2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Modeling of Cyclists' Decision for Left-Turn Vehicle at Unsignalized Intersection using Logistic Regression Model and Gaussian Mixuter Model

Ryo Wakisaka[1] Takuma Yamaguchi[1] Kazunori Ban[1] Hiroyuki Okuda[2] Tatsuya Suzuki[2]

Abstract— **In this paper, the cyclists' decision-making behavior in the interaction with a car at unsignaled intersection is measured and analyzed. Based on the measured data, the cyclists' decision-making model is identified by using logistic regression model. Since the data collection in the real world is hard to realize, we have used an interactive simulator in which the cycling simulator and the driving simulator are connected via network, and share the same virtual traffic environment. The cyclist's decision states are defined by three states regarding their operation, pedaling-on, pedaling-off and brake-on. The models to estimate these three states were constructed using the logistic regression model and the Gaussian mixture model, respectively. Finally the accuracy of the constructed models are verified, and compared with each other.**

I. INTRODUCTION

With the growing emphasis on environmental sustainability and health promotion, the use of bicycles is expected to increase. Therefore, the demand for safety evaluation of traffic environments that include bicycles will also increase. However, evaluation and verification of traffic systems involving bicycles in real environments is difficult due to safety issues and high costs in terms of both time and money. Therefore, evaluation and verification in a virtual environment using simulation is expected.

In order to evaluate the safety functions and operational rules of various types of mobility, including bicycles, through simulation, it is necessary to recreate realistic traffic environments within a virtual space. In a real traffic environment, traffic participants interact with each other to determine their behavior. Therefore, it is necessary to model behaviors that take into account interactions among traffic participants.

In modeling the behavior of vulnerable road users such as pedestrians and bicyclists, different approaches are often taken from those used for driver models in autonomous driving. For pedestrians, various motion and decision models have been studied, such as social force models and Bayesian network models [1]–[4]. On the other hand, there have been fewer studies on bicycles than on pedestrians. In some cases, bicycles are defined as pedestrians with different speeds [5]. However, bicycles have wider speed range than pedestrians, and their speed is finely adjusted by pedaling changes in addition to braking operations. Moreover, cyclists often coast to create time to assess the behavior of interacting traffic participants or to adjust their timing when passing through intersections. This behavioral intention differs from that of cars, where coasting is primarily aimed at reducing fuel consumption. Therefore, it is difficult to reproduce the behavior of bicycles and represent individual differences simply by changing the speed of a pedestrian or driver model. In addition, because bicycles are subject to a large load when accelerating from a standstill, some people may unconsciously adjust their speed to avoid coming to a complete stop. This suggests that bicycles make different decisions on acceleration and deceleration than pedestrians. Furthermore, bicycles do not have a strictly defined riding area like automobile lanes, and they travel on both sidewalks and roadways, which causes frequent interactions with both pedestrians and cars. For these reasons, accurately modeling the unique decisions and movements of cyclists is crucial for conducting effective traffic simulations.

Developing accurate cyclist models containing the decisions and motions in real-world environments allows for integration with existing pedestrian and driver models to create realistic traffic simulations. These simulations are expected to provide an efficient and effective evaluation of driver assistance systems, autonomous driving systems and transportation infrastructure. Such an approach holds significant potential for advancing the development of safe and comfortable mobility spaces within complex and diverse traffic conditions.

Measured data is often used to analyze and model the behavior of cyclists. The data for modeling cyclist behavior can be real-world data collected through sensors such as cameras [6]–[8] or virtual environment data obtained using a cycling simulator (CS) [9], [10]. In particular, data collection using a CS allows for safe and efficient measurement, making the development of CSs [11]–[13] a topic of significant interest. In a previous study on behavioral models of cyclists (bicycles), Alsaleh et al. [14] modeled interactions in mixed traffic with pedestrians and bicycles. Using inverse reinforcement learning with video data, they predict the following and overtaking behavior of cyclists toward pedestrians. However, the interaction with cars is not modeled because cars do not exist in the assumed traffic space. X. Li et al. [15] studied the differences in cyclist behavior across three countries with varying income levels and bicycle prevalence. The results showed that differences in road environments, policies, enforcement, and attributes such as age and gender lead to

[1]R.Wakisaka, T.Yamaguchi and K.Ban are with Toyota Technical Development Corporation, 1-9, Imae, Hanamoto-cho, Toyota, Aichi, Japan ryo.wakisaka@mail.toyota-td.jp

[2]H.Okuda and T.Suzuki are with Department of Mechanical System Engineering, Naogya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, Japan

variations in cyclist behavior. This suggests the importance of developing models capable of representing differences in cyclist behavioral characteristics to accurately reflect real-world cyclist behavior. Y. Li et al. [16] modeled illegal lane-changeing (ILC) behavior of bycicles by cellular automaton (CA) model. The simulation demonstrates the validity of the proposed model and analyzes the impact of ILC on traffic flow. While the proposed model considers the differences in ILC behavior between electric bicycles and regular bicycles, it does not address the individual characteristics of cyclists.

In this paper, we consider the construction of a behavioral model capable of expressing cyclist-specific behaviors based on interactions with others, as part of reproducing real traffic environments for simulation evaluation. Specifically, the cyclist's decision-making for a left-turning car at the unsignalized intersection is analyzed and modeled. The decision-making states of cyclists are defined as "Pedaling On," "Pedaling Off," and "Braking On." The model to estimate cyclist-specific behavioral intentions is constructed from pedaling and braking operations. To this end, we first measure and analyze cyclists' bicycling behavior through data measurement experiments using our originally developed CS. Then, using the measured data, the cyclist's decision-making model is constructed using logistic regression models (LRM) and Gaussian mixture models (GMM), respectively. Finally, constructed models were verified using the numerical simulation. Although the model proposed in this paper is a cyclist model, a similar analysis is expected to be applicable to the analysis and modeling of other small mobility devices.

II. CYCLIST MODEL

A. Model Structure

In reproducing the behavior of a cyclist in a traffic environment, the overall structure of the model in the simulation is defined as shown in Figure 1. In general, human behavior, including that of drivers of automobiles, is considered to be carried out through the processes of "cognition," "judgment," and "operation." The behavior model of a cyclist in Fig. 1 assumes this structure. First, it is assumed that the cyclist decides how to act based on the recognition results of the surrounding situation and determines the amount of operation on the bicycle interface such as handlebars and pedals. The bicycle moves according to the equation of state (bicycle model) according to the given operation amount. Then, the surrounding traffic participants react to the bicycle's movement and the situation changes. The cyclist then recognizes the changed surrounding situation again, makes a decision, and operates the bicycle again.

B. Modeling Target

In real-world traffic spaces, bicycles travel in a variety of environments and interact with a variety of traffic participants. In this paper, we focus on the interaction between bicycles and cars. It examines the modeling of cyclists' decision in dangerous bicycle-car scenarios. Intersections are the most likely locations for bicycle accidents to occur. Examples of accidents that may occur are head-on collisions

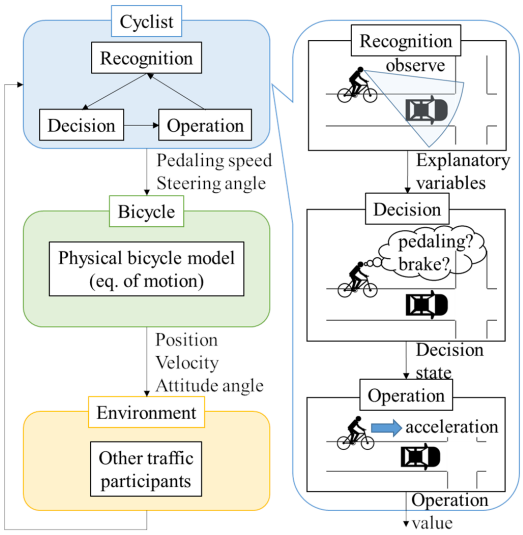


Fig. 1: Cyclist-bicycle-environment model

and left-turn entrapment accidents. In the case of a head-on collision, the cyclists do not recognize each other until just before the accident occurs, and their movements are similar to those of a run-off, so it is possible to represent their movements as a scenario. On the other hand, in the case of a left-right turn, the other party can be recognized before the accident occurs, and interaction occurs.

Therefore, it is necessary to model the interactive behavior in order to represent the motion in the simulation. One of the information that can be used to represent a cyclist's decision is the state of pedaling. Usually, the decision of a moving object moving in a straight line is often expressed by the acceleration and deceleration of the moving object. However, in the case of bicycles, cyclists often ride without pedaling or braking (coasting without operation input), and by skillfully using this state, they reduce the effort required for driving. In other words, to estimate the decision of cyclists, it is necessary to define the decision that combines pedaling and braking rather than acceleration/deceleration itself.

Therefore, in this study, we first improve the CS so that the pedaling state can be measured. Then, we measure the cyclist's behavior toward a left-turning car at an unsignalized intersection where interaction occurs between the cyclist and the car, and model the cyclist's decision.

III. EXPERIMENT

A. Experimental Environment

To measure the interaction between a bicycle and a car at an unsignalized intersection, we constructed a simulator environment that links a CS and a car driving simulator (DS). Figure 2 shows each simulator. In this experimental environment, the CS and DS share each other's position in the virtual space via a server program, making it possible to measure behaviors involving interactions between cyclists and drivers. The CS is equipped with displays surrounding the user, except for the rear left, allowing visual confirmation of the side and rear views. The DS, while lacking a dedicated rear view display, enables rear visibility through the rearview

(a) car driving simulator (b) cycling simulator

Fig. 2: Car driving simulator and cycling simulator

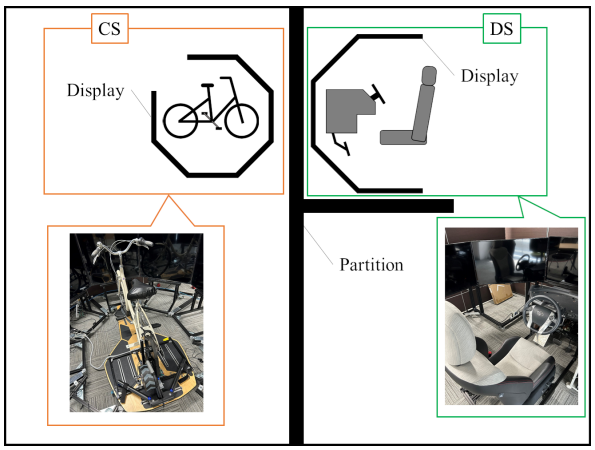


Fig. 3: Layout of simulators

and side mirrors. The cyclist in CS can recognize the turn signal lamps of the car in DS, and allowing the cyclist to know the driver's intention to make a left turn via the turn signals. The measurable data include positions, orientation, driving speeds, accelerations, and switch input of both simulators. Specifically, the CS records steering angles, brake inputs, and pedaling position, speed and torque, while the DS logs steering angles, gas and brake pedal input, and the status of the turn signals. In this experiment, the CS and DS are located in the same laboratory, and the cyclist and the driver are aware of each other that the other's movements are caused by real-time human operation. However, the simulators are separated from each other by a partition, so that they cannot recognize each other's operations. Figure 3 shows the layout of the simulator in the laboratory, where the CS can turn by leaning in addition to steering operation. In addition, a strain gage is attached to the crank of the pedals as a sensor for measuring pedaling information, enabling the presence or absence of pedaling to be obtained from the amount of strain on the crank.

B. Experimental Condition

Figure 4 shows the experimental scene. The experiment was conducted in pairs, with a cyclist operating the CS and a driver operating the DS. The initial position of the car

TABLE I: Initial conditions for experiment

v_{ref} [km/h]	l_{bi} [m]	l_{ci} [m]					
30	40 or 50	105	100	95	90	85	-
		80	75	70	65	60	45
40		138.3	133.3	128.3	123.3	118.3	-
		105	100	95	90	85	45

was changed so that the bicyclist and the car entered the intersection at around the same time when the speed of the car was set to 30[km/h] and 40[km/h], respectively. Table I shows the initial position patterns of the car. For the bicycles, the initial positions were determined based on the speed at which the bicycles were traveling during the practice ride before the start of the experiment. Specifically, the bicycles were set to be 50[m] from the intersection if the average speed during the practice run was 10[km/h] or higher, and 40[m] from the intersection if the average speed was less than 10[km/h].

On the DS side, the driver was instructed to set the target speed for each trial. The driver started driving at the signal to start the test and approached the intersection at the indicated target speed. After approaching the intersection, the driver operated his blinker at any time, watched the movement of the bicycles, and decided whether he would turn left first or yield to the bicycles. Although it is necessary to pull over to the left before making a left turn at an intersection by a car, in real traffic, it is not always possible to do so. In the CS, cyclists were instructed only on the psychological state (normal or hurried) of the cyclist during each trial. Cyclists were asked to ride assuming the psychological state they were told and to decide whether they would pass through the intersection first or yield to the car. The cyclists were asked to operate the simulator in the same way as when they normally ride a bicycle, including the gear settings, and were not instructed to set a target speed. In addition, instructions for each trial were displayed separately on the screen of each simulator to prevent the participants from predicting each other's behavior based on the instructions given to each other. Furthermore, since the cyclist and driver operated only one of the simulators, the pattern of instructions given by the other simulator was not known in advance.

To avoid familiarity with the experiment, a trial in which the car drove straight through the intersection was inserted once every 10 trials. There were 22 patterns for the initial position of the car and 2 patterns for the psychological state presented to the cyclist, for a total of 44 patterns in the test condition. The car moving straight ahead was added at a frequency of once every 10 times, making 48 patterns as one set, and three sets of data were measured for 14 pairs. This experiment was conducted with the approval of the Ethics Committee of Nagoya University and after obtaining informed consent from the subjects.

IV. Analysis of Measured Data

First, Table II shows each participant's basic information, such as the average speeds for the presented psychological state of the nine pairs' data, excluding those that could not be

TABLE II: List of participants' information

Pair No.	Cyclist						Dirver		
	ID	Age	Frequency of driving		Average velocity [m/s]		ID	Age	Frequency of driving (Car)
			Car	Bicycle	Normal	Hurry			
1	A	46	Everyday	Once per week	2.67	3.02	J	32	Everyday
2	B	23	Once per week	Once per yaer	3.42	3.70	K	44	Everyday
3	C	35	Rarely drive	2 times per month	1.62	2.43	L	58	Everyday
4	D	47	5 times per week	Rarely ride	3.69	4.69	M	33	Everyday
5	E	51	Everyday	Once per week	4.51	5.68	N	37	Everyday
6	F	33	Everyday	Rarely ride	2.39	2.74	O	36	6 times per week
7	G	55	Everyday	Rarely ride	3.88	4.89	P	52	Everyday
8	H	50	Everyday	Few times per year	3.83	4.95	Q	36	3 times per week
9	I	21	Once per week	Once per month	4.01	5.09	R	21	Once per week

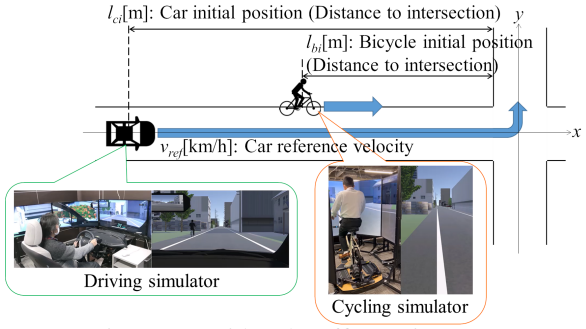


Fig. 4: Considered traffic environment

measured enough due to sensor errors or subjects' simulator sickness. The experiment participants included eight men and ten women, and the pairings were randomly assigned. Table II shows that the maximum difference in average speed among the subjects was approximately 3 [m/s], indicating a large difference in speed. In addition, experience as a driver may contribute to conservative behavior toward cars. However, all participants in the experiment had a driver's license, and all but one drove at least once a week. Therefore, the measured data did not reveal any differences in their behavior as cyclists based on their driving experience.

Next, Figure 5 shows the typical behaviors of cyclists from the measured data. In Fig. 5 (5a)-5a (5c), each graph shows, from the top, the coordinates of the bicycle and car on the x-axis (direction of travel), the bicycle speed, the sensor value of the strain gage attached to the crank portion of the pedals, and the amount of brake operation. Then, the light blue vertical line indicates the driver's blinker operation. The black vertical line indicates when the car overtook the bicycle. The A/D conversion values of each sensor are plotted for the sensor values of the strain gages and the amount of brake operation. For pedaling, a pedaling stop state is defined as a state in which the sensor value changes less than 3000 between samples for more than 1 second. For braking, a braking operation state is defined as a state in which the sensor value is 430 (the point at which the CS brake shoe makes contact with the rear wheel) or higher. These threshold values were determined based on actual measurement data. In Fig. 5, the pedaling stop state is shown in red, and the brake-on state is shown in purple. The figure on the right side of the graph roughly represents the changes in the relative positions of the bicycle and the car in the virtual space.

In Fig. 5 (5a), when the cyclist approaches the car about to make a left turn near an intersection, the cyclist stops pedaling and checks the car's movement while decelerating gently. Then, the cyclist passes through the intersection first by reading the driver's intention to yield from the car's deceleration and stop. In Fig. 5 (5b), the cyclist is a little behind the car when the car approaches the intersection, so he stops pedaling while the car is turning left to adjust the timing of his arrival at the intersection. In Fig. 5 (5c), the bicyclist is approaching the car near the intersection and shows his intention to yield by decelerating including braking earlier, and passes through the intersection after the car turns left.

The tendency of deceleration behavior was that braking was often performed when decelerating from proximity to a car. On the other hand, when decelerating at a short distance from a car, many cyclists decelerated gently by pedaling to a stop only. Cyclists E, H, and I, who were traveling faster than the other cyclists, often decelerated by braking. It is thought that the higher the speed, the more brake operations were performed to achieve higher deceleration. In most of the trials, the car turned left at the intersection first when the cyclist performed deceleration by braking. This can be interpreted as meaning that the cyclist's deceleration by braking was conveyed to the driver as an intention to yield.

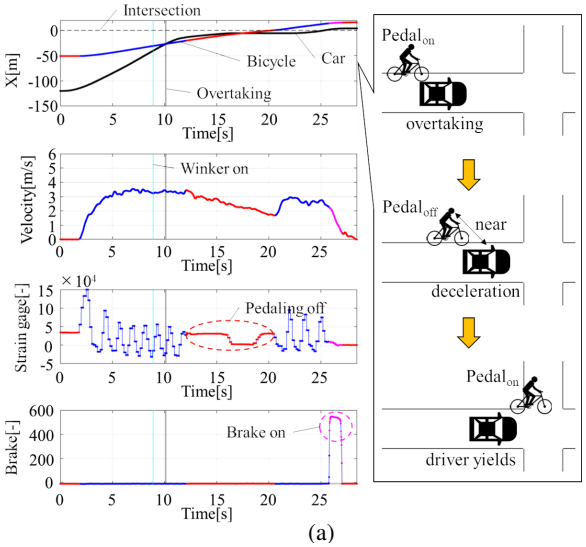
V. Construction of Cyclist Model

A. Logistic Regression Model

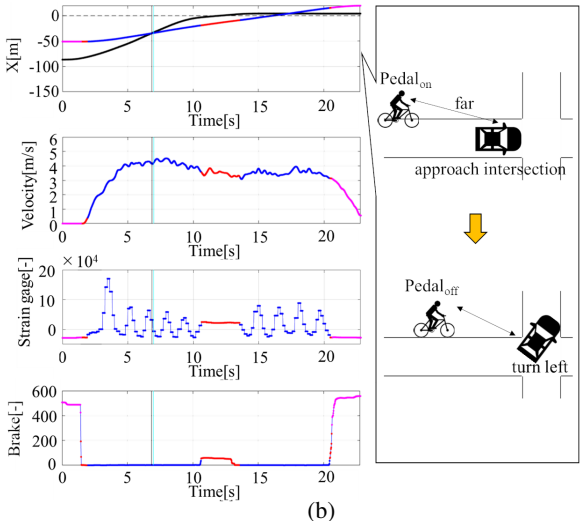
This section describes the modeling of cyclists' decision using logistic regression. The model was developed to estimate the probabilities of three target variables: "Pedal_{on}," "$\text{Pedal}_{\text{off}}$," and "$\text{Brake}_{\text{on}}$." The allocation of each decision state was performed according to the following conditions.

- $\text{Pedal}_{\text{off}}$
  - A segment that continuously satisfies the conditions of the following Eq. (1) for more than one second.

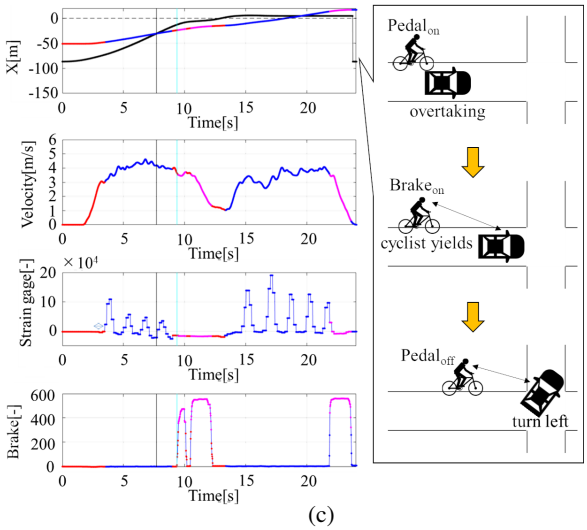
$$[|s(t)-s(t-1)| \leq s_{th}] \ \wedge \ [b(t) < b_{th}]. \tag{1}$$



(a)

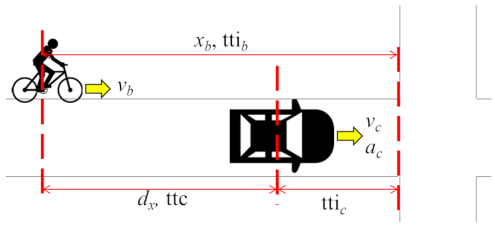


(b)



(c)

Fig. 5: Three typical interactions between cyclist and driver



d_x: x distance between car and bicycle [m]
x_b: x-coordinate of bicycle with intersection as origin[m]
v_b: bicycle velocity [m/s]
v_c: car velocity [m/s]
a_c: car acceleration [m/s^2]
ttc: time to collision[s]
tti$_b$: time to intersection[s] (bicycle)
tti$_c$: time to intersection[s] (car)

Fig. 6: Explanatory variables

- Brake$_{on}$
  - A segment that satisfies the conditions of Eq. (2).
  - A segment of less than one second that satisfies the conditions of Eq. (1) and is immediately before the segment that satisfies the conditions of Eq. (2).

$$b(t) \geq b_{th}. \tag{2}$$

- Pedal$_{on}$
  - A segment other than those corresponding to Pedal$_{off}$ or Brake$_{on}$.

Where, $s(t)$ represents the strain gauge sensor value at time t, s_{th} is the threshold for determining the pedaling state, $b(t)$ is the brake operation sensor value at time t, and b_{th} is the threshold for determining brake operation. The thresholds were set as $s_{th} = 3000$ and $b_{th} = 430$, and the cyclist's decision states were determined accordingly. Regarding the determination of Brake$_{on}$ it is assumed that when a cyclist applies the brake, they first stop pedaling. Therefore, a short period of pedaling cessation immediately before braking is considered preparatory behavior for braking and is included in the Brake$_{on}$ decision state. The probability of each decision $D \in (\text{Pedal}_{on}, \text{Pedal}_{off}, \text{Brake}_{on})$ at time t is described as follows:

$$P(D(t)|\boldsymbol{\phi}(t), \boldsymbol{\eta}) = \begin{cases} \dfrac{\exp(\boldsymbol{\eta}_{D(t)}^T \boldsymbol{\phi}(t))}{1 + \sum_{i \in I} \exp(\boldsymbol{\eta}_i^T \boldsymbol{\phi}(t))} & \text{if} \quad D(t) = \text{Pedal}_{on} \text{ or } \text{Pedal}_{off} \\ 1 - \sum_{i \in I} P(D(t) = i|\boldsymbol{\phi}(t), \boldsymbol{\eta}_i) & \text{otherwise,} \end{cases} \tag{3}$$

where, $I = \{\text{Pedal}_{on}, \text{Pedal}_{off}\}$ is the subset of cyclist's decisions, $\boldsymbol{\phi}(t)$ is the vector of explanatory variables, and $\boldsymbol{\eta}$ is the vector of regression coefficients. The explanatory variables used were those defined in Figure 6. By considering these explanatory variables, it becomes possible to explain the cyclist's interactive decision in response to the car and the environment. All variables were standardized for analysis. The probabilities of the decision states Pedalon, Pedaloff,

TABLE III: Logistic regression results

ID	A				B				C			
Decision state	Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
cnst	15.238	0.000	22.483	0.000	-47.941	0.000	-77.411	0.000	14.943	0.035	-79.800	0.000
d_x	-30.761	0.000	-24.722	0.000	-6.828	0.004	51.467	0.000	-64.179	0.000	136.190	0.000
d_x^2	39.332	0.000	30.548	0.000	169.950	0.000	9.720	0.213	221.220	0.000	-166.740	0.000
x_b	-22.587	0.000	-20.381	0.000	38.790	0.000	35.824	0.000	-13.915	0.022	36.071	0.000
x_b^2	-9.207	0.001	-17.120	0.000	37.470	0.000	28.011	0.000	30.672	0.000	11.125	0.016
v_c	6.675	0.000	3.372	0.000	9.163	0.000	5.417	0.000	28.517	0.000	15.924	0.000
a_c	5.506	0.000	3.652	0.000	1.639	0.004	0.816	0.135	-3.824	0.011	-3.786	0.060
ttc	-1.579	0.000	-1.745	0.000	-1.427	0.000	-1.907	0.000	-2.430	0.000	-4.684	0.000
tti_b	-1.383	0.630	1.289	0.643	16.194	0.000	17.665	0.000	-13.252	0.048	41.840	0.000
tti_c	2.792	0.000	0.208	0.346	17.076	0.000	13.949	0.000	6.959	0.000	6.385	0.000
v_b	10.251	0.000	0.774	0.372	18.291	0.000	5.015	0.001	-0.776	0.759	15.327	0.000

ID	D				E				F			
Decision state	Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
cnst	52.413	0.000	-59.415	0.000	-14.009	0.000	-2.558	0.307	-30.410	0.000	-32.374	0.000
d_x	-187.520	0.000	49.883	0.000	-24.881	0.000	4.731	0.063	-69.877	0.000	24.185	0.001
d_x^2	375.170	0.000	153.550	0.000	39.051	0.000	8.688	0.010	95.930	0.000	-18.333	0.066
x_b	-25.488	0.001	-5.325	0.566	8.823	0.000	0.913	0.725	37.111	0.000	26.327	0.000
x_b^2	2.356	0.739	54.881	0.000	16.647	0.000	-10.328	0.013	37.381	0.000	34.258	0.000
v_c	14.855	0.000	-90.969	0.000	9.652	0.000	5.413	0.000	1.419	0.146	-5.931	0.000
a_c	15.040	0.000	50.084	0.000	5.968	0.000	-0.285	0.568	2.558	0.022	13.099	0.000
ttc	0.148	0.873	5.487	0.000	2.442	0.000	2.250	0.000	-1.402	0.000	1.346	0.000
tti_b	-12.418	0.000	10.526	0.000	-4.187	0.000	1.555	0.000	18.690	0.000	3.585	0.005
tti_c	9.530	0.000	4.359	0.053	1.573	0.260	-3.947	0.000	-0.955	0.051	-3.692	0.000
v_b	3.635	0.023	20.866	0.000	-0.995	0.260	-7.322	0.000	24.209	0.000	-3.542	0.006

ID	G				H				I			
Decision state	Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$		Pedal_{on}		$\text{Pedal}_{\text{off}}$	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
cnst	-2.615	0.374	4.078	0.175	8.536	0.000	14.503	0.000	-36.983	0.000	-35.708	0.000
d_x	-66.244	0.000	17.779	0.000	-19.522	0.000	1.715	0.362	-13.961	0.000	-0.089	0.953
d_x^2	215.330	0.000	52.885	0.000	62.568	0.000	30.781	0.000	28.622	0.000	12.494	0.000
x_b	5.062	0.053	-14.885	0.000	-15.639	0.000	-20.611	0.000	43.835	0.000	34.407	0.000
x_b^2	17.447	0.000	-27.797	0.000	-4.306	0.002	-11.419	0.000	68.410	0.000	44.791	0.000
v_c	9.642	0.000	4.740	0.000	0.148	0.672	-0.834	0.020	6.884	0.000	6.560	0.000
a_c	3.920	0.000	3.541	0.000	9.898	0.000	5.157	0.000	-0.147	0.664	4.563	0.000
ttc	1.564	0.000	1.804	0.000	0.015	0.930	-0.036	0.810	-1.052	0.000	0.375	0.047
tti_b	-3.387	0.096	1.369	0.469	-1.111	0.000	1.169	0.000	0.042	0.945	4.888	0.000
tti_c	49.663	0.000	66.669	0.000	-22.461	0.000	-31.299	0.000	-24.337	0.000	-1.057	0.000
v_b	-0.200	0.815	-1.392	0.077	0.023	0.960	-3.012	0.000	-4.275	0.000	-3.893	0.000

and Brakeon were calculated by eq.3. Then, the state with the highest probability was adopted as the cyclist's decision. For parameter learning of the logistic regression model, four-fifths of the data were used for training, while the remaining one-fifth was used for model evaluation.

Table III shows the regression coefficients for each explanatory variable and the corresponding p-values obtained from the logistic regression analysis conducted on the data from Cyclists A to I. From Table III, comparing the magnitudes of the regression coefficients among the explanatory variables, it can be observed that the coefficients for d_x (the distance between the bicycle and a car) and x_b (the x-coordinate of the bicycle relative to the intersection) tend to be larger, either both or at least one of them.

B. Gaussian Mixture Model

This section describes the modeling of cyclists' decision using GMMs. In the modeling using GMMs, a GMM was fitted to each cyclist's decision state. The probability density function of the GMM is described as follows:

$$P(D(t)|\boldsymbol{x}(t),\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{k=1}^{K} m_k \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_{D(t),k}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}(t)-\boldsymbol{\mu}_{D(t),k})^T\boldsymbol{\Sigma}_{D(t),k}^{-1}(\boldsymbol{x}(t)-\boldsymbol{\mu}_{D(t),k})\right), \quad (4)$$

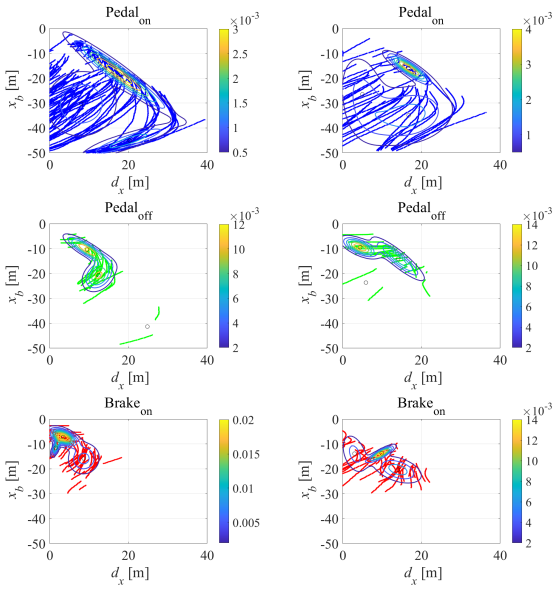
where $\boldsymbol{x}(t)$ is the random variable vector at time t, n is the number of random variables, $\boldsymbol{\mu}$ is the mean vector of the normal distribution estimated by fitting, $\boldsymbol{\Sigma}$ is the covariance matrix, m is the mixture ratio, and K is the number of mixtures.

The random variables were the same as the explanatory variables used in the LRM. The number of mixtures, one of the GMM parameters, was set to 3 by trial and error. The parameters estimated by GMM fitting about two of nine subjects (cyclist H and I) are shown in Table IV as representative examples. For reference, the distribution of each decision state for d_x and x_b, for which the regression coefficients in the LRM were large for cyclists H and I, and

TABLE IV: Estimated GMM parameters

ID	Decision state	Mixture ratio		d_x	d_x^2	x_b	x_b^2	v_c	a_c	ttc	tti_b	tti_c	v_b
H	Pedal_{on}	0.323	Mean	0.460	0.250	0.667	0.162	0.288	0.504	0.455	0.212	0.012	0.573
			Variance	0.049	0.046	0.044	0.030	0.014	0.030	0.174	0.048	0.010	0.072
		0.165	Mean	0.511	0.302	0.127	0.782	0.646	0.526	0.000	0.561	0.123	0.498
			Variance	0.050	0.052	0.020	0.037	0.032	0.052	0.010	0.087	0.024	0.061
		0.512	Mean	0.220	0.075	0.441	0.366	0.665	0.435	0.000	0.225	0.088	0.710
			Variance	0.037	0.018	0.044	0.052	0.030	0.019	0.010	0.019	0.012	0.022
	$\text{Pedal}_{\text{off}}$	0.170	Mean	0.154	0.025	0.904	0.018	0.238	0.636	0.000	0.821	0.027	0.041
			Variance	0.011	0.010	0.011	0.010	0.019	0.020	0.010	0.083	0.011	0.013
		0.562	Mean	0.302	0.094	0.656	0.149	0.483	0.382	0.000	0.160	0.038	0.599
			Variance	0.013	0.011	0.021	0.015	0.032	0.032	0.010	0.011	0.011	0.022
		0.269	Mean	0.371	0.152	0.673	0.165	0.311	0.446	0.643	0.153	0.026	0.601
			Variance	0.024	0.023	0.052	0.055	0.043	0.025	0.120	0.019	0.011	0.018
	Brake_{on}	0.108	Mean	0.177	0.040	0.833	0.046	0.164	0.376	0.576	0.234	0.336	0.408
			Variance	0.018	0.011	0.015	0.011	0.026	0.028	0.156	0.087	0.194	0.061
		0.247	Mean	0.078	0.008	0.922	0.014	0.164	0.571	0.000	0.993	0.104	0.003
			Variance	0.012	0.010	0.012	0.010	0.020	0.020	0.010	0.012	0.024	0.010
		0.644	Mean	0.156	0.032	0.779	0.074	0.398	0.326	0.019	0.130	0.056	0.550
			Variance	0.018	0.011	0.020	0.013	0.050	0.028	0.014	0.014	0.011	0.052

ID	Decision state	Mixture ratio		d_x	d_x^2	x_b	x_b^2	v_c	a_c	ttc	tti_b	tti_c	v_b
I	Pedal_{on}	0.239	Mean	0.552	0.290	0.554	0.231	0.434	0.486	0.000	0.257	0.028	0.560
			Variance	0.043	0.047	0.043	0.038	0.036	0.048	0.010	0.022	0.011	0.058
		0.682	Mean	0.368	0.119	0.317	0.505	0.676	0.529	0.000	0.284	0.095	0.711
			Variance	0.037	0.025	0.049	0.073	0.026	0.020	0.010	0.021	0.013	0.020
		0.079	Mean	0.653	0.392	0.491	0.270	0.233	0.416	0.656	0.200	0.037	0.744
			Variance	0.025	0.033	0.022	0.020	0.016	0.037	0.086	0.012	0.012	0.034
	$\text{Pedal}_{\text{off}}$	0.121	Mean	0.424	0.166	0.530	0.227	0.449	0.506	0.369	0.196	0.056	0.701
			Variance	0.041	0.028	0.017	0.016	0.068	0.015	0.156	0.012	0.011	0.024
		0.345	Mean	0.202	0.017	0.810	0.038	0.138	0.654	0.021	0.996	0.269	0.015
			Variance	0.014	0.011	0.012	0.010	0.023	0.023	0.023	0.011	0.102	0.011
		0.534	Mean	0.405	0.124	0.748	0.069	0.405	0.707	0.000	0.648	0.011	0.155
			Variance	0.020	0.016	0.016	0.012	0.023	0.025	0.010	0.117	0.010	0.034
	Brake_{on}	0.227	Mean	0.468	0.178	0.592	0.173	0.215	0.461	0.753	0.203	0.071	0.589
			Variance	0.022	0.018	0.017	0.014	0.013	0.029	0.079	0.011	0.012	0.028
		0.683	Mean	0.330	0.082	0.650	0.131	0.419	0.415	0.004	0.208	0.057	0.535
			Variance	0.026	0.015	0.019	0.014	0.050	0.043	0.011	0.016	0.012	0.046
		0.091	Mean	0.187	0.015	0.795	0.045	0.062	0.514	0.100	0.634	0.623	0.128
			Variance	0.016	0.011	0.013	0.010	0.016	0.025	0.060	0.129	0.165	0.028



(a) cyclist H (b) cyclist I

Fig. 7: Data distribution of each decision state and contour of GMM for cyclist H and I

the contours of the probability density function obtained by GMM fitting are shown in Figure 7. The probability of each decision state is calculated using Eq. (4), and the one with the highest probability is adopted as the result of estimating the cyclist's decision.

VI. Evaluation

As an evaluation of the constructed models, the match rate between the estimated decision states of the cyclist and the actual states in the evaluation data was calculated and compared. The match rate a between the estimated decision state $\hat{Y}$ and the actual state Y is defined as follows.

$$a = \frac{\sum_{l=1}^{L} [\hat{Y}_l = Y_l]}{L}, \tag{5}$$

where, L is the total number of samples in the evaluation data, l is the sample number in the evaluation data.

The match rates for both LRM and GMM are presented in Table V. From Table V, it can be observed that LRM has a slightly higher overall match rate. However, when comparing the match rates for each decision state, the accuracy for $\text{Pedal}_{\text{off}}$ is similar between the two models. For Brake_{on}, GMM shows a slightly higher match rate, while LRM performs better for Pedal_{on}. Notably, for Cyclists A, B, and F, the estimation accuracy of LRM for the Brake_{on} state is reduced due to the small number of data samples related to braking operations. In contrast, GMM achieves higher accuracy for these cases. As shown in Fig. 7, the data for the Brake_{on} state are concentrated in a specific region, allowing GMM to represent the distribution more effectively,

resulting in a higher matchi rate. On the other hand, the Pedal_{on} state data are widely and evenly distributed. This type of distribution is more challenging for GMM to model accurately, leading to a higher match rate for LRM. Based on the comparison of match rates, LRM has slightly higher overall accuracy. However, it was confirmed that GMM can achieve performance comparable to or even better than LRM for data with distribution characteristics that can be effectively represented by GMM, such as the Brake_{on} state.

TABLE V: Accuracy of cyclist decision estimation

Cyclist ID	LRM			
	Pedal_{on}	$\text{Pedal}_{\text{off}}$	Brake_{on}	Total
A	0.79	0.89	0.00	0.82
B	0.85	0.85	0.36	0.78
C	0.93	0.78	0.91	0.88
D	0.88	1.00	1.00	0.98
E	0.76	0.49	0.92	0.81
F	0.92	0.88	0.49	0.83
G	0.83	0.91	0.62	0.85
H	0.68	0.69	0.89	0.78
I	0.78	0.80	0.90	0.84
Average	0.82	0.81	0.68	0.84

Cyclist ID	GMM			
	Pedal_{on}	$\text{Pedal}_{\text{off}}$	Brake_{on}	Total
A	0.69	0.45	0.97	0.58
B	0.74	0.46	0.96	0.65
C	0.92	1.00	0.62	0.81
D	0.79	1.00	0.82	0.84
E	0.80	0.89	0.76	0.80
F	0.74	0.72	0.95	0.77
G	0.71	0.83	0.87	0.78
H	0.72	0.90	0.60	0.71
I	0.81	0.86	0.81	0.82
Average	0.77	0.79	0.82	0.75

VII. CONCLUSIONS

To model cyclists' decision, the cycling behavior of cyclists for a left-turning car at an unsignalized intersection was measured using the cycling simulator. Based on the measured data, models to estimate three decision states: Pedal_{on}, $\text{Pedal}_{\text{off}}$, and Brake_{on}, were developed using LRM and GMM, respectively. The comparison of the match rates of the constructed models revealed that GMM provided higher estimation accuracy for states like Brake_{on}, where the data distribution is concentrated in specific regions. Then, LRM showed higher accuracy for states like Pedal_{on}, where the data are distributed broadly and evenly. From these results, it is expected that constructing a model that combines the advantages of both the LRM and GMM will further improve the accuracy of estimating cyclists' decision states. As another modeling approach, the modeling using neural network models with deep learning will be also considered in future work. Neural network models are expected to achieve the highest estimation accuracy if sufficient training data is available. Furthermore, by using an autoencoder for feature extraction and employing the extracted features as explanatory variables in the LRM or GMM, it is anticipated that a simpler model with reduced input dimensionality can be realized while maintaining the estimation accuracy of the model.

REFERENCES

[1] D. Helbing and P. Molnár,"Social force model for pedestrian dynamics," Phys. Rev. E, vol. 51, no. 5, pp. 4282–4286, May 1995, doi: 10.1103/PhysRevE.51.4282.

[2] X. Chen, M. Treiber, V. Kanagaraj, and H. Li, "Social force models for pedestrian traffic - state of the art," Transport Reviews, vol. 38, no. 5, pp. 625–653, Sep. 2018, doi: 10.1080/01441647.2017.1396265.

[3] Y. Hashimoto, Y. Gu, L.-T. Hsu, M. Iryo-Asano, and S. Kamijo, "A probabilistic model of pedestrian crossing behavior at signalized intersections for connected vehicles," Transportation Research Part C: Emerging Technologies, vol. 71, pp. 164–181, Oct. 2016, doi: 10.1016/j.trc.2016.07.011.

[4] T. Nishimoto, T. Yamaguchi, H. Okuda, T. Suzuki, R. Wakisaka, and K. Ban, "Bayesian Network Model of Decision-Making for Pedestrians' Crossing Behavior Considering Gaze Information at Unsignalized Intersection," 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), pp.4408–4415, 2023, doi: 10.1109/ITSC57777.2023.10421888.

[5] Y. Han, Q. Chao, and X. Jin, "A simplified force model for mixed traffic simulation," Computer Animation and Virtual Worlds, vol. 32, no. 1, e1974, 12 pages, 2021, doi: 10.1002/cav.1974.

[6] X. Ma and D. Luo, "Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 40, pp. 130–144, Jul. 2016, doi: 10.1016/j.trf.2016.04.009.

[7] L. Bai and N. N. Sze, "Red light running behavior of bicyclists in urban area: Effects of bicycle type and bicycle group size," Travel Behaviour and Society, vol. 21, pp. 226–234, Oct. 2020, doi: 10.1016/j.tbs.2020.07.003.

[8] G. P. Castro, F. Johansson, and J. Olstam, "How to Model the Effect of Gradient on Bicycle Traffic in Microscopic Traffic Simulation," Transportation Research Record, vol. 2676, no. 11, pp. 609–620, Nov. 2022, doi: 10.1177/03611981221094300.

[9] D. P. Cobb, H. Jashami, and D. S. Hurwitz, "Bicyclists' behavioral and physiological responses to varying roadway conditions and bicycle infrastructure," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 80, pp. 172–188, Jul. 2021, doi: 10.1016/j.trf.2021.04.004.

[10] M. Nazemi, M. A. B. van Eggermond, A. Erath, D. Schaffner, M. Joos, and K. W. Axhausen, "Studying bicyclists' perceived level of safety using a bicycle simulator combined with immersive virtual reality," Accident Analysis & Prevention, vol. 151, 105943, 11 pages, Mar. 2021, doi: 10.1016/j.aap.2020.105943.

[11] X. Guo, E. Robartes, A. Angulo, T. D. Chen, and A. Heydarian, "Benchmarking the Use of Immersive Virtual Bike Simulators for Understanding Cyclist Behaviors," pp. 1319–1326, May 2022, doi: 10.1061/9780784483893.161.

[12] J. Haasnoot, R. Happee, V. van der Wijk, and A. L. Schwab, "Validation of a novel bicycle simulator with realistic lateral and roll motion," Vehicle System Dynamics, vol. 62, no. 7, pp. 1802–1826, Jul. 2024, doi: 10.1080/00423114.2023.2264418.

[13] M. M. Shoman and H. Imine, "Bicycle Simulator Improvement and Validation," IEEE Access, vol. 9, pp. 55063–55076, 2021, doi: 10.1109/ACCESS.2021.3071214.

[14] R. Alsaleh and T. Sayed, "Modeling pedestrian-cyclist interactions in shared space using inverse reinforcement learning," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 70, pp. 37–57, Apr. 2020, doi: 10.1016/j.trf.2020.02.007.

[15] X. Li, S. A. Useche, Y. Zhang, Y. Wang, O. Oviedo-Trespalacios, and N. Haworth, "Comparing the cycling behaviours of Australian, Chinese and Colombian cyclists using a behavioural questionnaire paradigm," Accident Analysis & Prevention, vol. 164, 106471, 12 pages, Jan. 2022, doi: 10.1016/j.aap.2021.106471.

[16] Y. Li, Y. Ni, J. Sun, and Z. Ma, "Modeling the illegal lane-changing behavior of bicycles on road segments: Considering lane-changing categories and bicycle heterogeneity," Physica A: Statistical Mechanics and its Applications, vol. 541, 123302, 15 pages, Mar. 2020, doi: 10.1016/j.physa.2019.123302.