To appear, 2018 Conference on Decision and Control (CDC)
https://www.cds.caltech.edu/~murray/preprints/gm18-cdc_s.pdf

Voluntary lane-change policy synthesis with reactive control improvisation

Jin I. Ge and Richard M. Murray

***Abstract*— In this paper, we propose reactive control improvisation to synthesize voluntary lane-change policy that meets human preferences under given traffic environments. We first train Markov models to describe traffic patterns and the motion of vehicles responding to such patterns using traffic data. The trained parameters are calibrated using control improvisation to ensure the traffic scenario assumptions are satisfied. Based on the traffic pattern, vehicle response models, and Bayesian switching rules, the lane-change environment for an automated vehicle is modeled as a Markov decision process. Based on human lane-change behaviors, we train a voluntary lane-change policy using explicit-duration Markov decision process. Parameters in the lane-change policy are calibrated through reactive control improvisation to allow an automated car to pursue faster speed while maintaining desired frequency of lane-change maneuvers in various traffic environments.**

I. Introduction

While automated vehicles have the potential to alleviate many traffic problems [1], in order to integrate well with the current road transportation system, it is necessary for an automated car to actively respond to other cars on road (which are largely human-driven) [2], [3]. Moreover, to be better accepted by users, an automated car may need to learn the preference of the human passengers on-board, and tune its behaviors accordingly.

Despite that adaptive cruise control and lane-keeping systems are relatively well researched [4], high-level decision-making is still an open question for automated cars. In particular, few research results exist regarding voluntarily lane-change policies that would allow the automated car to switch lanes for faster traffic flow [5]–[8]. It is important for both the overall traffic and the passengers on-board to enable an automated car to change a lane voluntarily in anticipation of slower traffic in front. It often creates less safety hazard and traffic perturbations to merge into a faster lane when the speed difference between two lanes is not yet too large [9]. Moreover, making voluntary lane-changes under similar conditions and with similar frequency a human driver would use is preferable.

To learn from human decision-making and build controllers that resemble human behaviors while meeting certain specifications, control improvisation has been shown to be a suitable method [10]. Randomized actions are generated that satisfy a set of hard constraints, while meeting a set of soft constraints in a probabilistic manner. While control improvisation has been used in composing music and controlling home appliances [11], [12], it is yet to be extended to problems with uncontrolled variables, where a controller is synthesized with respect to an environment. In particular, while lane-change maneuvers are decided by the automated vehicle, the speed and position of its neighboring vehicles cannot be directly controlled.

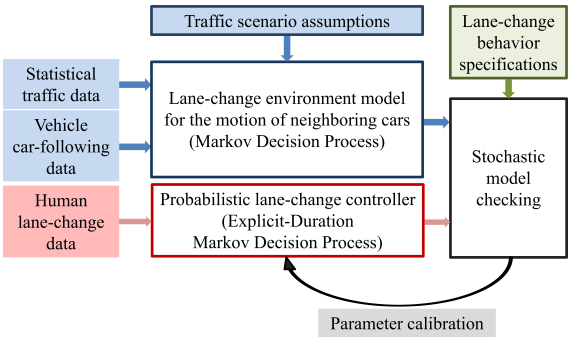

Fig. 1. Schematic for reactive control improvisation in voluntary lane-change synthesis.

Therefore, in this paper we integrate the notion of reactive synthesis and propose the method of reactive control improvisation. First, we train Markov models to describe a specified traffic pattern and the motion of vehicles responding to such patterns using traffic data. Using control improvisation, we ensure the trained models satisfy the given traffic environment assumptions. Then we train a voluntary lane-change policy as an explicit-duration Markov decision process based on human lane-change behavior. Finally, we formulate human preferences on voluntary lane-change behavior using reactive control improvisation. By calibrating parameters in the explicit-duration Markov decision process using stochastic model checking, a voluntary lane-change policy is synthesized that ensures safety and follows human preferences in a probabilistic manner; see Fig. 1.

II. Voluntary lane-change settings

Here we first set up the voluntary lane-change scenario on a two-lane road with symmetric lane-change rules, see Fig. 2. The ego vehicle (red) is in Lane 1, with its position and speed denoted s_e, v_e. The position and speed of its immediate predecessor (green) is denoted s_f, v_f. We project the longitudinal position of the ego car in Lane 2 (pink box). The position and speed of the projection's immediate leader (grey) are s_a, v_a, while the position and speed of two

*This work is supported by NSF VeHiCal project 1545126. The authors are with the Department of Computational and Mathematical Sciences, California Institute of Technology, Pasadena, California, 91125, USA. Email: jge@caltech.edu, murray@cds.caltech.edu

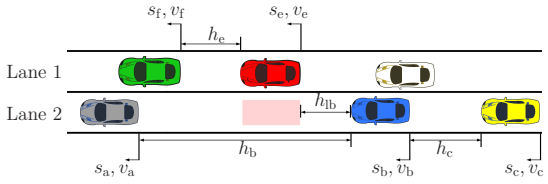

Fig. 2. The voluntary lane-changing scenario for two-lane roads.

cars behind the projection are denoted s_b, v_b (blue) and s_c, v_c (yellow), respectively.

A. Safety and incentive for voluntary lane change

Several models exist that describe human lane-changing behaviors. A shared feature of these models is that a lane-change maneuver is initiated if it is safe to change to the other lane lane, and the ego car has the incentive to do so. Here we describe the MOBILE (Minimizing Overall Braking Induced by Lane Changes) model, whose safety and incentive criteria are based on the acceleration of the ego car and its surrounding cars [13].

We assume that the acceleration a_e of the ego vehicle at time t is determined by its speed v_e, headway h_e, and its leader's speed v_f, that is:

$$a_e(t) = f_e\big(v_e(t), h_e(t), v_f(t)\big)\,, \tag{1}$$

where the headway $h_e = s_f - s_e - l$, l is the car length (assumed to be the same for all cars), and f_e can be various models as discussed in [16]. Similarly, the acceleration of the neighboring car B (blue) is

$$a_b(t) = f_b\big(v_b(t), h_b(t), v_a(t)\big)\,, \tag{2}$$

If the ego car were in Lane 2 (pink box), its acceleration and the acceleration of car B would be

$$\begin{aligned} \tilde{a}_e(t) &= f_e\big(v_e(t), h_b(t) - h_{lb}(t) - l, v_a(t)\big)\,, \\ \tilde{a}_b(t) &= f_b\big(v_b(t), h_{lb}(t), v_e(t)\big)\,. \end{aligned} \tag{3}$$

To ensure safety, if the ego vehicle merges onto Lane 2, it should not induce severe braking in vehicle B, that is,

$$\tilde{a}_b(t) > a^{\min}\,, \tag{4}$$

where $a^{\min}$ is the threshold for emergency braking. On the other hand, the incentive criterion requires certain 'overall benefit' for performing a lane-change at time t, i.e.,

$$\tilde{a}_e(t) - a_e(t) + p\big(\tilde{a}_b(t) - a_b(t)\big) > a^{\text{th}}\,, \tag{5}$$

where the politeness factor p and the benefit threshold a^{th} may differ among different drivers.

B. Limitations of safety and incentive criteria

While the lane-change model (4,5) seems to capture the human lane-change behavior well, it can generate an artificially high amount of lane changes in macroscopic traffic simulations [14]. In particular, if a vehicle would change its lane whenever the safety and incentive flags (4,5) are positive, a vehicle is likely to shift between two lanes with high frequency, leading to so-called 'ping-pong effect' [15].

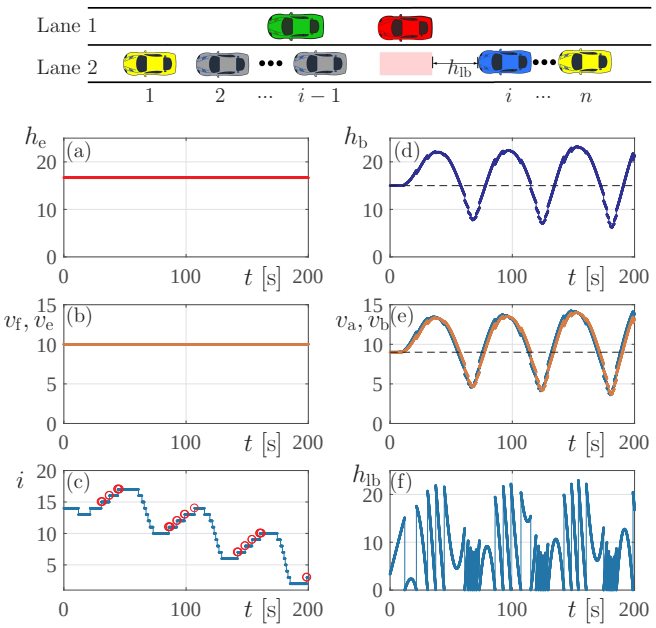

Fig. 3. An automated car (red) and its leader (green) travels on Lane 1, while cars $1,\ldots,n$ drives on Lane 2. (a) The bumper-to-bumper distance h_e between the automated car and its leader. (b) The speed of the automated car v_e and its leader v_f. (c) The curve shows the index i of the car in Lane 2 that immediately follows the projection of the automated car (pink box). The red circles indicate when the safety and incentive flags (4,5) are both positive. (d) The headway h_b between the current neighbors of the automated car in Lane 2. (e) The speeds v_a, v_b of the current neighbors of the automated car in Lane 2. (f) The distance h_{lb} between the automated car and its current neighbor vehicle i.

As a demonstration, we consider a two-lane scenario as shown in Fig. 3, where the ego car (red) in Lane 1 travels with constant speed $v_e = 10$ [m/s] while in Lane 2 cars $1,\ldots,n$ are led by car 1 with $v_1(t) = 9 + 6\sin(0.1t)$ [m/s]. In Fig. 3, panel (a) shows the headway h_e of the ego car (red), and panel (b) shows the speed v_e of the ego car and the speed v_f of its predecessor (green). The blue curve in panel (c) shows the index i of the car in Lane 2 in front of which the ego car is projected (pink box). Since the ego car maintains a higher average speed compared with cars in Lane 2, it keeps passing over cars in Lane 2 and the number i decreases from 14 to 2 in 200 seconds. Panel (d) shows the headway between the cars in Lane 2 which are the current neighbors of the ego car, i.e., $h_b = h_{i(t)}$. Correspondingly, panel (e) shows the speed $v_b = v_{i(t)}$ (orange curve) and $v_a = v_{i(t)-1}$ (blue curve). Note that while the speed and headway of cars in Lane 2 are continuous, the switchings in i introduce jumps in h_b and v_a, v_b. Similarly in panel (f), the distance h_{lb} between the ego car's projection and its current follower (blue) experiences jumps as the car number i changes.

While the ego car (red) travels with constant speed and passes over 12 cars in Lane 2, the red circles in Fig. 3(c) mark when the safety and incentive criteria (4,5) recommend voluntary lane changes to the ego vehicle. Here the safety

threshold $a^{\min} = -4$ [m/s^2], the incentive threshold $a^{\text{th}} = 2$ [m/s^2], and the politeness factor $p = 0.4$. By comparing panels (c,e), we see that the circles are given when the speed v_{a} in Lane 2 is temporarily higher than the ego car v_{e}. If the ego car did switch to Lane 2 and adopt the sinusoidal speed profile, the safety and incentive conditions (4,5) would prompt it to switch back to Lane 1 during the next lower half of the sinusoidal wave. While more restrictive thresholds in the safety and incentive criteria (4,5) can reduce the number of recommended lane changes in Fig. 3(c), overly restrictive thresholds can trap an automated car in a slower lane.

Through this simplified scenario, Fig. 3 demonstrates undesirable voluntary lane-change decisions that are solely based on safety and local incentives. An automated vehicle may eliminate overly frequent lane changes while responding to the attraction of faster lanes, if it incorporates in its decision-making statistical traffic information such as the average speed and speed variations of cars in front. Moreover, the performance of voluntary lane-change decisions needs to be evaluated stochastically, given the randomness in human-dominated traffic. Such stochastic criteria not only allows reactive synthesis of voluntary lane-change maneuvers responding to specific traffic situations, it can also learn from human lane-change behaviors, and may result in better acceptance from human passengers.

C. Control improvisation

Control improvisation is a framework to generate stochastic control sequences such that the control strategy obeys a set of hard constraints (conditions that are satisfied deterministically), a set of soft constraints (conditions that are satisfied with a probability), and certain randomness requirement on the richness of the controller behavior [10]. Specifically, given a finite alphabet Σ, we consider a randomized controller whose control sequence resides in the language $I \subset \Sigma^*$. Given a set $S \subset \Sigma^*$ for hard constraints, finitely many subsets $A_i \subset I$ with error probability bounds $\xi_i \in [0,1]$ for soft constraint $i \in \{1,\dots,n\}$, and a probability bound $\rho \in (0,1]$ for the randomness requirement, the distribution $D : \Sigma^* \to [0,1]$ with support set S such that

1) $S \subset I$,
2) $\forall i = 1,\dots,n, \quad P[\omega \in A_i | \omega \leftarrow D] \geq 1 - \xi_i$,
3) $\forall \omega \in S, \qquad D(\omega) \leq \rho$,

is called an (ξ, ρ)-improvising distribution. The hard constraint (1) is used to specify that behaviors with non-zero probabilities are within the safe set (e.g., no collisions on road); the soft constraint (2) is used to regulate the probability for certain sets of behaviors (e.g., the probability of consecutive lane-changes over 3 minutes apart is more than 90%); while the randomness requirement (3) ensures a low probability for any particular control sequence to be repeated.

III. Modeling in-lane traffic

In order to model the voluntary lane-change environment in two-lane traffic, we need to first describe vehicle motions in one lane. To model a specific level of traffic perturbations,

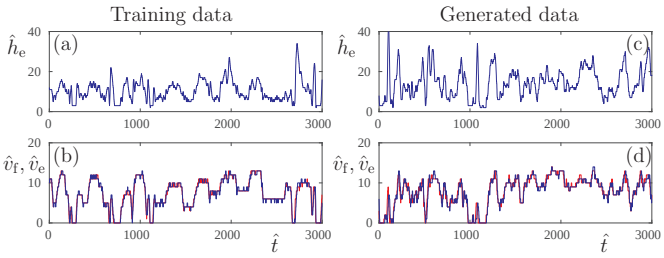

Fig. 4. (a,b) A 3000-step segment of training data used to model stop-and-go traffic. (c,d) A sample of traffic perturbations generated by the trained Markov model $\mathbf{M}_{\text{L}}$.

we train a Markov chain using corresponding speed and headway data. For vehicles responding to such a traffic pattern, we train a Markov decision process to model the car-following behaviors. Finally, we formulate the safety and volatility constraints of road traffic using control improvisation, and calibrate the Markov models such that their product can generate specified behaviors such as uniform traffic flow or stop-and-go waves.

A. Modeling traffic profiles with Markov chains

Assume that for a desired traffic pattern such as stop-and-go traffic waves, representative motion data can be collected from a segment of road. Without loss of generality, we consider Lane 1 in Fig. 2, and assume the speed v_{f} of the front vehicle and the speed v_{e} and headway h_{e} of the following vehicle are available through some types of vehicle-to-everything communication. As an example, Fig. 4(a,b) shows a segment of motion data $q(\hat{t}) = \big(\hat{h}_{\text{e}}(\hat{t}), \hat{v}_{\text{e}}(\hat{t}), \hat{v}_{\text{f}}(\hat{t})\big)$ for $\hat{t} = 1,\dots,\hat{T}$, which is representative in stop-and-go traffic. Here $\hat{h}_{\text{e}} = h_{\text{e}}/\eta_{\text{h}}$ is the quantized headway, $\hat{v}_{\text{e}} = v_{\text{e}}/\eta_{\text{v}}$ is the quantized speed, and $\hat{t} = t/\tau$ is the sampled time, with the quantization step $\eta_{\text{h}} = 2$ [m] and $\eta_{\text{v}} = 2$ [m/s], and the sampling period $\tau = 0.5$ [s].

For a specific traffic pattern, we define a discrete-time Markov chain

$$\mathbf{M}_{\text{L}} = \{\mathbb{X}_{\text{L}}, \pi_{\text{L}}, A_{\text{L}}, \mathbb{Y}_{\text{L}}, E_{\text{L}}\},$$

- the sample space $\mathbb{X}_{\text{L}} = \{s_m, m = 1,\dots,M_{\text{L}}\}$ where the state $s_m = (\hat{h}_{\text{e}}, \hat{v}_{\text{f}}, \hat{v}_{\text{e}})$,
- the initial distribution $\pi_{\text{L}} \in \mathbb{R}^{M_{\text{L}}}$,
- the transition matrix $A_{\text{L}} \in \mathbb{R}^{M_{\text{L}} \times M_{\text{L}}}$,
- the output space $\mathbb{Y}_{\text{L}} = \{o_k, k = 1,\dots,K\}$ where the output variable $o_k = \hat{v}_{\text{e}}$,
- the emission matrix $E_{\text{L}} \in \mathbb{R}^{M_{\text{L}} \times K}$.

While the emission matrix can be readily obtained through the projection of the state space $\mathbb{X}_{\text{L}}$ to the output space $\mathbb{Y}_{\text{L}}$, the transition probability

$$\big[A_{\text{L}}\big]_{ij} = P\big[q(\hat{t}+1) = s_j \big| q(\hat{t}) = s_i\big], \quad i,j = 1,\dots,M_{\text{L}}, \quad (6)$$

needs to be obtained from the traffic data $q(\hat{t})$, $\hat{t} = 1,\dots,\hat{T}$, using maximum likelihood estimation. Then the initial distribution π_{L} can be obtained as the stationary distribution with respect to the transition matrix A_{L}.

Fig. 4(c,d) shows an example of traffic perturbations generated by the Markov model trained with traffic data in panels (a,b). We note that the generated speed and headway profiles indeed display features of stop-and-go traffic.

We remark that the speed alone is insufficient to maintain Markov property for such traffic models. The car-following model (1) shows that a following car's acceleration depends on not only both cars' speeds but also the distance between them. Without the headway distance h_{e}, an extended history of v_{e} would influence v_{e} at the next time step.

B. Modeling vehicle response with Markov decision processes

Given the Markov chain describing traffic patterns in front, we propose a Markov decision process to describe the motion of vehicles responding to such traffic patterns. Without loss of generality, we assume in Fig. 2 vehicle A (grey) is the leader in Lane 2 and the motion of vehicles A and B (blue) can be described by the Markov model $\mathbf{M}_{\mathrm{L}}$. We consider the speed v_{c} and headway h_{c} of vehicle C (yellow) in response to the speed v_{b} of vehicle B.

For vehicle C responding to the motion of vehicles A and B modeled by $\mathbf{M}_{\mathrm{L}}$, we define a Markov decision process

$$\mathbf{M}_{\mathrm{r}} = \{\mathbb{X}_{\mathrm{r}}, \pi_{\mathrm{r}}, T_{\mathrm{r}}, \mathbb{U}_{\mathrm{r}}, \mathbb{Y}_{\mathrm{r}}, E_{\mathrm{r}}\} \tag{7}$$

- the sample space $\mathbb{X}_{\mathrm{r}} = \{s_m, m = 1, \ldots, M_{\mathrm{r}}\}$ where the state $s_m = (\hat{h}_{\mathrm{c}}, \hat{v}_{\mathrm{c}})$,
- the conditional initial distribution $\pi_{\mathrm{r}} \in \mathbb{R}^{M_{\mathrm{r}} \times K}$,
- the conditional transition matrix $T_{\mathrm{r}} \in \mathbb{R}^{M_{\mathrm{r}} \times M_{\mathrm{r}} \times K}$,
- the input space $\mathbb{U}_{\mathrm{r}} = \{u_k, k = 1, \ldots, K\}$ where the input variable $u_k = \hat{v}_{\mathrm{b}}$ is the predecessor's speed,
- the output space $\mathbb{Y}_{\mathrm{r}} = \{o_k, k = 1, \ldots, K\}$ where the output variable $o_k = \hat{v}_{\mathrm{c}}$ is the speed of the modeled vehicle,
- the emission matrix $E_{\mathrm{r}} \in \mathbb{R}^{M_{\mathrm{r}} \times K}$.

While the output map E_{r} can be obtained by projecting the sample space $\mathbb{X}_{\mathrm{r}}$ to the output space $\mathbb{Y}_{\mathrm{r}}$, the conditional transition matrix T_{r} needs to be obtained using a sequence of inputs $u(\hat{t}) \in \mathbb{U}_{\mathrm{r}}$ and states $q_{\mathrm{r}}(\hat{t}) \in \mathbb{X}_{\mathrm{r}}$ for $\hat{t} = 1, \ldots, \hat{T}$. In particular, the conditional transition probability for $s_i \to s_j$ given action u_k is

$$\left[T_{\mathrm{r}}\right]_{ijk} = P\left[q(\hat{t}+1) = s_j \middle| q(\hat{t}) = s_i, u(\hat{t}) = u_k\right], \tag{8}$$

for $i, j = 1, \ldots, M_{\mathrm{r}}$, $k = 1, \ldots, K$, and can be obtained using maximum likelihood estimation. And the initial distribution $\pi_{\mathrm{r}}(\cdot|u_k) \in \mathbb{R}^{M_{\mathrm{r}}}$ conditioned on the input u_k can be obtained through the stationary distribution of $T_{\mathrm{r}}(\cdot|u_k) \in \mathbb{R}^{M_{\mathrm{r}} \times M_{\mathrm{r}}}$.

As an example, we train a Markov decision process $\mathbf{M}_{\mathrm{r}}$ for vehicle C (yellow) in Fig. 2 and demonstrate the generated behavior in Fig. 5. Given the input $\hat{v}_{\mathrm{b}}$ (red curve in Fig. 5(b)), vehicle C (blue curve in Fig. 5(b)) is able to follow the speed profile well, and the headway $\hat{h}_{\mathrm{c}}$ (Fig. 5(a)) responds accordingly.

Note that training the vehicle response model $\mathbf{M}_{\mathrm{r}}$ may require more data than training the traffic profile model $\mathbf{M}_{\mathrm{L}}$. Nevertheless, when there are no sufficient data to span the product space $\mathbb{X}_{\mathrm{r}} \times \mathbb{U}$, car-following models [16] with bounded acceleration, speed, and headway can be used to obtain a priori transition probability matrix T_{r} for $\mathbf{M}_{\mathrm{r}}$.

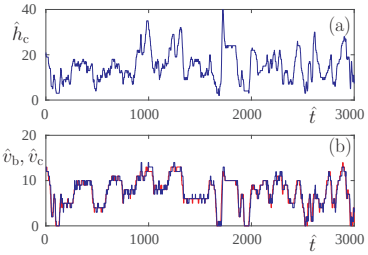

Fig. 5. A sample of the car-following behavior generated by the trained Markov decision process $\mathbf{M}_{\mathrm{r}}$. In the upper panel, the headway response of the vehicle is shown. In the lower panel, the input u is plotted in red, while the speed response is plotted in blue.

C. Calibrating in-lane traffic model

To ensure that the trained models $\mathbf{M}_{\mathrm{L}}$ and $\mathbf{M}_{\mathrm{r}}$ can generate diverse motion profiles that meet given traffic scenario assumptions, we consider N consecutive vehicles in one lane with the product model

$$\mathbf{M}_N = \mathbf{M}_{\mathrm{L}} \times (\mathbf{M}_{\mathrm{r}})^{N-2}, \tag{9}$$

and we formulate the traffic scenario assumptions in the framework of control improvisation:

1) Hard constraint

The product model $\mathbf{M}_N$ does not generate states that contain collisions or reversing cars, that is,

$$\begin{aligned} S &\subset \mathbb{X}_{\mathrm{L}} \times \mathbb{X}_{\mathrm{r}}^{N-2} \\ &= \{(\hat{v}_1, \hat{h}_i, \hat{v}_i) \mid \hat{v}_1 \geq 0,\, \hat{h}_i > 0,\, \hat{v}_i \geq 0,\, i = 2, \ldots, N\}, \end{aligned} \tag{10}$$

2) Soft constraints

The product model $\mathbf{M}_N$ generates a specified level of harshness in acceleration/deceleration maneuvers with high probability. That is, the one-step state variance

$$\Delta s_i(\hat{t}) = (\hat{h}_i(\hat{t}+1) - \hat{h}_i(\hat{t}), \hat{v}_i(\hat{t}+1) - \hat{v}_i(\hat{t})), \tag{11}$$

satisfies

$$P[\|\Delta s_i\| \leq \Delta^{\mathrm{a}}] \geq 1 - \xi^{\mathrm{a}}, \tag{12}$$

for $\hat{t} = 1, 2, \ldots$ where Δ^{a} is the volatility bound for headway and speed variations, and ξ^{a} is the corresponding concentration bound.

Moreover, headway and speed trajectories generated by $\mathbf{M}_N$ satisfy certain concentrations, that is,

$$\begin{aligned} P[|h_i - h^*| \leq \Delta^{\mathrm{h}}] &\geq 1 - \xi^{\mathrm{h}}, \\ P[|v_i - v^*| \leq \Delta^{\mathrm{v}}] &\geq 1 - \xi^{\mathrm{v}}, \end{aligned} \tag{13}$$

where Δ^{h} and Δ^{v} are the bounds for headway and speed variations, h^* and v^* are the mean headway and speed in given traffic scenario, while ξ^{h} and ξ^{v} are corresponding probability bounds.

3) Randomness

The randomness condition ensures richness in the generated motion of one-lane traffic. It requires that a particular

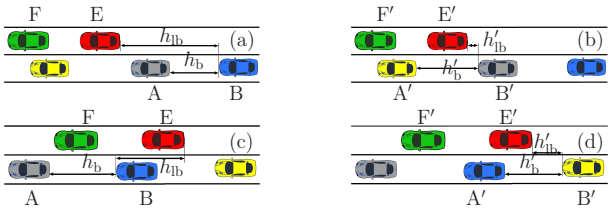

Fig. 6. (a,b) Configurations of neighboring cars during a pass-over scenario. (c,d) Configurations of neighboring cars during a fall-back scenario.

sequence of speed and headway profiles is related with very small probability. Since the product Markov model $\mathbf{M}_N$ is aperiodic and irreducible, this condition is automatically satisfied due to ergodicity.

Because the training data used in this paper are representative of the given stop-and-go traffic scenario, the trained models satisfy the hard constraint and the soft constraints, and are able to generate desired traffic profiles; see Fig. 4(c,d) and Fig. 5. However, if traffic scenarios with less representative data need to be modelled, further calibration is necessary after the transition matrices $A_{\rm L}$ and $T_{\rm r}$ are estimated from the data. For example, if the data contain irregular driving behaviors such as collisions, the support set of $\mathbf{M}_N$ may strictly contain the set S of hard constraint (10), then the states violating the hard constraint need to be pruned. On the other hand, if a traffic scenario with less speed and headway volatility than Fig. 4(a,b) is desired, $\mathbf{M}_N$ may not satisfy the soft constraints (12,13), then the transition probabilities can be tuned accordingly. In this way, the traffic scenario assumptions are enforced through control improvisation.

IV. Modeling Voluntary Lane-Change Environment

Based on the Markov models on the traffic pattern and vehicle responses, we may obtain a Markov model for the voluntary lane-change environment. Again we consider Fig. 2, and denote the lane the ego car is in as $\lambda \in \Lambda = \{1,2\}$. The projected headway $h_{\rm lb}$ satisfies

$$h_{\rm lb}(\hat{t}+1) = h_{\rm lb}(\hat{t}) + \tau\big(v_{\rm e}(\hat{t}) - v_{\rm b}(\hat{t})\big)\,. \tag{14}$$

While the right-hand side of equation (14) is smooth, $h_{\rm lb}$ experience jumps as the ego car passes by or falls behind vehicles in the other lane, resulting in jumps in $v_{\rm e}, v_{\rm b}$, see Fig. 3.

When the ego car falls behind vehicle B in the other lane, see Fig. 6(c), the neighboring vehicles A and B change

$$\begin{aligned} h_{\rm lb} \le -h^{\rm hys} \quad \Rightarrow \quad & v'_{\rm a} = v_{\rm b} \\ \wedge \quad & (h'_{\rm b}, v'_{\rm b}) \sim P_{\rm f}[\,\cdot\,|v'_{\rm a}] \\ \wedge \quad & h'_{\rm lb} = h'_{\rm b} + h_{\rm lb} + l\,, \end{aligned} \tag{15}$$

see Fig. 6(d), where $h^{\rm hys}$ is the hysteresis distance to avoid chattering in the switching. $P_{\rm f}[\,\cdot\,|v'_{\rm a}]$ is the distribution of the headway $h'_{\rm b}$ and speed $v'_{\rm b}$ of the newly emerged vehicle B′ conditioned on the speed $v'_{\rm a}$ of its predecessor

$$P_{\rm f}[(h'_{\rm b}, v'_{\rm b})|v'_{\rm a} = u_k] = \pi_{\rm r}\big(h'_{\rm b}, v'_{\rm b}|u_k\big)\,, \tag{16}$$

where $\pi_{\rm r}(\cdot|u_k)$ is the stationary distribution of the vehicle response model $\mathbf{M}_{\rm r}$ under input speed u_k for $k = 1,\ldots,K$. On the other hand, when the ego car passes over vehicle A in Fig. 6(a), the neighboring cars A and B have

$$\begin{aligned} h_{\rm lb} \ge h_{\rm b} + h^{\rm hys} \quad \Rightarrow \quad & h'_{\rm lb} = h_{\rm lb} - h_{\rm b} - l \\ \wedge \quad & v'_{\rm b} = v_{\rm a} \\ \wedge \quad & (v'_{\rm a}, h'_{\rm b}) \sim P_{\rm b}[\,\cdot\,|v'_{\rm b}]\,, \end{aligned} \tag{17}$$

see Fig. 6(b), where the probability distribution $P_{\rm b}[(v'_{\rm a}, h'_{\rm b})|v'_{\rm b}]$ for the newly emerged vehicle A′ is calculated through Bayesian rules such that

$$P_{\rm b}[(v'_{\rm a}, h'_{\rm b})|v'_{\rm b}] = \frac{P[v'_{\rm a}, h'_{\rm b}, v'_{\rm b}]}{P[v'_{\rm b}]} = \frac{P[h'_{\rm b}, v'_{\rm b}|v'_{\rm a}]P[v'_{\rm a}]}{\sum_{h'_{\rm b}}\sum_{v'_{\rm a}} P[h'_{\rm b}, v'_{\rm b}|v'_{\rm a}]P[v'_{\rm a}]}\,, \tag{18}$$

where the stationary distribution

$$\begin{aligned} P[h'_{\rm b}, v'_{\rm b}|v'_{\rm a}] &= \pi_{\rm r}(h'_{\rm b}, v'_{\rm b}|u_k), \\ P[v'_{\rm a}] &= \sum_{h'_{\rm b}}\sum_{v'_{\rm a}} \pi_{{\rm L}\lambda^{\rm c}}(v'_{\rm a}, h'_{\rm b}, v'_{\rm b})\,, \end{aligned} \tag{19}$$

based on the stationary distributions $\pi_{\rm r}$ of the response model $\mathbf{M}_{\rm r}$ and $\pi_{{\rm L}\lambda^{\rm c}}$ of the traffic pattern model $\mathbf{M}_{{\rm L}\lambda^{\rm c}}$ in the other lane $\lambda^{\rm c} \in \Lambda \oplus \{\lambda\}$. Here the 'exclusive or' operation $\oplus$ is used to select the lane without the ego car.

Using the switching conditions (15,17), we may also define a counter c on the number of cars in the adjacent lane that the ego car passes over:

$$\begin{aligned} h_{\rm lb} \ge h_{\rm b} + h^{\rm hys} \quad &\Rightarrow \quad c' = c+1\,, \\ h_{\rm lb} \le -h^{\rm hys} \quad &\Rightarrow \quad c' = c-1\,. \end{aligned} \tag{20}$$

We consider the set of lane-change decisions $\Phi = \{0,1\}$, where $\phi = 1$ indicates the decision to change a lane, then

$$\begin{aligned} \phi = 1 \quad \Rightarrow \quad & v'_{\rm e} = v_{\rm e} \\ \wedge \quad & v'_{\rm f} = v_{\rm a} \\ \wedge \quad & h'_{\rm e} = h_{\rm b} - h_{\rm lb} - l \\ \wedge \quad & v'_{\rm a} = v_{\rm f} \\ \wedge \quad & (h'_{\rm lb}, v'_{\rm b}) \sim P_{\rm f}[\,\cdot\,|v_{\rm e}] \\ \wedge \quad & h'_{\rm b} = h'_{\rm lb} + h_{\rm e} \\ \wedge \quad & \lambda' = \lambda^{\rm c} \\ \wedge \quad & c' = 0\,, \end{aligned} \tag{21}$$

where $P_{\rm f}[\,\cdot\,|v_{\rm e}]$ is given as in equation (16). Note that the counter c on the number of passing vehicles is reset as the ego car switches to a new lane.

Based on the motion $v_{\rm a}, h_{\rm b}, v_{\rm b}$ of neighboring cars in the adjacent lane, the motion $v_{\rm f}, h_{\rm e}, v_{\rm e}$ of the ego car and its leader, and the projected headway $h_{\rm lb}$ between two lanes, we can define the safety flag c_1 and the incentive flag c_2 as

$$\begin{aligned} c_1 &\models \big(\tilde{a}_{\rm b}(t) > a^{\rm min}\big)\,, \\ c_2 &\models \Big(\tilde{a}_{\rm e}(t) - a_{\rm e}(t) + p\big(\tilde{a}_{\rm b}(t) - a_{\rm b}(t)\big) > a^{\rm th}\Big)\,, \end{aligned} \tag{22}$$

which are used to indicate potential lane-change opportunities.

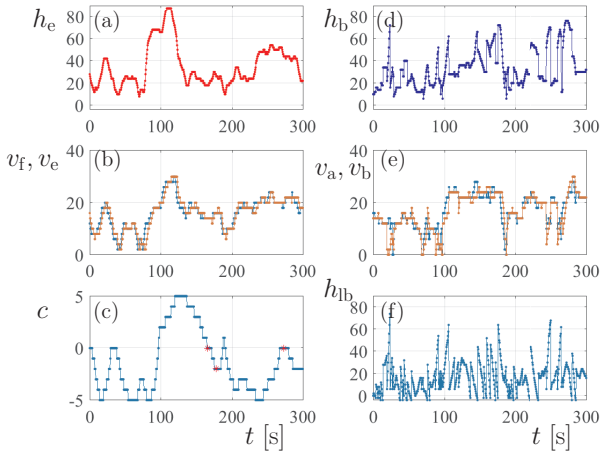

Fig. 7. (a) The bumper-to-bumper distance h_e between the ego car and its leader. (b) The speed of the automated car v_e and its leader v_f. (c) The number c of cars the ego car passes over. The red stars indicate instances when the safety and incentive flags (c_1, c_2) are true. (d) The headway h_b between the current neighbors of the ego car in the other lane. (e) The speeds v_a, v_b of the current neighbors of the ego car in the other lane. (f) The distance h_{lb} between the automated car and its current neighbor vehicle b.

Using equations (15 – 22), we can define a Markov decision process to describe the lane-change environment

$$\mathbf{M} = \{\mathbb{X}, \pi, T, \Phi, \mathbb{Y}, E\} \tag{23}$$

- the sample space $\mathbb{X} = \{s_m, m = 1, \ldots, M\}$ where the state $s_m = (\hat{v}_f, \hat{h}_e, \hat{v}_e, \hat{v}_a, \hat{h}_b, \hat{v}_b, \hat{h}_{lb}, \lambda, c)$,
- the initial distribution $\pi \in \mathbb{R}^M$,
- the input space Φ where the input variable $\phi = 0$ indicates no lane-change, while $\phi = 1$ indicates lane-change maneuver,
- the transition matrix $T \in \mathbb{R}^{M \times M \times |\Phi|}$, where $|\Phi|$ is the cardinality of the input space Φ,
- the output space $\mathbb{Y} = \{(c, c_1, c_2)\}$,
- the emission matrix $E \in \mathbb{R}^{M \times |\mathbb{Y}|}$, where $|\mathbb{Y}|$ is the cardinality of the output space $\mathbb{Y}$.

Note that the transition matrix T can be calculated based on the traffic model $\mathbf{M}_{L\lambda}$ for the ego lane, the traffic model $\mathbf{M}_{L\lambda_c}$ and response model $\mathbf{M}_r$ for the next lane, the projected headway model (14) and switching rules (15 – 19). And the emission matrix E can be calculated using (22).

As an example, we show the trajectory of a neighbor model $\mathbf{M}$ in Fig. 7, where the traffic patterns on both lanes are as shown in Fig. 4, and the vehicle response model is as shown in Fig. 5. Fig. 7(c) shows that c varies between 5 and -5 within 300 seconds, that is, while the ego car passes over up to 5 cars temporarily, it can fall back quickly. This is well expected as the traffic patterns on both lanes are the same. The safety flag c_1 and the incentive flag c_2 are satisfied three times during these 300 seconds, as shown by the red stars in Fig. 7(c). Different from in Fig. 3, the three potential lane-changes within realistic traffic pattern here are not evenly distributed in time ($t = 166.5, 178.5, 273.5$ [s]). Therefore,

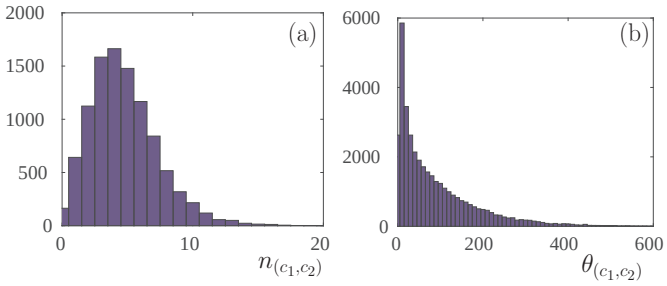

Fig. 8. (a) Distribution of the number of lane-change opportunities in 600 seconds, over 10000 runs. (b) Distribution of the time gap between two consecutive lane-change opportunities in 600 seconds, over 10000 runs.

consecutive lane changes can happen within short periods of time if the ego car acts on every lane-change opportunities.

Furthermore, Fig. 8 shows the corresponding distribution of the number of lane-change opportunities and the time gap between two consecutive lane-change opportunities in 10000 runs. In average, there are 4.83 lane-change opportunities in 600 seconds, while the time gap between two opportunities follows an exponential distribution (expected for a Markov model). If an automated car acts on every lane-change opportunity, it would perform a lane-change every 2 minutes in average, and it is highly likely to change lanes again within a short period of time. Such frequent lane-change behavior is not only potentially detrimental to the overall traffic, it also may not be appreciated by the passengers on-board.

While the ego car does not lose speed by avoiding voluntary lane-changes in this simulation, allowing voluntary lane-changes is still beneficial for most traffic conditions when traffic patterns on two lanes are not symmetric. Therefore, it is necessary to build a probabilistic lane-change policy to avoid frequent lane-changes while providing the human passengers on-board the satisfaction that the automated car travels no slower than the other cars on road.

V. Voluntary Lane-Change Controller Synthesis Using Reactive Control Improvisation

In this section, we propose a stochastic lane-change policy using explicit-duration Markov decision process. While this explicit-duration Markov decision process can be trained using human lane-change data, to ensure the performance of the lane-change controller in various traffic situations, we propose a reactive control improvisation framework to calibrate its parameters.

A. Explicit-duration Markov decision process

Here we propose an explicit-duration Markov decision process as a voluntary lane-change controller

$$\mathbf{C} = (\Phi, D, P_\phi, p_d, A_c) \tag{24}$$

- the sample space $\Phi = \{0, 1\}$ where $\phi = 1$ indicates initiating a lane-change maneuver,
- the duration set $D = \{1, \ldots, d^{\max}\}$, where $d^{\max}$ is the maximum duration for refusing lane-change chances.

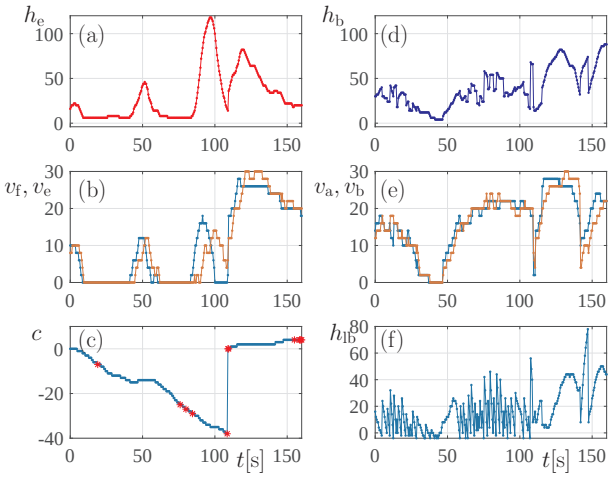

Fig. 9. (a) The bumper-to-bumper distance h_e between the ego car and its leader. (b) The speed of the automated car v_e and its leader v_f. (c) The number c of cars the ego car passes over. The red dots indicate when the safety and incentive flags are positive. (d) The headway h_b between the current neighbors of the ego car in the other lane. (e) The speeds v_a, v_b of the current neighbors of the ego car in the other lane. (f) The distance h_{lb} between the automated car and its current neighbor vehicle b.

– the transition matrix $P_\phi \in \mathbb{R}^{2\times 2}$ where

$$[P_\phi]_{ij} = P[\phi' = j | \phi = i, d = 1],$$

– the duration distribution $p_d \in \mathbf{R}^{d^{\max}\times 2}$, where

$$p_d(k) = P[\phi' = 0, d' = k | \phi = 1, d = 1], \quad k = 1, \ldots, d^{\max},$$

– the transition matrix $T \in \mathbb{R}^{M\times M\times|\Phi|}$, where $|\Phi| = 2$,
– the input space $A_c = \{(c, c_1, c_2)\}$ which is the same as the output space E of the lane-change environment.

Here the stochastic matrix P_ϕ is an identity matrix, since there are only two states in Φ and remaining in the same state is not allowed with explicit duration.

The action $a_c = (c, c_1, c_2)$ determines whether the Markov decision process is allowed to transit into $\phi' = 1$. That is,

$$\begin{aligned} &\neg(c < 0 \wedge c_1 = 1 \wedge c_2 = 1) \Rightarrow \phi' = 0, \\ &(c < 0 \wedge c_1 = 1 \wedge c_2 = 1) \Rightarrow (\phi', d') \sim P_{\phi,d}(\phi, d). \end{aligned} \tag{25}$$

When $\phi = 1$, d has Dirac distribution centered on 1, i.e., $p_d[k = 1 | \phi = 1] = 1$, ensuring no immediately consecutive lane-changes. However, the duration distribution $p_d[k|\phi = 0]$ can have the entire set D as its support, and need to be trained using human lane-change behaviors using maximum likelihood estimate methods presented in [17], [18].

To demonstrate the behavior of the proposed controller, we plot the behavior of the ego car and its neighboring vehicles in Fig. 9 in a 150-second simulation. In panel (c), the counter c jumps from around -40 to 0 at $t \approx 120$ [s], indicating a lane-change maneuver. While multiple lane-change opportunities exist during $t < 120$ [s], as marked by the red dots, in these cases the lane-change decision remains $\phi = 0$. This behavior is similar to human lane-change decisions in the sense that a human driver does not immediately switch to a faster adjacent lane, but rather wait for the 'cost' of remaining in the slower lane to accumulate and execute the lane-change maneuver after some time.

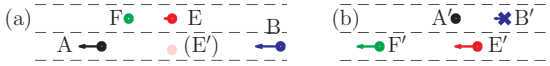

Fig. 10. (a) Configuration of neighboring cars before the lane-change maneuver. The pink dot (E′) is the projection of the ego car in the other lane. (b) Configuration after the lane-change maneuver, where car F′ is car A in (a), car A′ is car F in (a), and car B′ is newly emerged for the lane-change environment.

In Fig. 10 we show the configuration of the lane-change environment before and after the lane-change maneuver. In panel (a), the ego car (red) has very small speed in response to its leader (green) while cars in the next lane travel faster. Since there is a large gap between car A and car B, vehicle E is able to switch over, as shown in panel (b). Note that vehicle B′ in panel (b) is a newly emerged vehicle whose initial position and speed follows the Bayesian rule $P_f[\cdot|v_e]$.

B. Reactive control improvisation

In order to ensure desired lane-change behavior in the specific environment, we need to verify the product Markov model $\mathbf{M} \times \mathbf{C}$ satisfies safety constraints all the time and meets human preferences probabilistically. Therefore, we formulate the safety and human preference constraints in the framework of control improvisation:

1) Hard constraint

A voluntary lane-change only happens when it is safe and locally motivated for pursuing higher speed across the lanes:

$$c_1 = 0 \wedge c_2 = 0 \Rightarrow \phi' = 0. \tag{26}$$

Note that this constraint is satisfied by the safety and incentive criteria (25).

2) Soft constraints

The human preference on the frequency of lane-change maneuvers can be written as

$$\begin{aligned} P[|n_\phi - N_\phi| < \Delta^N] > 1 - \xi_N, \\ P[|\theta_\phi - \Theta_\phi| < \Delta^\theta] > 1 - \xi_\theta, \end{aligned} \tag{27}$$

where n_ϕ is the number of lane changes in a fixed amount of time, N_ϕ is the number of lane changes suitable for the particular traffic patterns, θ_ϕ is the time gap between two consecutive lane changes, Θ_ϕ is the average time gap human passengers prefer. The bounds Δ^N, Δ^θ and ξ_N, ξ_θ control the concentration of the lane-change number n_ϕ and time gap θ_ϕ.

3) Randomness

The randomness requirement specifies that the probability of generating a particular decision sequence $\{\phi(\hat{t})\}$ is low. Here this condition is also met by the construction of the controller $\mathbf{C}$.

Then we check that the production $\mathbf{M} \times \mathbf{C}$ satisfies the constraints (26,27), i.e. $\mathbf{M} \times \mathbf{C}$ is a (ξ, ρ)-improvising distribution conditioned on the assumptions (10,12,13) on $\mathbf{M}$. Note that while tuning the trained duration distribution p_d directly

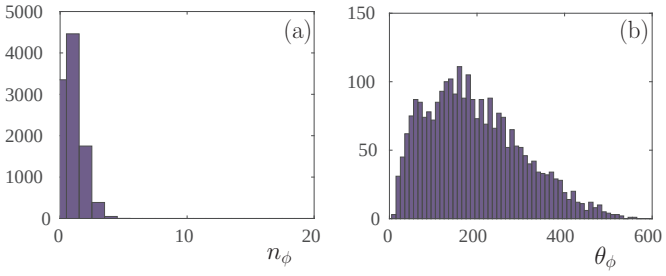

Fig. 11. (a) Distribution of the number of lane-change maneuvers in 600 seconds, over 10000 runs. (b) Distribution of the time gap between two consecutive lane-change maneuvers in 600 seconds, over 10000 runs.

influences the distribution of the number n_ϕ and the time gap θ_ϕ of lane changes, the safety and incentive parameters $a^{\min}$ and a^{th}, p can also be tuned to satisfy the soft constraint (27) through the influence of $\mathbf{M}$.

To demonstrate the distribution of n_ϕ and θ_ϕ in the soft constraint (27), we plot their distributions in Fig. 11 using the same traffic pattern setting as in Fig. 8. In panel (a), the average number of lane-change maneuvers in 600 seconds is 0.93 among 10000 simulations. In panel (b), the average time gap between two consecutive lane changes is 195 seconds. By comparing Fig. 11 and Fig. 8, we see that the proposed lane-change controller is able to avoid frequent lane changes while allowing an automated car to pursue speed gains in another lane.

VI. Conclusion

In this paper, we synthesized randomized decisions for voluntary lane changes that would meet human preferences under given traffic environment. The traffic environment was described using Markov models trained by traffic data. Parameters in the environment model were calibrated through control improvisation to be stochastically representative of the specified traffic patterns. An explicit-duration Markov decision process was proposed as the voluntary lane-change controller, whose parameters were trained using human lane-change data. Then, reactive control improvisation was formulated to ensure that the lane-change controller was tuned to satisfy human preferences stochastically under given traffic environment.

Acknowledgement

The authors would like to thank the support from NSF CPS Frontier project 1545126, and thank the insightful discussions with Prof. Sanjit Seshia and Prof. Gabor Orosz.

References

[1] R. Stern, S. Cui, M. L. D. Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, H. Pohlmanng, F. Wu, B. Piccolih, B. Seiboldb, J. Sprinkle, and D. B. Work, "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation research Part C: Emerging technologies*, vol. in press, 2018.

[2] D. Sadigh, K. Driggs Campbell, A. A. A. Puggelli, W. Li, V. Shia, R. Bajcsy, A. L. Sangiovanni-Vincentelli, S. S. Sastry, and S. A. Seshia, "Data-driven probabilistic modeling and verification of human driver behavior," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2013-197, Dec 2013.

[3] K. Driggs-Campbell, R. Dong, S. S. Sastry, and R. Bajcsy, "Informative human-in-the-loop predictions via empirical reachable sets," *Transactions on Intelligent Vehicles*, vol. in press, 2018.

[4] M. Aeberhard, S. Rauch, M. Bahram, G. Tanzmeister, J. Thomas, Y. Pilat, F. Homm, W. Huber, and N. Kaempchen, "Experience, results and lessons learned from automated driving on Germany's highways," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 42–57, 2015.

[5] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "Mpdm: Multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1670–1677.

[6] S. Ulbrich and M. Maurer, "Towards tactical lane change behavior planning for automated vehicles," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 989–995.

[7] M. Wang, S. P. Hoogendoorn, W. Daamen, B. van Arem, and R. Happee, "Game theoretic approach for predictive lane-changing and car-following control," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 73 – 92, 2015.

[8] Y. Du, Y. Wang, and C. Y. Chan, "Autonomous lane-change controller via mixed logical dynamical," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014, pp. 1154–1159.

[9] H. Jula, E. B. Kosmatopoulos, and P. A. Ioannou, "Collision avoidance analysis for lane changing and merging," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 6, pp. 2295–2308, 2000.

[10] D. J. Fremont, A. Donzé, S. A. Seshia, and D. Wessel, "Control improvisation," in *35th IARCS Annual Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS)*, ser. LIPIcs, vol. 45, December 2015, pp. 463–474.

[11] I. Akkaya, D. Fremont, R. Valle, A. Donzé, E. A. Lee, and S. A. Seshia, "Control improvisation for probabilistic temporal specifications," in *Proceedings of the 1st IEEE International Conference on Internet-of-Things Design and Implementation (IoTDI)*, April 2016, pp. 55–70.

[12] A. Donzé, R. Valle, I. Akkaya, S. Libkind, S. A. Seshia, and D. Wessel, "Machine improvisation with formal specifications," in *Proceedings of the 40th International Computer Music Conference (ICMC)*, September 2014.

[13] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model mobil for car-following models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, pp. 86–94, 2007.

[14] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, "A realistic two-lane traffic model for highway traffic," *Journal of Physics A: Mathematical and General*, vol. 35, no. 15, pp. 3369–3388, 2002.

[15] K. Nagel, D. E. Wolf, P. Wagner, and P. Simon, "Two-lane traffic rules for cellular automata: A systematic approach," *Physical Review E*, vol. 58, no. 2, pp. 1425–1437, 1998.

[16] D. Chowdhury, L. Santen, and A. Schadschneider, "Statistical physics of vehicular traffic and some related systems," *Physics Reports*, vol. 329, pp. 199–329, 2000.

[17] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[18] S.-Z. Yu and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1947–1951, 2006.