The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Symbolic Modeling of Driving Behavior based on Hierarchical Segmentation and Formal Grammar

Ato Nakano, Hiroyuki Okuda, Tatsuya Suzuki, Shinkichi Inagaki, Soichiro Hayakawa

***Abstract*— This paper presents a new hierarchical segmentation of the observed driving behavioral data based on the multiple levels of abstraction of the underlying dynamics. By synthesizing the ideas of a feature vector definition revealing the dynamical characteristics and an unsupervised clustering technique, the hierarchical segmentation is achieved. The identified mode can be regarded as a kind of symbol in the abstract model of the behavior. Second, the grammatical inference technique is introduced to develop the context-dependent grammar of the behavior, i.e., the symbolic dynamics of the human behavior. In addition, the behavior prediction based on the obtained symbolic model is performed.**

I. Introduction

Recently, Many ideas have been exploited for the driver modeling from viewpoint of the control technology and the information processing to realize the safety and human-friendly cars [1][2][3][4].

In the driving behavior, it is often found that the driver appropriately switches between simple control laws instead of adopting the complex nonlinear control law. This idea can be verified by executing a 'segmentation' of the observed driving data according to the classification of the dynamical characteristics underlying the behavioral data [7][8][9]. This strategy also can be regarded as one of the solutions for the 'symbolic grounding' problem by assigning each obtained mode to each symbol. Furthermore, the transition between modes can be regarded as a kind of driver's decision-making in the complex driving task [9]. Thus, the introduction of the segmentation leads to higher level understanding of the driving behavior wherein the motion control and decision making aspects are synthesized.

Another important characteristics in the human behavior is described by its hierarchical structure, i.e., many behaviors can be understood by a hierarchical modeling characterized by the different level of abstraction of dynamics. From this viewpoint, it is quite natural to introduce the 'hierarchical segmentation' in the analysis of the human behavior. As a consequence, a hierarchical symbolization of the human behavior can be realized based only on the observed behavioral data (without any prior knowledge). The hierarchical symbolization is expected to play an essential role in the design of intelligent human support system thanks to its high describability and understandability of the complex behavior.

A.Nakano, H.Okuda, T.Suzuki and S.Inagaki are with the Department of Mechanical Science and Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan t_suzuki@nuem.nagoya-u.ac.jp

S.Hayakawa is with the Department of Mechanical Engineering, Mie University, Kurimamachiya-cho 1577, Tsu, Mie, Japan hayakawa@ss.mach.mie-u.ac.jp

Based on these considerations, first of all, we propose a new hierarchical segmentation of the observed driving behavioral data based on the multiple levels of abstraction of the underlying dynamics. In order to realize this idea, a PieceWise AutoRegressive eXogenious (PWARX) model is implemented. The PWARX model is often used as the identification model of the hybrid dynamical systems [5][6] wherein each ARX model represents the corresponding dynamics of each mode. In our problem setting, the number of modes (the number of symbols) is supposed to be controllable to obtain the hierarchical structure although it is assumed to be fixed in the standard framework of the hybrid system identification. By synthesizing the definition of the feature vector revealing the dynamical characteristics[5] and an unsupervised clustering technique, the hierarchical segmentation is achieved. The usefulness of the hierarchical segmentation is demonstrated by applying to the driving behavioral data on the expressway. Second, the grammatical inference technique[11] is introduced to develop the context-dependent grammar of the behavior, i.e., the symbolic dynamics of the human behavior. The vector quantized environmental information and the identified mode obtained by the clustering are regarded as the environment symbol and the mode symbol, respectively. Then, the production rules to express the relation between the environment and mode symbols are identified. Finally, the behavior prediction based on the obtained symbolic model is performed and discussed.

II. Hierarchical segmentation

In this section, we discuss how to define the 'mode' in the driving behavioral data and how to obtain the hierarchical structure. First of all, the driver input and output are defined.

A. Definition of input and output

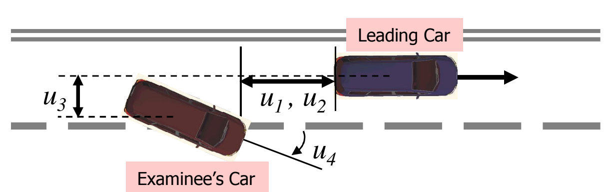

Fig. 1. Definition of input signals.

Throughout this paper, we focus on the driving behavior on the expressway which consists of 'following the leading vehicle', 'lane changing', 'overtaking', and so on. The driver input, i.e., the sensory information of the driver is defined as follows (Fig.1):

- Range from the leading car: u_1
- Range rate between the leading and examinee's cars: u_2
- Lateral displacement from the leading car: u_3
- Yaw angle of examinee's car: u_4
- Index for approaching (KdB): u_5
- Amount of time duration that the examinee looks at the left side mirror in the latest 10 [sec] (TL): u_6
- Amount of time duration that the examinee looks at the right side mirror in the latest 10 [sec] (TR): u_7

KdB is an index which represents the logarithm of the time derivative of the area of the back of the leading car projected on the driver's retina [10]. The KdB can be expressed by using u_1 and u_2 as follows:

$$KdB = \begin{cases} -10 \times \log(|-2 \times \frac{u_2}{u_1^3} \times \frac{1}{5\times 10^{-8}}|) & : u_2 > 0 \\ 10 \times \log(|-2 \times \frac{u_2}{u_1^3} \times \frac{1}{5\times 10^{-8}}|) & : u_2 < 0 \end{cases} \tag{1}$$

The large KdB implies that the driver is facing dangerous situation. Also, the driver output is defined as follows:

- Steering angle: y_1
- Pedal operation: y_2

These input and output variables are chosen so that the resulting model can express the behavioral characteristics underlying the observed data. Furthermore, these variables can be observed in the real driving situation by using existing sensors.

B. PWARX model as mathematical representation of multi-mode driving behavior

In this subsection, the PWARX model is implemented as a mathematical model of the driving behavior. The PWARX model consists of the several ARX sub-models, i.e., modes, and can express a complex input-output relationship with any approximation level by appropriately controlling the number of modes. We consider the following first order PWARX model which has s modes:

$$y(k) = f(r(k)) + \epsilon(k)$$

$$f(r(k)) = \begin{cases} \theta_1 r(k) & if\ r(k) \in \mathcal{R}_1 \\ \theta_2 r(k) & if\ r(k) \in \mathcal{R}_2 \\ & \vdots \\ \theta_s r(k) & if\ r(k) \in \mathcal{R}_s \end{cases} \tag{2}$$

where $y(k)$ and $r(k)$ are defined as follows:

$$y(k) = (y_1(k)\ y_2(k))^T \tag{3}$$

$$r(k) = (u_1(k-1)\ u_2(k-1)\ \cdots\ u_7(k-1)\ y_1(k-1)\ y_2(k-1))^T \tag{4}$$

The subscript k denotes the sampling index ($k = 1, 2 \ldots, n$). Furthermore, θ_i ($i = 1, \cdots, s$) is a (2×9) unknown matrix to be identified from the data, and is supposed to have a form:

$$\theta_i = \begin{pmatrix} \theta_{i,1}^T \\ \theta_{i,2}^T \end{pmatrix} \tag{5}$$

In the PWARX model, not only parameters θ_i but also the partitions of the subspaces $\mathcal{R}_1, \cdots, \mathcal{R}_s$ are unknown. Therefore, it is not straightforward to assign each observation $(y(k), r(k))$ at sampling instant k to the corresponding mode. To resolve this problem, a clustering based technique is developed in [5] under the definition of interesting feature vector which represents the local dynamical characteristics underlying $(y(k), r(k))$. In the next subsection, this feature vector is introduced.

C. Definition of feature vector

1) Assume that the set of sample data $\{(y(j), r(j))\}$, $(j = 1, 2 \ldots, n)$ is given. For each sample data $(y(j), r(j))$, collect the neighboring c data in the (y, r) space, generate the local data set LD_j, and calculate the feature vector ξ_j (Fig.2). Note that the index j indicates the order not in the time space but in the data space. The feature vector ξ_j consists of the local parameters $((\theta_{j,1}^{LD})^T, (\theta_{j,2}^{LD})^T)^T$ in the local ARX model for the LD_j and the mean value m_j of the data r in the LD_j. $(\theta_{j,l}^{LD})^T$ $(l = 1, 2)$ and m_j are calculated as follows:

$$\theta_{j,l}^{LD} = (\Phi_j^T \Phi_j)^{-1} \Phi_j^T y_{LD_j,l} \tag{6}$$

$$m_j = \frac{1}{c} \sum_{r \in LD_j} r \tag{7}$$

where $y_{LD_j,l}$ $(c \times 1;\ l = 1, 2)$ is the output samples in the LD_j, and Φ_j is given by

$$\Phi_j = (r_1\ r_2\ \cdots\ r_c)^T \quad (r \in LD_j). \tag{8}$$

As the result, $\xi_j = ((\theta_{j,1}^{LD})^T, (\theta_{j,2}^{LD})^T, m_j^T)^T$

2) For each feature vector ξ_j, the following covariance matrix R_j is calculated:

$$R_j = \begin{pmatrix} V_{j,1} & 0 & 0 \\ 0 & V_{j,2} & 0 \\ 0 & 0 & Q_j \end{pmatrix} \tag{9}$$

where

$$V_{j,l} = \frac{SSR_{j,l}}{c - (9+1)} (\Phi_j^T \Phi_j)^{-1} \tag{10}$$

$$SSR_{j,l} = y_{LD_j,l}^T (I - \Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T) y_{LD_j,l} \tag{11}$$

$$Q_j = \sum_{r \in LD_j} (r - m_j)(r - m_j)^T \tag{12}$$

The feature vector ξ_j represents the combination of the local dynamics and data. By this definition, the data is classified based not only on the value of data but also on the similarity of the underlying dynamics. Furthermore, the covariance matrix R_j represents the confidence level of the corresponding feature vector ξ_j. R_j is used as the weighting matrix in the calculation of the dissimilarity between feature vectors in the clustering procedure.

D. Unsupervised hierarchical clustering

The unsupervised hierarchical clustering is applied to the feature vectors ξ_j $(j = 1, \cdots, n)$. The clustering algorithm is listed below:

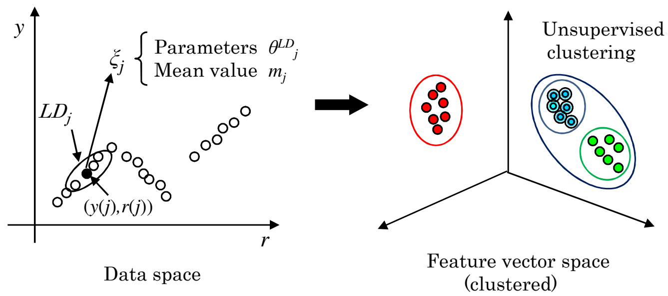

Fig. 2. Transformation from data space to feature vector space.

1) Regard each feature vector ξ_j as each cluster C_j, i.e., each cluster consists only of one feature vector. Calculate the dissimilarity $D_{p,q}$ between any two clusters C_p and C_q by using the following dissimilarity measure:

$$\begin{aligned} D_{p,q} &= \| \xi_p - \xi_q \|^2_{R^{-1}_{p,q}} \\ &= (\xi_p - \xi_q)^T R^{-1}_{p,q} (\xi_p - \xi_q) \end{aligned} \tag{13}$$

where

$$R^{-1}_{p,q} = R^{-1}_p + R^{-1}_q. \tag{14}$$

2) Unify two clusters C_x and C_y which shows the smallest $D_{x,y}$. The unified cluster is denoted by C_r. If all clusters are unified, terminate the algorithm. Otherwise, go to step 3).
3) Calculate the dissimilarity $D_{r,t}$ between C_r and C_t for all $t\,(t \neq r)$ by using the following dissimilarity measure:

$$D_{r,t} = \frac{n_r n_t}{n_r + n_t} \sum_{\xi_{i_r} \in C_r} \sum_{\xi_{i_t} \in C_t} \| \xi_{i_r} - \xi_{i_t} \|^2_{R^{-1}_{i_r,i_t}} \tag{15}$$

where n_r and n_t are numbers of feature vectors belonging to clusters C_r and C_t, respectively. Go to step 2).

After this clustering procedure, the classification of the feature vector space is achieved together with a dendrogram which shows the hierarchical classification for different number of modes. Since the transformation from the feature vector (ξ) space to the original observed data $(y,\, r)$ space is straightforward, the segmentation of the observed data is obtained together with the hierarchical structure.

Note that once segmentation of the data is achieved, the identification of the parameters θ_i and the partitions of the subspaces $\mathcal{R}_1, \cdots, \mathcal{R}_s$ in the PWARX model (2) is straightforward.

III. Analysis of Driving Behavioral Data

A. Driving environment

In this paper, the following driving environment on the expressway was designed on the driving simulator which provides a stereoscopic immersive vision [9].

- The expressway is endless, and has two lanes, the cruising lane and the passing lane.
- There are 10 cars on the cruising lane. Five of them are running ahead of the examinee's car. The remaining five cars are running behind the examinee's car. Their velocities vary from 70 to 85[km/h]. Once the examinee's car overtakes the leading car, then the tale-end car on the cruising lane is moved to the head of the cars running on the cruising lane. The examinee is not aware of this change.
- There are 10 cars on the passing lane. Five of them are running ahead of the examinee's car. The remaining five cars are running behind the examinee's car. Their velocities vary from 90 to 110[km/h]. Once the examinee's car is overtaken by the car on the passing lane, then the top car on the passing lane is moved to the tale-end of the cars running on the passing lane. The examinee is not aware of this change.
- The range between cars is set to be 50 to 300[m], and there is no collision between cars except the examinee's car.
- There is no lane change of the cars except the examinee's car.

Under this driving environment, five examinees performed the test driving. Note that the examinees were provided with the instruction 'Drive the car according to your usual driving manner'. Since this instruction is 'loose' instruction, the examinees do not concern much about the environmental information. As the result, each examinee can drive as his/her usual manner.

B. Observed behavioral data and clustering results

The unsupervised clustering based on the feature vector shown in the previous section has been applied to the observed driving behavioral data. The dendrogram obtained from the proposed strategy is shown in Fig.3. In Fig.3, the vertical axis represents the dissimilarity between clusters. When the two clusters are unified, the corresponding dissimilarity is designated by the horizontal bar. The horizontal axis represents the data which is rearranged after the clustering to show the hierarchical structure clearly. From this figure, we can clearly understand the hierarchical structure in the driving behavior. As the typical example, the two dashed horizontal lines are superimposed. The upper line shows the case that the number of modes (clusters) s, i.e., the number of the ARX models in (2) is set to be two. On the other hand, the lower line shows the case that s is set to be five.

In Fig.4, the observed driving (input-output) profiles are shown. All profiles are normalized before clustering. In the profile of the lateral displacement, it takes positive value when the examinee's vehicle is on the right side of the leading car. The steering angle takes positive value when the examinee turns it clockwise. Also, the pedal operation takes positive value when the accelerator is stepped on, and takes negative value when the braking pedal is stepped on. Note that the range, the range rate and the lateral displacement profiles show discontinuity. Since these variables are defined by the relative displacement from the leading car, if the examinee's car changes the driving lane, these variables change discontinuously.

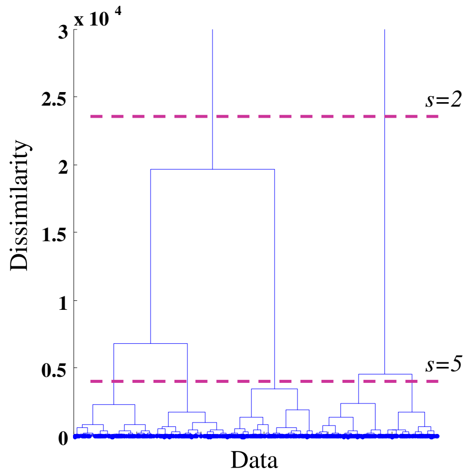

Fig. 3. Dendrogram of clustering (Examinee A).

In addition, the clustering results in the case of two-mode modeling are indicated by colors in Fig.4. Thus, the segmentation works well. In order to investigate the behavioral meaning of each mode, a part of the profile of the lateral displacement is enlarged in Fig.5. As shown in Fig.5, the meaning of two modes can be understood as the 'Following on Cruising Lane + Passing' (Mode 1: FC+P mode) and 'Following on Passing Lane + Returning' (Mode 2: FP+R mode), respectively. This result implies that the symbolization of the behavior can be achieved based on the 'dissimilarity' of the underlying dynamics.

C. Discussion

In order to analyze the hierarchical structure of the behavior, the clustering results in the case of five-mode modeling are shown in Fig.6, and the enlarged lateral displacement is shown in Fig.7. From Fig.7, we can see that the two-mode model is further decomposed into the local behaviors; they are 'Long Range Following on Cruising Lane' (Mode 1: LRFC mode), 'Short Range Following on Cruising Lane' (Mode 2: SRFC mode), 'Passing' (Mode 3: P mode), 'Following on Passing Lane' (Mode 4: FP mode), and 'Returning' (Mode 5: R mode). The switching between these modes is caused by the driver's decision making. The hierarchical relationship between these modes found in the dendrogram is depicted in Fig.8. Thus, the hierarchical structure of the driving behavior can be obtained in a quite consistent manner. One of the significant contributions of this work is that this hierarchical structure is obtained automatically based only on the observation (including the definition of the input and output signals) and data processing. Since this hierarchy clearly expresses the multiple abstraction level of the human behavior, the proposed framework is expected to be a basis for the design of many human centric systems.

IV. Development of Symbolic Behavior Model and Its Application to Behavior Prediction

In this section, the human behavior is considered as an entity (linguistic source) capable of generating a specific

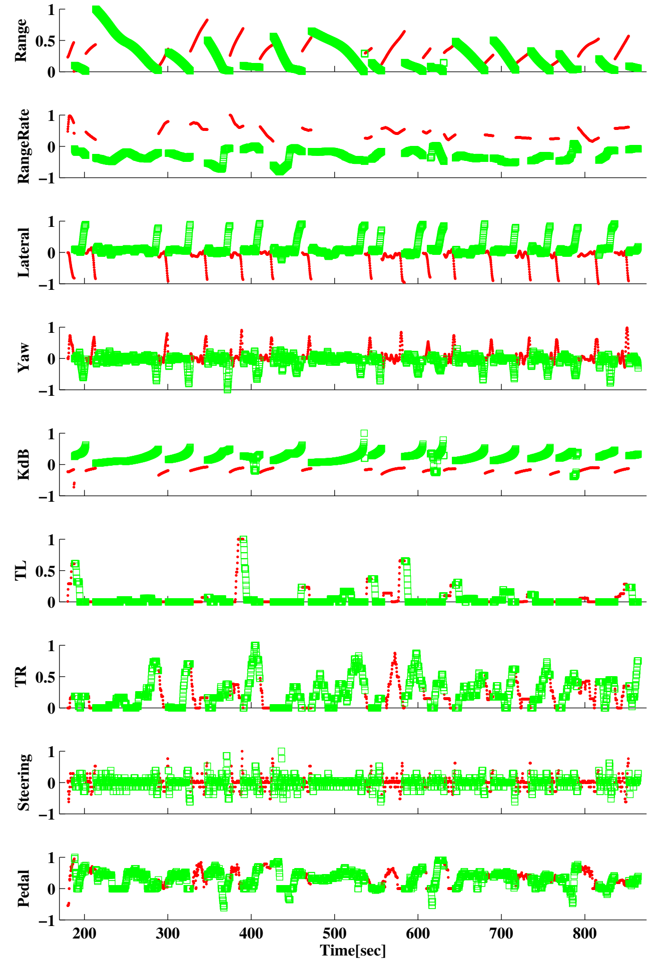

Fig. 4. Observed profiles and segmentation result (Examinee A, 2modes).

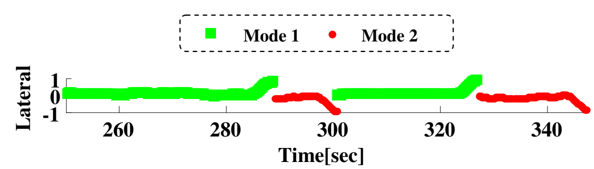

Fig. 5. Enlarged profile of lateral displacement (Examinee A, 2modes).

language (set of symbol strings). The grammar G of the language is the set of production rules that specifies all the strings in the language and their relationships. Once the grammar is found, the grammar itself is a model for the source of the behavior.

A. Definition of behavioral grammar

First of all, the behavioral grammar G is defined as follows:

$$G = \{\Sigma_m, \Sigma_e, S, P\} \tag{16}$$

Σ_m is a mode alphabet, i.e., the set of mode symbols defined by the clustering introduced in section III. Therefore, the number of mode symbols $|\Sigma_m| = s$, i.e., the number of ARX models. Σ_e is an environment alphabet, i.e., the set of

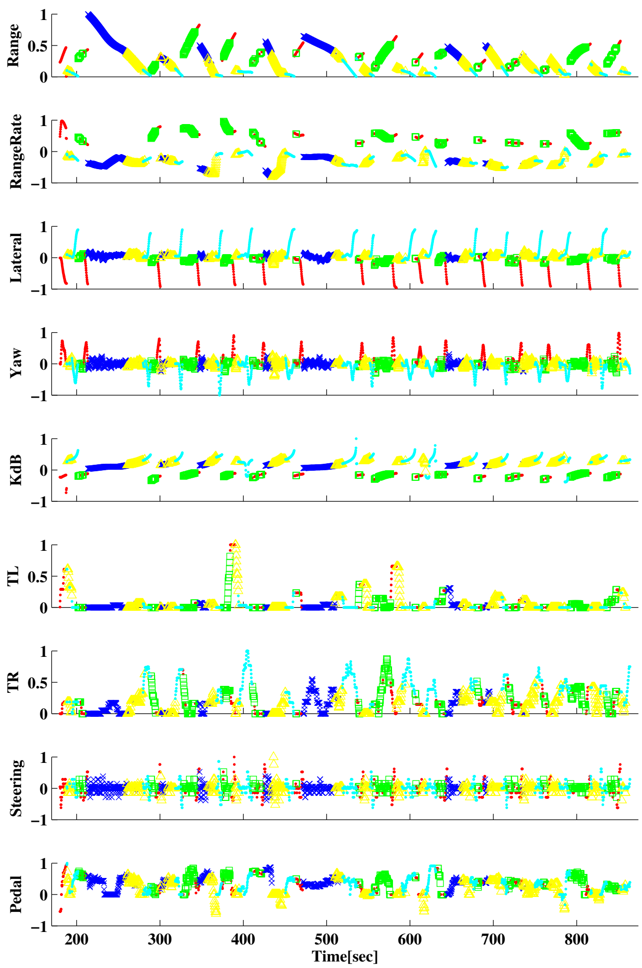

Fig. 6. Observed profiles and segmentation result (Examinee A, 5modes).

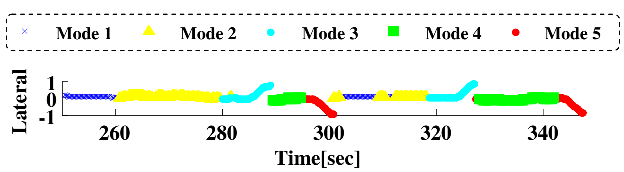

Fig. 7. Enlarged profile of lateral displacement (Examinee A, 5modes).

symbols created by a vector quantization of the environmental information. The number of environment symbols $|\Sigma_e|$ depends on the quantization. In the proposed framework, Σ_m and Σ_e are regarded as a terminal alphabet and a nonterminal alphabet in the standard grammar, respectively. S is a special nonterminal symbol used to start the generation of string. P is a set of production rules, i.e., the substitution rules (denoted by $a \rightarrow b$) used to generate the strings. The $n-type$ production rules are defined as substitution rules of the form

$$m_{k-n} \cdots m_{k-1} E_k \rightarrow m_{k-n} \cdots m_{k-1} m_k \delta \qquad (17)$$

where $m_{k-n} \cdots m_{k-1}$ is a sequence of mode symbols, E_k is an environment symbol, and δ is a special nonterminal symbol. δ is used to indicate the conclusion, or not, of a generated string. The $n-type$ production rule encodes the

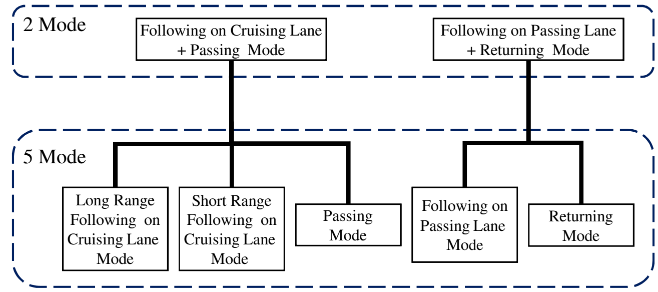

Fig. 8. Identified hierarchical structure of driving behavior (Examinee A).

evolution of the mode depending on its n past modes and on the environment symbol E. Therefore, the $n-type$ production rule can be regarded as a symbolic dynamics whose order is specified by n. Once G is identified, the symbolic behavior can be computed by executing the production rules.

B. Grammatical inference

Development of the symbolic behavior model can be formulated as the grammatical inference problem [11] under the suitable definitions of the mode and environment alphabets. Grammatical inference, in general, is the identification of a grammar from a set of examples. The main part pf the grammatical inference is the generation of the production rules based on the observation, and is realized by the following procedure (See [11] for detail).

1) A $0-type$ production rule is assumed for every newly occurring environment symbol.
2) A new $(n+1)-type$ production rule is generated whenever the data conflicts with the previously established $n-type$ production rules. The conflicting $n-type$ production rules are also promoted to $(n+1)-type$ production rules or are deleted if there is not sufficient information in the past.

C. Application to symbolic behavior modeling and prediction

1) Definition of environment symbol: First of all, the environment symbols are defined by the vector quantization of the the relative position (X_i) and velocity (V_i) of the six surrounding cars as shown in Fig.9. Since the goal is to realize the long-term prediction based on the symbolic model, the wider range of cars are considered as the environment than the definition of the input variables for the PWARX model. The CSL (Competitive and Selective

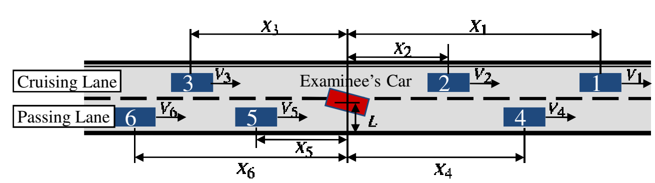

Fig. 9. Definition of environment.

Learning) algorithm was used for the quantization. The necessary number of symbols depends on the complexity of the environment. Here, 10 symbols were defined by trial and error.

2) Behavior prediction based on production rules: By applying the Grammatical inference to the two-mode model and the five-mode model, we have developed the two symbolic behavior models with different definition of the mode symbol.

The number of identified rules and the average type of the rule are shown in Table I. In the two-mode model, the number of identified rules is smaller, but the average type of the rule is higher compared with the five-mode model. This implies that these factors depend on the 'resolution' of the symbolic representation. Another interesting inquiry is that the number of identified rules varies from examinee to examinee. The examinee who has great number of rules (like the examinee E) can be considered to have an inconsistent driving manner.

In addition, the prediction of the behavior based on the symbolic model was performed. In order to predict the future behavior, the prediction of the environment symbol must be considered. In this work, the prediction of the environment symbol was realized by a simple first-order prediction of X_is and V_is. Figure 10 shows the success rate of the prediction for various prediction horizon using the several models with different number of modes (1 step is 240 [msec]). From Fig.10, the success rate goes down as the prediction horizon becomes longer. However, even in the five-mode model, about 70% success rate is achieved for 10step (2.4[sec]) ahead prediction. This long-term prediction has never been realized in the conventional behavior model based on the controller model or the information processing model. Furthermore, the low-mode model shows higher success rate than the high-mode model. Thus, the proposed framework can control the prediction accuracy by choosing the 'resolution' of the symbolic representation.

TABLE I

GENERATED PRODUCTION RULES (FIVE EXAMINEES)

Examinee	Number of Data	two-mode model Number of Rules	two-mode model Average of Rule Type	five-mode model Number of Rules	five-mode model Average of Rule Type
A	2851	90	27.6	1026	18.4
B	2908	120	22.6	996	17.7
C	2424	273	21.0	962	18.2
D	3008	448	21.0	1251	18.4
E	3063	670	20.6	1640	14.8

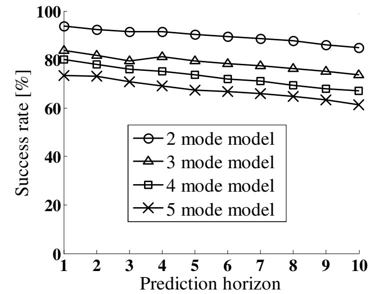

Fig. 10. Success rate of the prediction for various prediction horizon (all examinees).

V. CONCLUSION

This paper has presented a new hierarchical segmentation of the observed driving behavioral data based on the multiple levels of abstraction of the underlying dynamics. By synthesizing the ideas of the feature vector definition revealing the dynamical characteristics and the unsupervised clustering technique, the hierarchical segmentation has been achieved. The identified mode can be regarded as a kind of symbol in the abstract model of the behavior. Second, the grammatical inference technique was introduced to develop the context-dependent grammar of the behavior, i.e., the symbolic dynamics of the human behavior. In addition, the behavior prediction based on the obtained symbolic model was performed and discussed. The proposed framework enables us to make a bridge between the signal space and the symbolic space in the understanding of the human behavior. The design of the environment symbol with hierarchical structure and application to anomaly detection are our future works.

REFERENCES

[1] C.MacAdam: Application of an optimal preview control for simulation of closed-loop automobile driving, *IEEE Trans. on Syst. Man and Cyber.*, Vol.11, No.9, pp.393-399, 1981.

[2] A.Modjtahedzadeh and R.Hess: A Model of Driver Steering Control Behavior for Use in Assessing Vehicle Handling Qualities, *ASME J. of Dynam. Syst. Meas. Contr.*, Vol.15, Sept., pp.456-464, 1993.

[3] T.Pilutti and A.G.Ulsoy: Identification of Driver State for Lane-Keeping Tasks, *IEEE Trans. on Syst. Man and Cyber., Part A*, Vol.29, No.5, pp.486-502, 1999.

[4] M.C.Nechyba, Y.Xu: Human Control Strategy: Abstraction, Verification, and Replication, *IEEE Control Systems Magazine*, October pp.48-61, 1997.

[5] G.Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari: A clustering technique for the identification of piecewise affine system, *Automatica*,Vol.39, pp.205-217, 2003

[6] T.Roll, A.Bemporad, L.Ljung: Identification of piecewise affine systems via mixed-integer programming, *Automatica*, Vol.40, pp.37-50, 2004

[7] J.H.Kim, S.Hayakawa, T.Suzuki, K.Hayashi, S.Okuma, N.Tsuchida, M.Shimizu, S.Kido: Modeling of Driver's Collision Avoidance Maneuver based on Controller Switching Model, *IEEE Trans. on System, Man and Cybernetics*, part B, Vol.35, No.6, pp.1131-1143, 2005.

[8] S.Sekizawa, S.Inagaki, T.Suzuki, S.Hayakawa, N.Tsuchida, T.Tsuda, H.Fujinami: Modeling and Recognition of Driving Behavior Based on Stochastic Switched ARX Model, *IEEE Trans. on Intelligent Transportation Systems*, Vol. 8, NO. 4, pp.593-606, 2007

[9] T.Akita, T.Suzuki, S.Hayakawa, S.Inagaki: Analysis and Synthesis of Driving Behavior based on Mode Segmentation, *Proc. of Int. Conf. on Control Automation and Systems*, pp.2884-2889, 2008

[10] T.Wada, S.Doi, K.Imai, N.Tsuru, K.Isaji, H.Kaneko: Analysis of Drivers' Behaviors in Car Following Based on a Performance Index for Approach and Alienation, *Proc. of SAE2007 World Congress*, SAE paper 2007-01-0440.

[11] K.S.Fu, T.L.Booth: Grammatical Inference: Introduction and Survey - Part II, *IEEE Trans. on System, Man and Cybernetics*, Vol.5, pp.409-423, 1975