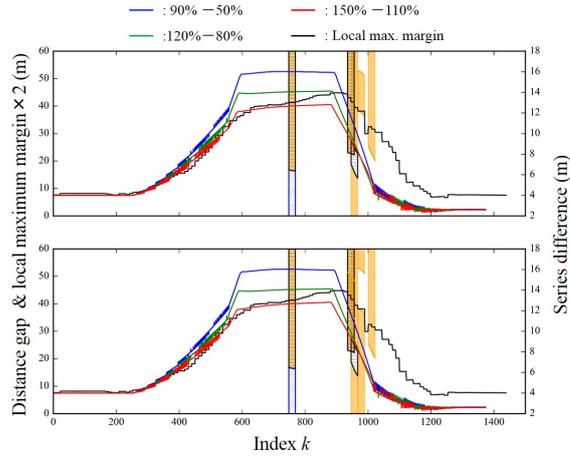
(a) Evolution on index k vs. the distance gap plane.
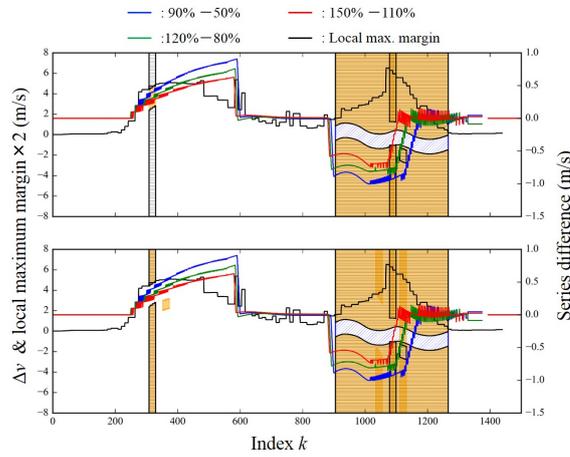
(b) Evolution on index k vs. the velocity difference plane.
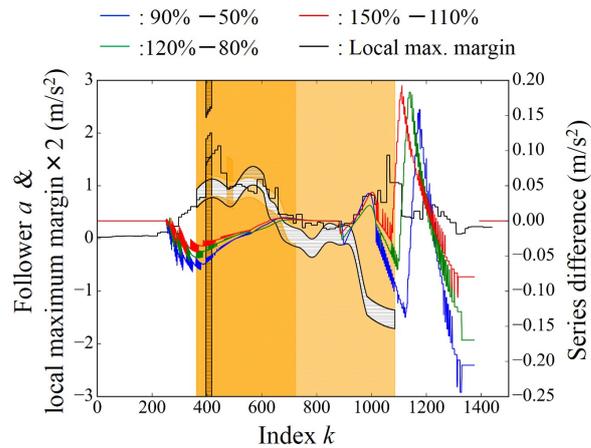
(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.111: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter $s_{\min}$.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.112: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter L^{f} or L^{l}.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.113: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter A.

3.4.3.7 For variation of parameter B

Figure 3.114 shows the trend diagrams for the variation of parameter B. Regarding the distance gap in Figure 3.114a, we observed turnover of the sign in each series difference at $k = 1000$. For the velocity difference shown in Figure 3.114b, although rough trends of the shapelets and the local maximum margin are captured, we observe timing differences of the peaks, sign inversions in the latter deceleration phase, and perturbations at the fixed range. Because the parameter B is related to the comfortable deceleration magnitude as mentioned in Section 3.1.2, we can observe the variations mainly in the deceleration phase. However, the difference peaks were observed at $1200 \leq k \leq 1300$ where the local maximum margin decreases its value from the early deceleration phase. Although the deceleration difference in the early deceleration phase can be replicated, we need to be patient with the variation in the latter deceleration phase when we adjust the parameter B to control the deceleration magnitude.

3.4.3.8 Summary: The replication performance of the ID model

From Table 3.14 to 3.16, we summarized the results of the comparison for each parameter in the ID model. From Table 3.14, most of the parameters except for L^{f} and L^{l} have the ability to replicate all the features of the distance gap. In particular, T seems to be the best parameter to replicate the features in the distance gap as it does not cause any shifts or unnecessary features. Regarding the followers' acceleration in Table 3.16, although there is a limitation for the range, it would be a choice to adjust V_{m} for the replication of the features in followers' acceleration. This is because we cannot keep the independence between physical values if we choose δ, A, or B, which affect the other physical values. Regarding the velocity difference in Table 3.15, it would be preferable to choose s_{min} from the perspective of the independence. We also need to be aware that all the features in the velocity difference cannot be replicated by any parameters in the ID model.

Table 3.14: Replicablity of respective parameters in the ID model for the distance gap of various following vehicles.

	Distance gap
V_{m}	**For some param. range, fully replicable w/o shift w/o UF.** V_{m} less than $V_{\mathrm{max}}^{\mathrm{l}}$ is required.
δ	**For all param. range, fully replicable w/ shift w/o UF.**
T	**For all param. range, fully replicable w/o shift w/o UF.**
s_{min}	**For all param. range, fully replicable w/ shift w/o UF.**
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**No effect.**
A	**For all param. range, fully replicable w/ shift w/o UF.**
B	Sign alterntion at $k = 1000$ (ED). **For all param. range, fully replicable w/ shift w/ UF.**
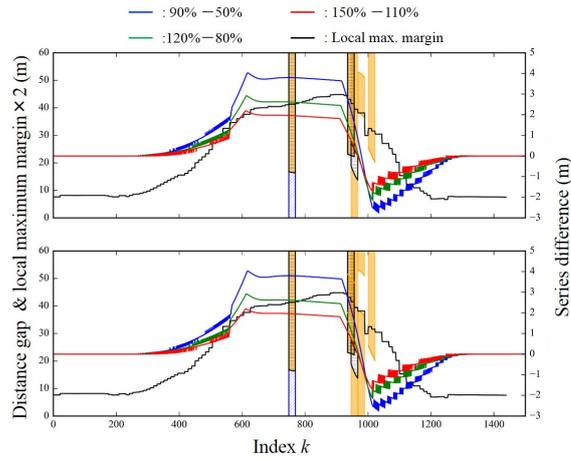
(a) Evolution on index k vs. the distance gap plane.
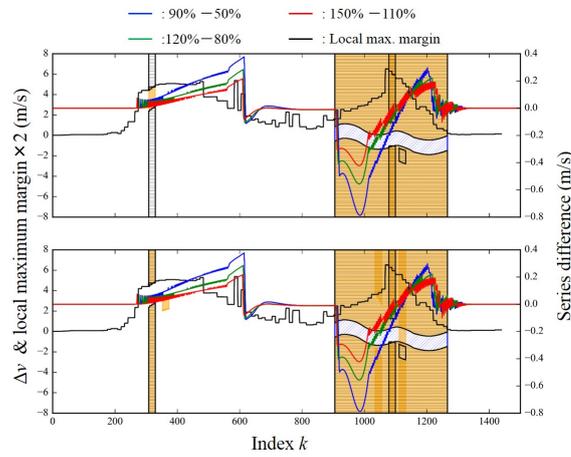
(b) Evolution on index k vs. the velocity difference plane.
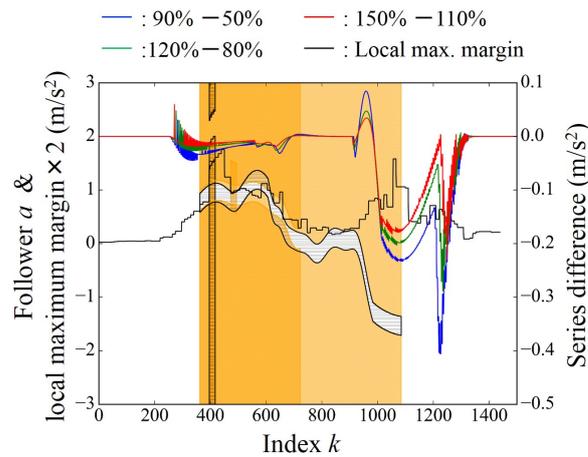
(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.114: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter B.

Table 3.15: Replicablity of respective parameters in the ID model for the velocity difference of various following vehicles.

	Velocity difference
V_{m}	Unnecessary variation at fixed range at ST. Unnable to replicate variation in EA. **For all param. range, no necessary variation w/ UF.**
δ	Variation in D replicated. Unnecessary variation in 400-800 (A & ST). **For all param. range, partly replicable w/o shift w/ UF.**
T	Difference remains at fixed range $1254 \leq k$ (SP). Different trend in A. Basic trend is consistent but feature timing shifted. **For all param. range, partly replicable w/ shift w/ UF.**
$s_{\min}$	Small perturbation in fixed range at SP. Trend roughly replicated. Acc. difference appears not in EA but LA. **For all param. range, partly replicable w/ shift w/ UF.**
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**No effect.**
A	Small perturbation at fixed range $k = 750$ (ST). Peak timing shifted. Trend roughly replicated. **For some range of param, partly replicable w/ shift w/o UF.**
B	Trend roughly replicated. Small perturbation at fixed range. Sign inverstion in LD. Timing shift of peaks. **For all param. range, partly replicable w/ shift w/ UF.**

Table 3.16: Replicablity of respective parameters in the ID model for followers' acceleration of various following vehicles.

	Followers' acceleration
V_{m}	With proper Vm, shapelet of truck replicated in dec. V_{m} less than $V^{\mathrm{l}}_{\max}$ required. Unnable to replicate large acc. of cars. **For some param. range, partly replicable w/o shift w/ UF.**
δ	Unnecessary variation at $k = 1200$ (LD). Rough trends replicated but shifted. **For all param. range, fully replicable w/ shift w/ UF.**
T	Totally different especially at beginning of SP. **For all param. range, no necessary variation w/ UF.**
$s_{\min}$	**For all param. range, no necessary variation w/ UF**
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**No effect.**
A	**For some param. range, fully replicable w/ shift w/ UF.**
B	Dec. difference in ED replicated. Remaining variation in LD. **For all param. range, partly replicable w/o shift w/ UF.**

3.4.4   Gazis-Herman-Rothery Model

In this section, we discuss the replicability of the GHR model to observed features of the characteristic leaves. Before we proceed with the comparison, we note that the series after the velocity became less than zero were removed. In addition, the common index $k = 0$ was not assigned to the first index of the simulated series, but assigned to the index corresponding to the simulation time $t > 0$. However, the simulation time assigned $k = 0$ was before the follower started to move, i.e., almost the initial state. Therefore, we decided to proceed with the comparison using these DTW series.

3.4.4.1   For variation of parameter C

Figure 3.115 shows the trend diagrams for the variation of parameter C. As long as we focus on the series differences before the velocity becomes less than zero, parameter C seems to have the ability to replicate the variation around the shapelets in Figure 3.115a. The trends of the series differences are also consistent with those of the local maximum margin.

As for the velocity difference shown in Figure 3.115b, the series differences become almost zero except for the combination $90\% - 50\%$. However, although the rough trends of the series differences follow those of the local maximum margin, the peaks of the variation are observed not in the early acceleration phase and the latter deceleration phase but in the latter acceleration phase and the early deceleration phase. We need to be aware of these shifts when we fit the velocity difference with parameter C.

Regarding the followers' acceleration shown in Figure 3.115c, although the fixed ranges in the waiting phase and the steady phase are replicated, the series differences are varied at fixed range in the stopping phase, which is directed to the negative velocity. Furthermore, we observe several alternations of the signs in the acceleration phase. The major signs in the acceleration phase and the deceleration phase cannot be compared. We conclude that parameter C does not replicate the observed features and causes unnecessary variations in the followers' acceleration.

3.4.4.2   For variation of parameter T

Figure 3.116 shows the trend diagrams for the variation of parameter T. The trends of the series differences observed in Figure 3.116 are totally different from those of the local maximum margin. Although the time delay T seems to be the crucial parameter replicating response time of respective vehicles, we need to conclude that the variation of T does not replicate the vehicle differences.

(a) Evolution on index k vs. the distance gap plane.

(b) Evolution on index k vs. the velocity difference plane.
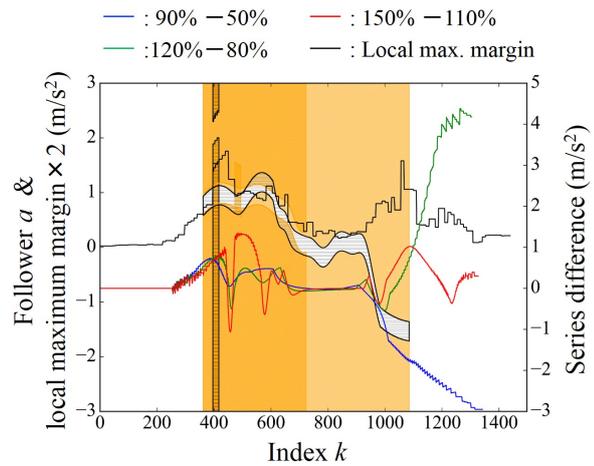
(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.115: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter C.
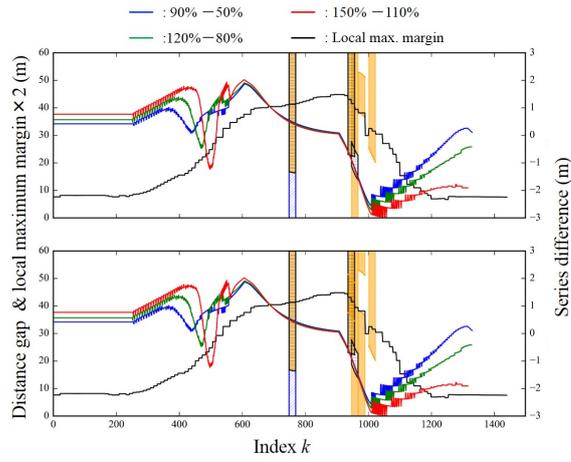
(a) Evolution on index k vs. the distance gap plane.
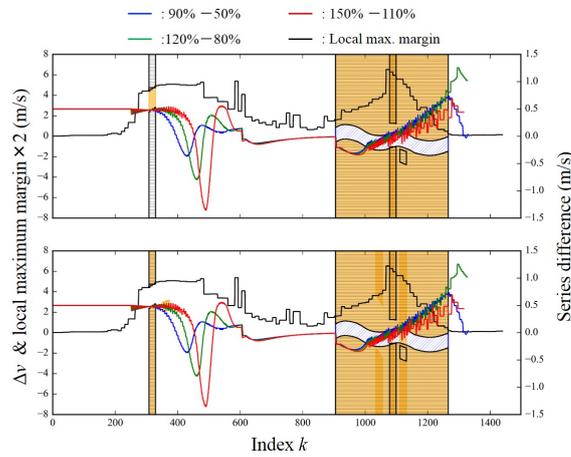
(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.116: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter T.

3.4.4.3   For variation of parameter $s_{\min}$

Figure 3.117 shows the trend diagrams for the variation of parameter $s_{\min}$. Regarding the distance gap shown in Figure 3.117a, parameter $s_{\min}$ seems to replicate the trends of the distance gap similar to parameter C.

As for the velocity difference, the series differences decrease where the fixed ranges are. However, although the trends of the series differences are roughly consistent with the local maximum margin, the positions of the peaks of the series differences are different from where the shapelets exist. We conclude that parameter $s_{\min}$ can replicate the rough trends of the variation but its performance for the replication is not high.

Regarding the followers' acceleration shown in Figure 3.117c, even when we ignore the acceleration directed to the negative velocity after $k = 1000$, the variation trends of the series difference are totally different from those of the local maximum margin. We need to conclude that parameter $s_{\min}$ does not have any ability to replicate the variation in followers' acceleration when the following vehicles are changed.

3.4.4.4   For variation of parameter L^{f} or L^{l}

Figure 3.118 shows the trend diagrams for the variation of parameter L^{f} or L^{l}. From Figure 3.118a, we can conclude that L^{f} and L^{l} have the ability for replication of the distance gap in the cases of various vehicles. The series differences capture the variations at where the shapelets are and the trends of the local maximum margin.

Regarding the velocity difference shown in Figure 3.118b, we observed the large oscillations at the fixed ranges in the stopping phase ($1254 \le k$), although we could see the variation in the range $900 \le k \le 1200$ where the shapelets were observed. In addition, the sign alternations in the acceleration phase $300 \le k \le 700$ and few variation in $300 \le k \le 400$ were observed. The variation in the velocity difference caused by the vehicle change cannot be replicated properly by parameter L^{f} and L^{l}.

Regarding the followers' acceleration, we cannot observe significant trends of the series differences except for the acceleration directed to the negative velocity after $k = 1100$. Parameter L^{f} and L^{l} cannot replicate the variation of the followers' acceleration.
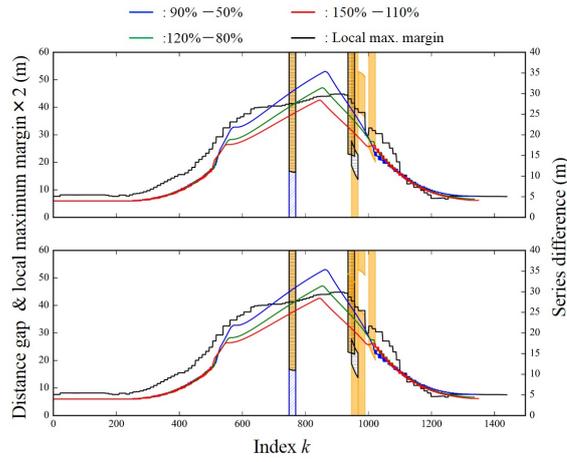
(a) Evolution on index k vs. the distance gap plane.
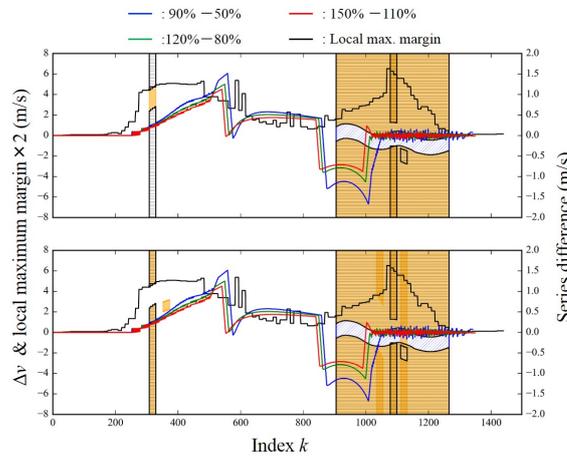
(b) Evolution on index k vs. the velocity difference plane.
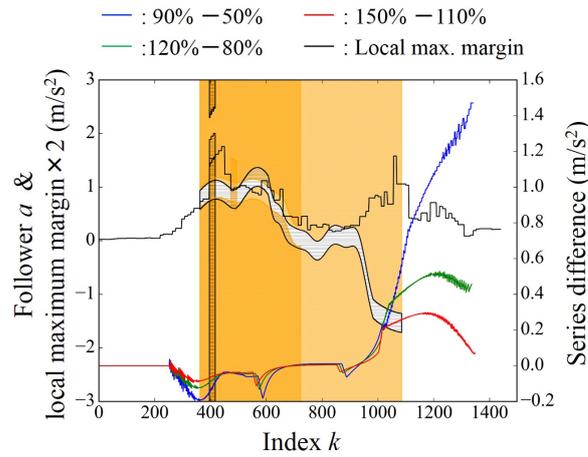
(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.117: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter $s_{\min}$.
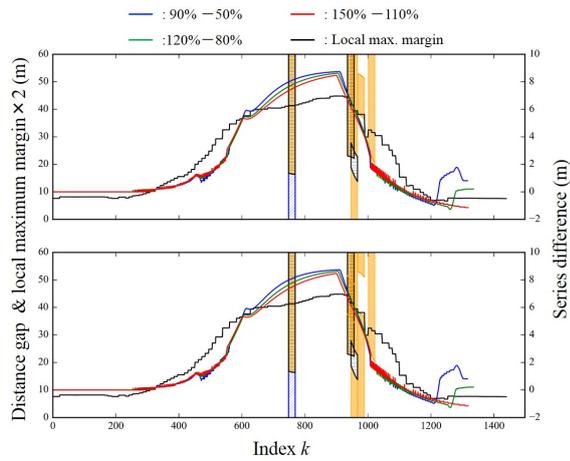
(a) Evolution on index k vs. the distance gap plane.
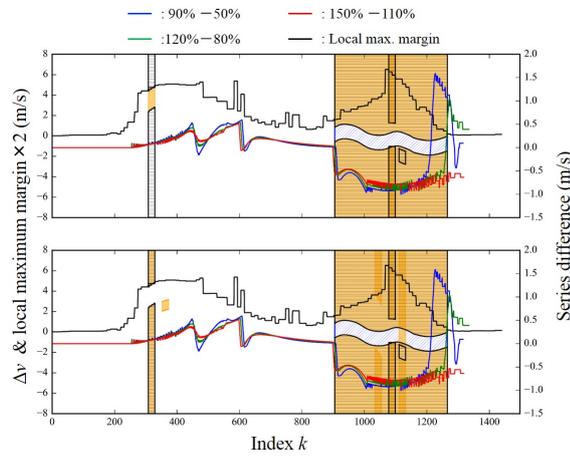
(b) Evolution on index k vs. the velocity difference plane.

(c) Evolution on index k vs. the followers' acceleration plane.

Figure 3.118: Evolution of the shapelets, the local maximum margin and difference between simulated series with the variation of parameter L^{f} or L^{l}.

3.4.4.5 Summary: The replication performance of the GHR model

From Table 3.17 to 3.19, we summarized the results of the comparison for each physical value and parameter. For the distance gap in Table 3.17, every parameters except for T have the ability to replicate the features without any shift and the unnecessary features. Regarding the velocity difference, these parameters, i.e., C, s_{min}, L^f and L^l have the ability to replicate some of the features. However, none of the parameters in the GHR model has capability for the replication of the variation in followers' acceleration. We conclude that the GHR model, in particular the model presented by Herman *et al.* [69] cannot be utilized for the analysis of mixed traffic including microscopic investigation from the acceleration perspective.

Table 3.17: Replicablity of respective parameters in the GHR model for the distance gap of various following vehicles.

	Distance gap
C	**For all param. range, fully replicable w/o shift w/o UF.**
T	**For all param. range, no necessary variation w/ UF.**
s_{min}	**For all param. range, fully replicable w/o shift w/o UF.**
$L^f(L^l)$	**For all param. range, fully replicable w/o shift w/o UF.**

Table 3.18: Replicablity of respective parameters in the GHR model for the velocity difference of various following vehicles.

	Velocity difference
C	**For some param. range, partly replicable w/ shift w/ UF.** Combination of 90-50% is not applicable. Peak position different.
T	**For all param. range, no necessary variation w/ UF.**
s_{min}	**For all param. range, partly replicable w/ shift w/ UF.** Trends roughly replicated. Peak position different.
$L^f(L^l)$	**For all param. range, partly replicable w/o shift w/ UF.** Unnecessary sign alternation observed.

Table 3.19: Replicablity of respective parameters in the GHR model for followers' acceleration of various following vehicles.

	Followers' acceleration
C	Variation directed to negative velocity & sign oscillation in A observed. **For all param. range, no necessary variation w/ UF.**
T	**For all param. range, no necessary variation w/ UF.**
$s_{\min}$	**For all param. range, no necessary variation w/ UF.** Only small effect.
$L^{\mathrm{f}}(L^{\mathrm{l}})$	**For all param. range, no necessary variation w/ UF.** Only small effect.

Chapter 4

Discussions

4.1 Characteristic Behaviors of Respective Vehicles and Their Causes

In this section, we integrate the results obtained in Chapter 2 and 3. A multiple regression analysis was used to clarify that the height of the following vehicles affected their acceleration and deceleration. This tendency implied that drivers prefer the similar magnitudes for the pitching direction jerk when they drive different vehicles. On the other hand, the magnitudes of the acceleration and deceleration increased with the size of leading vehicles. Furthermore, the operational delay of followers decreased when the height of the leading vehicles increased. These tendencies implied that followers' response to the driving behaviors of the leading vehicles was improved by the better visibility of the leaders. In addition, large leaders caused shorter distance gaps. The following drivers could unintentionally approach the leaders as a result of carelessness; they could assume that they were still safe if they approached the leader because they could recognize the small deceleration of the leader.

Based on the decision tree for the distance gap, we clarified that motorcycles have a small distance gap in the steady phase. The riders would approach the leaders because they assume that motorcycles are agile. The trucks had smaller distance gaps than the cars in the early deceleration phase. This was because of the delay in the deceleration, i.e., inertia of the trucks.

Regarding the velocity difference, the motorcycles had a smaller velocity difference than the other vehicles in the steady and deceleration phases. This was because the motorcycles could adjust their behavior to the leaders as a result of their small inertia. At the beginning of the acceleration phase, the velocity difference of the cars was larger than that of the trucks. This meant that the response of the cars to the leaders' acceleration was slow. In addition, the velocity difference of the trucks was larger than that of the cars in the latter deceleration phase. This meant that the trucks finished their deceleration earlier than the cars. These tendencies implied that the drivers compensated for the slowness of the trucks by their response speed.

In the acceleration, we observed timing difference of the peak between the cars and trucks, and small decelerations of the cars in the early deceleration phase. These characteristics could be explained by the carelessness of the car drivers.

148

Based on these interpretations, the drivers' recognition of the agility of a vehicle affected the distance gap and timings of the acceleration and deceleration. At the same time, the visibility of the leaders affected the distance gap. The drivers of motorcycles could adjust their driving to the leaders but tended to have short distance gaps because of their recognition of their agility. The drivers of trucks compensated for the slowness of their vehicles by their response speed. In particular, they tended to act safely when they were in the deceleration phase. The drivers of cars tended to be careless in reaction to the behaviors of the leaders because of their confidence in their agility. This tendency resulted in large magnitudes for the acceleration and deceleration.

Based on these ideas, we theorize that the auto-rickshaws would have a short distance gap, and the distance gap behind them would be short. We also theorize that electric normal passenger cars would have large magnitudes for their acceleration and deceleration as a result of the drivers' carelessness. In addition, electric cars might have short distance gaps because they have quick acceleration like motorcycles. There is a possibility that the electric cars might not show this quick acceleration because the drivers could prefer similar jerks on each vehicle they drive.

4.2 Proposals Based on Investigations

4.2.1 Improvement of Car-Following Models

We presented the main factors affecting the followers' driving in Section 4.1. There were the drivers' recognition of their vehicles' agility and the visibility of the leaders. These factors affect the respective vehicles in different phases and timings. Therefore, we propose that a model with asymmetric parameters for each phase would replicate the characteristic behaviors of the respective vehicles in mixed traffic. In other words, the model for mixed traffic should have terms for the steady state, action-starting state, controlling state, and action-finishing state. In the steady state, vehicles stand still or travel at a steady velocity. In this state, they have achieved their targets for the velocity, distance gap, etc. In the action-starting state, the drivers receive stimulation to start their operation. We need some parameters indicating the response time and threshold for this stimulation. In the controlling state, the drivers control their vehicles based on their attitude toward safety, their driving preference, the behaviors of the leaders, etc. In the action-finishing state, the drivers smoothly finish their control and transfer to the steady state. Model parameters for each state or term are necessary, but these parameters should be affected by the drivers' recognition of their vehicles' agility and the visibility of the leaders.

4.2.2 Jam Reduction for Mixed Traffic in Developing Countries

In general, traffic jams are caused by delays in drivers' responses and short distance gaps. A traffic jam grows because of a lack of acceleration at its head. We expect that these causes could be reduced by controlling of the drivers' recognition and visibility of the leaders.

A "staggered formation" of motorcycles would be one solution to realize large distance gaps. Figure 4.1 shows a schematic of such a staggered formation. Respective motorcycles shift their traveling lines from the nearest leader and follow the next-nearest leader. This formation can realize a high density with the large distance gaps and is already popular with riders. However, promoting this formation through driver education, traffic signs, etc. could prevent both the occurrence and growth of traffic jams.
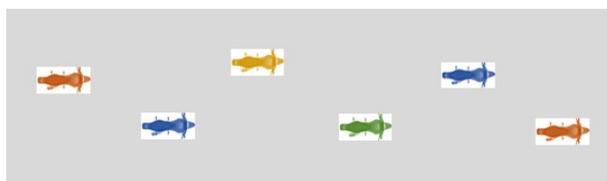
Figure 4.1: A schematic of staggered formation of motorcycles.

Large distance gaps would also be promoted if it becomes difficult to measure the distance gaps from the leaders. Displaying a visual pattern such as an optical illusion on the backs of trucks would be one solution. In addition, brake lights mounted on the top of a vehicle in the back might produce a shorter distance gap. Improving of such safety equipment could help to reduce traffic jams.

On the other hand, in order to improve the response speed through agility recognition, we need to reduce the acceleration and deceleration performances of cars at the beginning of the acceleration and deceleration. The suspension setting and the engine control could be used to control the jerk that drivers experience.

The market penetration of electric vehicles will increase in the near future, even in developing countries. However, as we discussed in Section 4.1, it is not clear whether electric vehicles can reduce traffic jams. Their quick acceleration might cause short distance gaps and drivers' carelessness.

4.2.3 Jam Reduction Method for Mixed Traffic Comprising Manual and Autonomous Driving Vehicles: Crowd Optimization with Emergent Formation-Control

In Section 4.2.2, we proposed some ideas to reduce traffic jams of current mixed traffic. On the other hand, the reduction of traffic jams involving autonomous driving vehicles is being investigated by many researchers, as we mentioned in Section 1.1. In a few decades, autonomous driving vehicles will be introduced in mixed traffic.

Driving strategies for the distance gap, acceleration, etc. will be the main factors to reduce traffic jams involving autonomous driving vehicles. However, if an autonomous driving strategy also considers and utilizes the characteristics of respective manual driving vehicles, jam reduction will be more efficient, and the comfort of all the drivers will also be improved.

For example, most motorcycles will be manually operated in such an era. Although autonomous driving motorcycles are being developed, they will not be very popular because the autonomous driving technology does not contribute to the convenience, comfort, and riding pleasure of motorcycles. If autonomous driving vehicles do not isolate each motorcycle in a traffic stream, it will be possible to promote the staggered formation proposed in Section 4.2.2.

We demonstrated various driving changes caused by a combination of vehicles in this thesis. If an autonomous driving strategy controls the order and formation of vehicles, manual driving vehicles will show the preferable characteristics needed for jam reduction. Furthermore, the order and formation of the vehicles can realize a safer and more comfortable traffic stream, which will relieve all the drivers from any unnecessary strain caused by mixed traffic.

This idea can be rephrased in a more general description: crowd optimization with emergent formation-control. Imagine that there is a self-driving particle system. If we change the characteristics of a portion of the particles, this will affect the entire system, i.e., the formation and configuration of the particles will be changed. Such an environmental change will affect other particles and change their characteristics. If we control these chained effects, we can control the state of the system, i.e., the macroscopic behavior of the crowd of the particles. We expect this concept to be utilized for other crowd systems, as well as for vehicle traffic.

Chapter 5

Conclusions

To address the traffic jam which has become a serious and worldwide problem, the effectiveness of the jam reduction relying on the improvement of driving methods has been confirmed by many researchers. This approach for jam reduction is effective especially in the developing countries as well as in developed countries because it does not require the allocation of a huge amount of resources for infrastructure improvement. However, mixed traffic, which comprises many types of vehicles in these developing countries, shows phenomena different from uniform traffic owing to its heterogeneity, and has recently attracted the interest of many researchers.

In this thesis, we clarified the difficulty of replicating the behavioral changes of various types of vehicles in mixed traffic with variable adjustments of existing car-following models. It was also proposed that the engine performance of the following vehicles need to be taken into account as well as the vehicle height, weight, and length of the leading and following vehicles when we replicate the behavioral differences of various types of vehicles. Based on these insights, it was implied that the environmental factors that drivers received from the leading and their vehicles, i.e., drivers' recognition of their vehicles' agility and the visibility of the leaders, mainly affected the followers' driving. Because these factors affect the respective drivers in different phases and timings, a model for mixed traffic should have asymmetric parameters for each operation phase. Furthermore, we discussed jam reduction methods for not only current mixed traffic but also future mixed traffic comprising manual and autonomous driving vehicles. In this discussion, we proposed a concept of "crowd optimization with emergent formation-control."

In Chapter 2, we first conducted a series of car-following experiments in a test circuit with motorcycles, passenger cars, and trucks in order to obtain all the behaviors of known vehicles in the acceleration, steady driving, and deceleration phases. Because we obtained all the vehicle and driving characteristics of the leading and following vehicles, we were able to analyze the behaviors of the following vehicles by the multiple regression analysis. Our experiment with the motorcycles was enabled by the miniaturization of the data recording devices, i.e., tablets.

Through the analysis of the experimental results, in Chapter 2, we found that

1. the maximum velocity and acceleration of the following vehicles are mainly affected by the driving of the leading vehicle, and

2. the maximum deceleration, distance gaps when the platoons starts, the maximum distance gap during a trial, and delay in maximum acceleration timing are also affected by the vehicle characteristics of the leading and following vehicles.

In Chapter 3, we determined the characteristic behaviors of the respective vehicles through all the procedures, from the acceleration to deceleration, i.e., the characteristic leaves. By the dynamic time warping and the decision tree learning based on the shapelets, we were able to extract the features of these characteristic leaves and compared them to those caused by the variation of car-following model parameters. With this approach, we investigated the responses of the car-following models under the transient phases, which are ignored in the error minimization of the macroscopic values (e.g., the flow) and of the acceleration over the whole observed time range.

Based on the comparison of the observed characteristic leaves of three types of vehicles and simulated characteristic leaves varied by the car-following model parameters, we clarified desired and unnecessary features replicated and caused by the parameter variation. From these results, we made the following conclusions:

1. There are no parameters in the investigated car-following models that can replicate the differences observed in all the distance gaps, velocity differences, and followers' acceleration when we change the types of the followers.

2. Although features in the distance gap can be replicated by some parameters in the models, respective models are weak in different physical values regarding the feature replication.

In Chapter 4, we integrated the results obtained in Chapter 2 and 3, and concluded that the drivers' recognition of their vehicles' agility and the visibility of the leaders were the main factors affecting the followers' driving. We also proposed formulation of a car-following model replicating mixed traffic and jam reduction methods for current and future mixed traffic. Through the discussion, we reached a concept of crowd optimization with emergent formation-control, which can be utilized for other crowd systems, as well as for vehicle traffic.

Regarding future works and prospects, it will be easy to estimate the driving characteristics of the new types of vehicles introduced in the consideration of mixed traffic based on our regression models. On the other hand, for the parameters determining the maximum deceleration, minimum and maximum distance gaps and the operational delay, we need to take the vehicle height and engine performance into account. In addition, it was suggested that the conflicting behaviors in the distance gaps need to be thoroughly investigated for better understanding of the psychological aspects of the drivers in relation to safety training, comfortable autonomous driving, etc.

It was implied that when we focus on the transient phenomena, there is the possibility that errors are included in the discussions based on the existing car-following models. Therefore, it is necessary to propose and validate new car-following models that accurately replicate the differences in the velocity difference and the acceleration of various types of vehicles. The detailed discussion regarding the replication performance of each parameter will be the basis of new models replicating driving behaviors in mixed traffic.

While we removed the effects of individual drivers in the regression analysis by dummy variables, we could not perfectly remove these effects in the analysis based on the characteristic leaves. Because it needs a huge number of trials with many subjects and types of vehicles in order to ignore the individual tendencies, some data processing method removing these effects from time series will make our discussion more accurate. On the other hand, if we obtain the characteristic leaves in real traffic of various countries, the difference between them, i.e., tendencies of drivers in respective countries, will be clarified by our method.

Furthermore, in order to realize the concept of crowd optimization with emergent formation-control in vehicle traffic, we need to investigate the order and formation of vehicles in current mixed traffic.

In this thesis, we analyzed and discussed variations in the driving of the following drivers on various types of vehicles from the perspectives of phenomena, causes, and replication. The first contribution of this thesis is capturing that drivers' recognition of their vehicles' agility and the visibility of the leaders mainly affect the followers' driving in different phases and timings. The other contributions are the proposals for model formulation and jam reduction methods for mixed traffic. Based on our research, we hope that studies on mixed traffic in developing countries, on safety driving, on jam reduction for heterogeneous traffic, and on comfortable autonomous driving under various market penetration rates of the autonomous vehicles will be advanced in the future.

Bibliography

[1] Development Policy and United Nations Analysis Division. World Economic Situation and Prospects as of mid-2016, 2016. URL `http://www.un.org/en/development/desa/policy/wesp/wesp_current/2016wesp_update.pdf`. Last visited: December 12, 2016.

[2] Department of Economics Population Division and United Nations Social Affairs. World Population Prospects, 2016. URL `https://esa.un.org/unpd/wpp/Publications/Files/WPP2015_DataBooklet.pdf`. Last visited: December 12, 2016.

[3] Dirk Helbing. Traffic and related self-driven many-particle systems. *Reviews of modern physics*, 73(4):1067, 2001.

[4] Mark Brackstone and Mike McDonald. Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2(4): 181–196, 1999.

[5] Ryosuke Nishi, Akiyasu Tomoeda, Kenichiro Shimura, and Katsuhiro Nishinari. Theory of jam-absorption driving. *Transportation Research Part B: Methodological*, 50:116–129, 2013.

[6] Yohei Taniguchi, Ryosuke Nishi, Akiyasu Tomoeda, Kenichiro Shimura, Takahiro Ezaki, and Katsuhiro Nishinari. A demonstration experiment of a theory of jam-absorption driving. In *Traffic and Granular Flow'13*, pages 479–483. Springer, 2015.

[7] Robert Herman, Elliott W Montroll, Renfrey B Potts, and Richard W Rothery. Traffic dynamics: analysis of stability in car following. *Operations research*, 7(1):86–106, 1959.

[8] Takashi Nagatani. Stabilization and enhancement of traffic flow by the next-nearest-neighbor interaction. *Physical Review E*, 60(6):6395, 1999.

[9] Akihiro Nakayama, Yūki Sugiyama, and Katsuya Hasebe. Effect of looking at the car that follows in an optimal velocity model of traffic flow. *Physical Review E*, 65(1):016112, 2001.

[10] Dirk Helbing, Ansgar Hennecke, Vladimir Shvetsov, and Martin Treiber. MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model. *Transportation Research Part B: Methodological*, 35(2):183–211, 2001.

[11] Nils Eissfeldt and Peter Wagner. Effects of anticipatory driving in a traffic flow model. *The European Physical Journal B-Condensed Matter and Complex Systems*, 33(1):121–129, 2003.

[12] HX Ge, SQ Dai, LY Dong, and Yu Xue. Stabilization effect of traffic flow in an extended car-following model based on an intelligent transportation system application. *Physical Review E*, 70(6):066134, 2004.

[13] Michael Sivak and Brandon Schoettle. Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles. 2015.

[14] Pravin Varaiya and Steven E Shladover. Sketch of an IVHS systems architecture. In *Vehicle Navigation and Information Systems Conference, 1991*, volume 2, pages 909–922. IEEE, 1991.

[15] U Hanebutte, E Doss, T Ewing, and A Tentner. Simulation of vehicle traffic on an automated highway system. *Mathematical and computer modelling*, 27(9-11):129–141, 1998.

[16] Robert A Ferlis. The Dream of the Automated Highway. *Public roads*, 71 (1), 2007.

[17] Bart Van Arem, Cornelie JG Van Driel, and Ruben Visser. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, 7(4):429–436, 2006.

[18] Gerrit JL Naus, Rene PA Vugts, Jeroen Ploeg, Marinus JG van de Molengraft, and Maarten Steinbuch. String-stable CACC design and experimental validation: A frequency-domain approach. *IEEE Transactions on vehicular technology*, 59(9):4268–4279, 2010.

[19] Meng Wang, Winnie Daamen, Serge P Hoogendoorn, and Bart van Arem. Rolling horizon control framework for driver assistance systems. Part II: Cooperative sensing and cooperative control. *Transportation research part C: emerging technologies*, 40:290–311, 2014.

[20] Alireza Talebpour and Hani S Mahmassani. Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies*, 71:143–163, 2016.

[21] Jeroen Ploeg, Bart TM Scheepers, Ellen Van Nunen, Nathan Van de Wouw, and Henk Nijmeijer. Design and experimental evaluation of cooperative adaptive cruise control. In *2011 14th International IEEE Conference on intelligent transportation systems (ITSC 2011),*, pages 260–265. IEEE, 2011.

[22] Tien-Pen Hsu, Ahmad Farhan Mohd Sadullah, and Nyugen Xuan Dao. A comparison study on motorcycle traffic development in some Asian countries–case of Taiwan, Malaysia and Vietnam. *The eastern Asia society for transportation studies (EASTS), Int. Cooperative Res. Activity*, 2003.

[23] James R Sayer. The effect of lead-vehicle size on driver following behavior. 2000.

[24] Majid Sarvi. Heavy commercial vehicles-following behavior and interactions with different vehicle classes. *Journal of advanced transportation*, 47(6): 572–580, 2013.

[25] Kayvan Aghabayk, William Young, Majid Sarvi, and Yibing Wang. Examining vehicle interactions during a vehicle-following manoeuvre. In *Australasian Transport Research Forum (ATRF), 34th, 2011, Adelaide, South Australia, Australia*, volume 34, 2011.

[26] B Bhavathrathan and C Mallikarjuna. Evolution of macroscopic models for modeling the heterogeneous traffic: an Indian perspective. *Transportation Letters*, 4(1):29–39, 2012.

[27] Hari Krishna Gaddam, Anjaneyulu Chinthireddy, and K Ramachandra Rao. Comparison of Numerical Schemes for LWR Model under Heterogeneous Traffic Conditions. *Periodica Polytechnica. Transportation Engineering*, 44(3):132, 2016.

[28] Caleb Ronald Munigety, Prakkhar Alok Gupta, Krishna Murthy Gurumurthy, Srinivas Peeta, and Tom V Mathew. Vehicle-type dependent car following model using spring-mass-damper dynamics for heterogeneous traffic. In *Transportation Research Board 95th Annual Meeting*, number 16-5025, 2016.

[29] Saskia Ossen and Serge P Hoogendoorn. Heterogeneity in car-following behavior: Theory and empirics. *Transportation research part C: emerging technologies*, 19(2):182–195, 2011.

[30] Danjue Chen, Soyoung Ahn, Soohyuk Bang, and David Noyce. Car-Following and Lane-Changing Behavior Involving Heavy Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, (2561):89–97, 2016.

[31] Danjue Chen, Jorge Laval, Zuduo Zheng, and Soyoung Ahn. A behavioral car-following model that captures traffic oscillations. *Transportation research part B: methodological*, 46(6):744–761, 2012.

[32] Anthony D Mason and Andrew W Woods. Car-following model of multi-species systems of road traffic. *Physical Review E*, 55(3):2203, 1997.

[33] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, 62(2):1805, 2000.

[34] Da Yang, Jing Jin, Bin Ran, Yun Pu, and Fei Yang. Modeling and analysis of car-truck heterogeneous traffic flow based on intelligent driver car-following model. In *Transportation Research Board 92nd Annual Meeting*, number 13-2358, 2013.

[35] Tzu-Chang Lee. *An agent-based model to simulate motorcycle behaviour in mixed traffic flow.* PhD thesis, Imperial College London (University of London), 2007.

[36] KI Wong and Tzu-Chang Lee. Investigating the heterogeneity in driving behavior of powered two-wheelers under mixed traffic flow. In *Transportation Research Board 94th Annual Meeting*, number 15-1165, 2015.

[37] Yasuhiro Shiomi, Teruaki Hanamori, Nobuhiro Uno, and Hiroshi Shimamoto. Modeling mixed traffic flow with motorcycles based on discrete choice approach. In *Transportation Research Board 93rd Annual Meeting*, number 14-4846, 2014.

[38] Timothy Oketch. New modeling approach for mixed-traffic streams with nonmotorized vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, (1705):61–69, 2000.

[39] Yuki Sugiyama, Minoru Fukui, Macoto Kikuchi, Katsuya Hasebe, Akihiro Nakayama, Katsuhiro Nishinari, Shin-ichi Tadaki, and Satoshi Yukawa. Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. *New journal of physics*, 10(3):033001, 2008.

[40] Rui Jiang, Mao-Bin Hu, HM Zhang, Zi-You Gao, Bin Jia, Qing-Song Wu, Bing Wang, and Ming Yang. Traffic experiment reveals the nature of car-following. *PloS one*, 9(4):e94351, 2014.

[41] Rui Jiang, Mao-Bin Hu, HM Zhang, Zi-You Gao, Bin Jia, and Qing-Song Wu. On some experimental features of car-following behavior and how to model them. *Transportation Research Part B: Methodological*, 80:338–354, 2015.

[42] Akihito Nagahama, Daichi Yanagisawa, and Katsuhiro Nishinari. Dependence of driving characteristics upon follower–leader combination. *Physica A: Statistical Mechanics and its Applications*, 483:503–516, 2017. doi: 10.1016/j.physa.2017.04.136. URL https://doi.org/10.1016/j.physa.2017.04.136.

[43] M Yu Byron, Krishna V Shenoy, and Maneesh Sahani. Derivation of Kalman Filtering and Smoothing Equations. URL http://www.ece.cmu.edu/~byronyu/papers/derive_ks.pdf. Last visited: March 15, 2017.

[44] Japan Automobile Standards Internationalization Center. Road Vehicles Act. In *Automobile type approval handbook for Japanese certification*, volume 1, pages 33–103. 1997.

[45] Trade Ministry of Economy and Industry in Japan. 2007 Report on size-JPN 2004-2006, 2007. URL http://warp.da.ndl.go.jp/info:ndljp/pid/281883/www.meti.go.jp/press/20071001007/20071001007.html. Last visited: July 3, 2016.

[46] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.

[47] Michael J Lighthill and Gerald Beresford Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 229, pages 317–345. The Royal Society, 1955.

[48] Paul I Richards. Shock waves on the highway. *Operations research*, 4(1): 42–51, 1956.

[49] Harold J Payne. Models of freeway traffic and control. *Mathematical models of public systems*, 1971.

[50] GB Whitham and Richard G Fowler. Linear and nonlinear waves. *Physics Today*, 28:55, 1975.

[51] Boris S Kerner and Peter Konhäuser. Cluster effect in initially homogeneous traffic flow. *Physical review E*, 48(4):R2335, 1993.

[52] Ilya Prigogine. A Boltzmann-like approach to the statistical theory of traffic flow. 1961.

[53] H Lenz, CK Wagner, and R Sollacher. Multi-anticipative car-following model. *The European Physical Journal B-Condensed Matter and Complex Systems*, 7(2):331–335, 1999.

[54] Katsuya Hasebe, Akihiro Nakayama, and Yūki Sugiyama. Equivalence of linear response among extended optimal velocity models. *Physical Review E*, 69(1):017103, 2004.

[55] B Chopard and M Droz. *Cellular automata*. Springer, 1998.

[56] Debashish Chowdhury, Vishwesha Guttal, Katsuhiro Nishinari, and Andreas Schadschneider. A cellular-automata model of flow in ant trails: non-monotonic variation of speed with density. *Journal of Physics A: Mathematical and General*, 35(41):L573, 2002.

[57] Debashish Chowdhury, Andreas Schadschneider, and Katsuhiro Nishinari. Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms. *Physics of Life reviews*, 2(4):318–352, 2005.

[58] Marco Bartolozzi and Anthony William Thomas. Stochastic cellular automata model for stock market dynamics. *Physical review E*, 69(4):046112, 2004.

[59] Louis A Pipes. An operational analysis of traffic dynamics. *Journal of applied physics*, 24(3):274–281, 1953.

[60] Robert E Chandler, Robert Herman, and Elliott W Montroll. Traffic dynamics: studies in car following. *Operations research*, 6(2):165–184, 1958.

[61] Gordon Frank Newell. Nonlinear effects in the dynamics of car following. *Operations research*, 9(2):209–229, 1961.

[62] Masako Bando, Katsuya Hasebe, Akihiro Nakayama, Akihiro Shibata, and Yuki Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical review E*, 51(2):1035, 1995.

[63] Masako Bando, Katsuya Hasebe, Ken Nakanishi, and Akihiro Nakayama. Analysis of optimal velocity model with explicit delay. *Physical Review E*, 58(5):5429, 1998.

[64] Masako Bando, Katsuya Hasebe, Ken Nakanishi, Akihiro Nakayama, Akihiro Shibata, and Yūki Sugiyama. Phenomenological study of dynamical model of traffic flow. *Journal de Physique I*, 5(11):1389–1399, 1995.

[65] Rui Jiang, Qingsong Wu, and Zuojin Zhu. Full velocity difference model for a car-following theory. *Physical Review E*, 64(1):017101, 2001.

[66] Peter G Gipps. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2):105–111, 1981.

[67] Martin Treiber and Arne Kesting. *Traffic Flow Dynamics*. Springer, 2013.

[68] Eiji Kometani and Tsuna Sasaki. On the stability of traffic flow (report-I). *J. Oper. Res. Soc. Japan*, 2(1):11–26, 1958.

[69] Robert Herman and Renfrey B Potts. Single lane traffic theory and experiment. In *Proceedings Symposium on Theory of Traffic Flow*, pages 120–146. Elsevier, 1959.

[70] Leslie C Edie. Car-following and steady-state theory for noncongested traffic. *Operations research*, 9(1):66–76, 1961.

[71] Denos C Gazis, Robert Herman, and Richard W Rothery. Nonlinear follow-the-leader models of traffic flow. *Operations research*, 9(4):545–567, 1961.

[72] Adolf D May Jr and EM Harmut. Non-integer car-following models. *Highway Research Record*, (199), 1967.

[73] MP Heyes and R Ashworth. Further research on car-following models. *Transportation Research*, 6(3):287–291, 1972.

[74] Joseph Treiterer and Jeffrey Myers. The hysteresis phenomenon in traffic flow. *Transportation and traffic theory*, 6:13–38, 1974.

[75] Avishai Ceder and Adolf D May. Further evaluation of single-and two-regime traffic flow models. *Transportation Research Record*, (567), 1976.

[76] Maurice Aron. Car following in an urban network: simulation and experiments. *Planning And Transport Research and Computation*, 1988.

[77] H Ozaki. Reaction and anticipation in the car-following behavior. *Transportation and traffic theory*, 12:349–366, 1993.

[78] Walter Helly. Simulation of bottlenecks in single-lane traffic flow. In *Proc. Symposium on Theory of Traffic Flow, 1959*, pages 207–238. Research Laboratories, 1959.

[79] RM Michaels. Perceptual factors in car following. In *Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow (London, England), OECD*, 1963.

[80] Shinya Kikuchi and Partha Chakroborty. Car-following model based on fuzzy inference system. *Transportation Research Record*, pages 82–82, 1992.

[81] Evelyn Fix and Joseph L Hodges Jr. Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, California Univ Berkeley, 1951.

[82] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[83] Jan Salomon Cramer. The origins of logistic regression. 2002.

[84] Vladimir Vapnik. Pattern recognition using generalized portrait method. *Automation and remote control*, 24:774–780, 1963.

[85] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[86] Earl B Hunt, Janet Marin, and Philip J Stone. Experiments in induction. 1966.

[87] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[88] Wei-Yin Loh. Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348, 2014.

[89] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[90] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[91] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[92] W. L. Buntine. *A Theory of Learning Classification Rules.* PhD thesis, University of Technology, Sydney, 1990.

[93] Igor Kononenko. Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition. *Current trends in knowledge acquisition*, pages 190–197, 1990.

[94] Pat Langley and Stephanie Sage. Oblivious decision trees and abstract cases. In *Working notes of the AAAI-94 workshop on case-based reasoning*, pages 113–117. Seattle, WA, 1994.

[95] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[96] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[97] Lexiang Ye and Eamonn Keogh. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data mining and knowledge discovery*, 22(1):149–182, 2011.

[98] Li Liu, Yuxin Peng, Ming Liu, and Zigang Huang. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowledge-Based Systems*, 90:138–152, 2015.

[99] Simon Malinowski, Brigitte Chebel-Morello, and Noureddine Zerhouni. Remaining useful life estimation based on discriminating shapelet extraction. *Reliability Engineering & System Safety*, 142:279–288, 2015.

[100] R. Bellman and R. Kalaba. On adaptive control processes. *IRE Transactions on Automatic Control*, 4(2):1–9, November 1958. ISSN 0096-199X. doi: 10.1109/TAC.1959.1104847.

[101] *Yattakoto no setsumei* (Explanation of what I did), 2017. URL http://ksknw.hatenablog.com/entry/2017/03/26/234048. Last visited: November 11, 2017. in Japanese.

[102] J Xing. A Parameter Identification of a Car-Following Model. In *Steps Forward. Intelligent Transport Systems World Congress*, volume 4, 1995.

Acknowledgements

I would like to thank Prof. Katsuhiro Nishinari for giving me an opportunity to conduct this interesting study and for many valuable suggestions. I would also like to express my special and cordial appreciation for his encouragement to my career as a researcher. I am deeply grateful to Dr. Daichi Yanagisawa for giving many constructive advices for my research and a lot of encouragement during daily meetings. I am also grateful for his practical guidance regarding academic activities.

I wish to express my special thanks to Prof. Kenryu Nakamura for giving several valuable comments from an interdisciplinary point of view and for providing basic knowledge of human support. I would like to express my gratitude to Prof. Kumiko Tanaka for several useful comments, which helped in developing this thesis objective. I am deeply grateful to Dr. Kiichiro Hatoyama for his constructive advices from the perspective of traffic engineering. I would like to express my cordial appreciation to Dr. Shuichi Matsumoto for his constructive comments on my research.

I would like to deeply thank Mr. Naotaka Ishizawa, Mr. Prashant Walawalkar, and other members of NYK Auto Logistics (India) Private Limited. for providing a great support for conducting field observation in India. I am deeply grateful to my ex-colleagues of Continental Automotive Corporation and Continental Automotive Japan for giving me an opportunity to proceed to the doctoral course.

I would like to express my cordial appreciation to Dr. Jun Sato for his helpful comments from the perspective of mathematical physics. I wish to express my cordial appreciation to Dr. Kenichiro Shimura for providing essential comments about data collection. I would like to deeply thank Dr. Claudio Feliciani for his practical advices on data processing and his continued friendship. I am deeply grateful to Dr. Takahiro Ezaki and Dr. Satori Tsuzuki for their constructive advices on my research and career. I cordially thank Dr. Shingo Ichiki, Dr. Doan Phi, and Dr. Takahiro Tannnai for their hearty encouragements in the laboratory.

I would like to express my cordial appreciation to Mr. Keiichiro Fukuda for his continued encouragement in daily life. I express my appreciation to Mr. Haruki Ishikawa and Mr. Yu-Li Tsai for their friendship. I cordially express my thanks to Ms. Takako Hoshiyama, Ms. Xiaolu Jia, Mr. Hidetaka Ito, Mr. Hikaru Miyashita, Mr. Ryoya Awata, Mr. Kohei Kikuchi, Mr. Akio Muro, Mr. Masahiro Tanaka, Mr. Hiroki Yamamoto, Mr. Yusuke Miyoshi, Mr. Koki Nagao, Mr. Fumitaka Sumiyama, Mr. Akihiro Fujita, Ms. Kasumi Hatori, Mr. Jun Hattori, Mr. Sho Ishii, Mr. Takahumi Omori, Mr. Daiki Watanabe, Mr. Wenlong Li, and Ms. Riho Kawaguchi for their continuous help and warm encouragements in the laboratory. I express my hearty appreciation to Ms. Chika Masujima, Ms. Yuko Uchiyama, Ms. Yuko Tsumori, Ms. Kanako Nakagawa, Ms. Yoriko Yokoyama, and Ms. Erika Shiihara for their kind assistances in the laboratory. I would like to deeply thank my friends at Mitaka International Hall of Residence for their warm support and friendship. Finally, I specially thank my grandmothers, my father, my mother, and my sister for their continued encouragement and warm support, and for giving me various opportunities throughout my life.