Traffic Measurement Using Linear Predictive Coding of Road Environmental Sound from Microphone on The Vehicle

Hironori Hara and Shinji Ozawa
Faculuty of Information and Computer Science, Dept. of Science and Technology, Keio University
Hiyoshi 3–14–1, Kouhoku-ku, Yokohama-shi, Kanagawa 223–8522, Japan
Email: hironori@ozawa.ics.keio.ac.jp and ozawa@ozawa.ics.keio.ac.jp

***Abstract*— Traffic measurement is one of the important development project of ITS. Present traffic measurement system generally uses signal obtained from ultrasonic sensor, loop coil, and infra-red sensor. However, these sensor's sensing region is a limitation. Therefore, the development of sensors which are able to obtain a wide range information is needed. Image sensors and sound sensors are some selected examples of such a sensor. Although the development of image sensor is performed at various organizations, examples of research using sound sensor are limited. This is one of the reasons that sound sensor is not put to practical use, even though it is expected. Therefore, this paper, we propose the method that performs traffic measurement using a sound sensor. The sound sensor is on the vehicle, and traffic measurement is performed by counting the vehicles which are passing each other coming from the opposite lane. This method is based on Linear Predictive Coding analysis and Self-Organizing Maps algorithm. The method is divided into the learning phase and the recognition phase. During the learning phase, the feature map is created by SOM algorithm using spectral peaks estimated by the LPC analysis. At the recognition phase, vehicles passing each other are recognized from the track of winner's position on the feature map.**

I. INTRODUCTION

ITS is being actively promoted by many countries as a means to improve traffic safety and efficiency.

Traffic measurement which grasps the traffic flow is one of the important development project of ITS. Present traffic measurement system generally uses signal obtained from ultrasonic sensor, loop coil, and infra-red sensor. However, these sensor's sensing region is a limitation. Therefore, the development of the sensors which are able to obtain a wide range of information is needed. Image sensors and sound sensors are some selected example of such a sensor. Between both, research on traffic measurement using image sensor has been especially conducted at various organizations[1], [2]. However, there are few examples of research using sound sensor, this sensor has not been put to practical use yet.

There is a fair possibility that in some cases sound sensor is more useful than image sensor because of the merits which the image does not have. For example, sound is not influenced by a change of lighting. In addition, there is no resolution and picture angle issue. It is noted that sound data is generally smaller in comparison with image data. These advantages, plus low computation cost and hardware resource, in comparison with image sensor, are the main reasons that the development of sound sensor is expected.

In previous work, many research treat environmental sound recorded from sound sensors as noise. For example, the measurement of road traffic noise volume[3] and the research on improvement of the quality of noise[4]. On the other hand, Taguchi[5] proposed using environmental sound for identifying vehicle type based on time-varying spectral pattern of traffic noise.

Road information is generally obtained from sensors installed at a fix place, e.g. mounted at the roadside or tall buildings. Therefore, we will not be able to collect information about the traffic flow for roads that is not installed with sensors. Any new sensor installed requires planning, which is time consuming and often incurred high cost, because the possibility of installing sensors on all road is not possible. Therefore, we proposed using a probe-car that works as a mobile sensor. Here, we develop a 'not-specialized' vehicle to work as a probe-car. These probe-cars collect the road information data and then transfer it to the control center. With this, the probe-car is not only used to perform sensing, but is also able to obtain road information at any place automatically. Therefore, it is not necessary to install new sensor on the road. This idea required us to develop sensing algorithms.

Therefore, this paper proposed using sound sensor installed in a vehicle as mobile sensors which are able to collect traffic measurement. Traffic measurement is performed by counting the vehicles which are passing each other on the opposite lane. We develop an algorithm especially to recognize the passing of vehicle from each other using road environmental sound.

II. THE TARGET ROAD ENVIRONMENTAL SOUND

The situations of traffic or surroundings influence road environmental sound. It is difficult for the system to adapt to all the situation. Thus, it is necessary to make the target situations clear. TableI shows the target situations that we recognieze of in this paper.

The type of road shows that the road is a general road or a superhighway. A general road has high traffic daily and changes or installing new sensors requires traffic re-routing which is time consuming and high cost. Nevertheless, these roads are usually congested and detail traffic measurement is needed. We may say that it is suitable for traffic measurement at a general road using mobile sensors on the vehicle.

This paper treats of the roads which have one lane in each direction. Traffic measurement is performed by counting the vehicles which are passing each other on the opposite lane.

It is difficult to measure traffic jam quantitatively. Therefore, this paper assumes a flowing traffic. In some cases, vehicles stop at traffic lights or move slowly, but throughout the experiments there was no traffic jam.

This paper assumes the weather is good with no rain or strong winds. Although numerical data shows instantaneous wind velocity from different direction, it is difficult to categorize and analyze quantitatively.

TABLE I

TARGET SITUATIONS THAT WE TREAT OF IN THIS PAPER.

Situation	Target
Type of roads	Not a super highway
Number of lanes	One lane in each direction
Traffic jam	No
Weather	Good
Strong wind	No

III. FREQUENCY ANALYSIS BY LINEAR PREDICTIVE CODING

A. Frequency domain features

When a vehicle passes each other, road environmental sound is influenced by both the frequency domain, and the time domain. In the time domain, it is difficult to recognize only the influence of two vehicles passing each other because of noise. Therefore, features of the frequency domain are analyzed.

B. Linear Predictive Coding

Sound waves are segmented into frames in order to extract the frequency domain features. Then, we employ the frequency analysis method in each frame. The frequency analysis is performed using Linear Predictive Coding (LPC)[6]. LPC assumes that the present sample x_t can be represented as a linear combination of previous p sample. The spectral estimation is performed by determining the linear prediction coefficients a_i $(1 \leq i \leq p)$. This technique has been widely used in speech analysis because the spectral estimation can be performed by using a very small number of parameters.

In LPC, it is important to determine the number of LPC coefficients p. If p is in high-order, the model by LPC generally adapts to input signal. However, model overfitting is performed if p is made too high. In this case, the prediction coefficients that do not adapt to the input signal are determined. Then, the Final Prediction Error (FPE) technique proposed by Akizuki et al[7] is employ to determine p value dynamically.

The predicted sample $\hat{x}_t$ of x_t is shown in equation (1).

$$\hat{x}_t = -\sum_{i=1}^{p} a_i x_{t-i} \tag{1}$$

The linear prediction error $e_t(p) = x_t - \hat{x}_t$ at the p-th order is shown in equation (2), applying equation (1), assuming $a_0 = 1$.

$$e_t(p) = \sum_{i=0}^{p} a_i x_{t-i} \tag{2}$$

The prediction error variance $\sigma_e^2(p)$ is shown in equation (3) using equation (2). N is the number of samples per frame.

$$\sigma_e^2(p) = \frac{1}{N}\sum_{i=0}^{N-1} e_t(p)^2 \tag{3}$$

FPE technique adopts p which minimizes the final prediction error $FPE(p)$ as the number of LPC coefficients. $FPE(p)$ is shown in equation (4).

$$FPE(p) = \frac{N+(p+1)}{N-(p+1)}\sigma_e^2(p) \tag{4}$$

Finally, LPC is performed by using p value that is calculated by FPE, and a_i value which is calculated by a covariance method.

C. The spectral estimation result by LPC

Fig.1 shows the peak positions that is determined by the LPC. The parameters of LPC are shown in TableII. The road environmental sound in Fig.1 fulfills the conditions of TableII. The change of the number of peaks is shown in Fig.2. The probe-car was running at about 50km/h on the uncovered and straight line road. The time of vehicles passed each other is denoted by closed triangles in Fig.1 and Fig.2. In this paper, the group of nodes that respond to vehicle's passing each other is called *the passing each other state*, and other groups are called *the other state*.

Fig.1 and Fig.2 show that several peaks appear when the vehicles are passing through each other, and few peaks appear in the case of the other state. Moreover, peaks frequently appear around 1kHz, 4kHz, or 10kHz in the case of the passing each other state.

In some cases, peaks, however, appear in some bands of frequency above, and some cases show the similar tendency in the passing each other if the noise was added. The noise, for example, is the sound produces when something hits vehicle's body, and the influence of the wind. Therefore, it is difficult to extract the features of the passing each other state using only the result of the spectral estimation. Thus, we introduce learning process, and the features of each frame are classified into the passing each other state and the other state.

IV. SELF-ORGANIZING MAPS

Kohonen's Self-Organizing Maps (SOM)[8] is a neurocomputational algorithm of unsupervised learning. This algorithm is considered an artificial neural network model of the brain and is widely used.

Road environmental sound is reflected by not only the influence of two vehicles passing each other but also various situation. In order to recognize the passing each other state, it

TABLE II

THE PARAMETERS OF SOUND ANALYSIS.

Parameters	Set value
Sampling rate	48000 [Hz]
Quantization bit rate	16 [bits]
Frame length	1024 [point]
Frame shift	1024 [point]

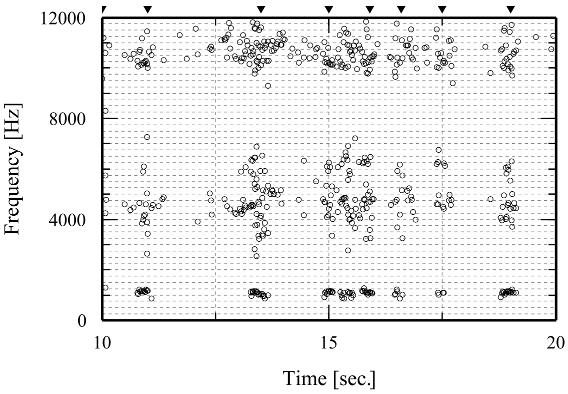


Fig. 1. The result of peak detection using LPC.

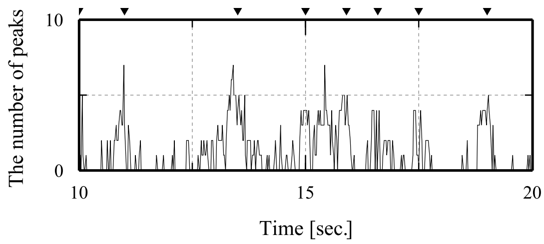


Fig. 2. The change of the number of peaks.

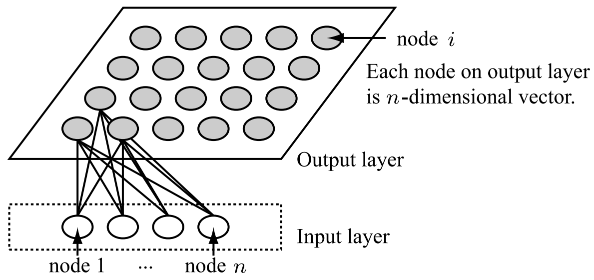


Fig. 3. SOM.

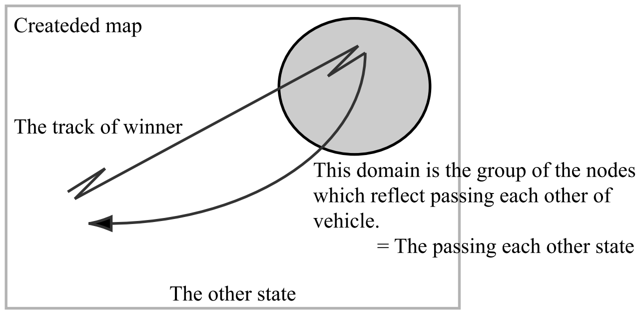


Fig. 4. Created map and winner's track.

is necessary that the situations are classified into the passing each other situation or others. SOM is able to classify the property of the input signal topologically. In addition, the preliminary knowledge about the input signal is unnecessary. Therefore, in this paper, we employ the SOM algorithm.

The SOM is the network that consists of two layers, shown in Fig.3. These layers are called input layers and output layers. Although each neuron of both the input and output layer is called node, both properties are different. The nodes of the input layer are a scalar, and each the input layer consists of n nodes. The each node of the output layer is n-dimensional vector. These nodes are generally arranged in two dimensions, and the input layer is fully connected to the output layer nodes. The learning process of SOM is shown in equation (5).

$$\boldsymbol{m}_i(t+1) = \begin{cases} \boldsymbol{m}_i(t) + \alpha(t)(\boldsymbol{x}(t) - \boldsymbol{m}_i(t)) & \\ & \text{if } i \in N_c(t) \\ \boldsymbol{m}_i(t) & \text{if } i \notin N_c(t) \end{cases} \quad (5)$$

$\boldsymbol{m}_i(t)$ is the n-dimensional vector of the node i on the output layer at the time t, and $\boldsymbol{x}(t)$ is n-dimensional input data. $\alpha(t)$ is called the learning-rate factor ($0 \leq \alpha(t) < 1$). $N_c(t)$ is called neighborhood. Both $\alpha(t)$ and the radius of $N_c(t)$ are usually decreasing monotonically in time.

Given $\boldsymbol{x}(t)$ value, each distance betweeen $\boldsymbol{x}(t)$ and $\boldsymbol{m}_i(t)$ can be calculated, and the node that has nearest $\boldsymbol{m}_i(t)$ to $\boldsymbol{x}(t)$ is determined. This node is called the winner node. $\boldsymbol{m}_i(t)$ of the winner and winner's neighbor nodes are updated so that the distance to $\boldsymbol{x}(t)$ becomes small using equation (5). This process is performed to nodes that exist within N_c from the winner. Both N_c and $\alpha(t)$ become small with progress of learning. This learning process is repeated T times. After the learning process is completed, nodes of the output layer make several groups that reflected the topology of the input vectors, and nodes that belong to the same group are similar.

So, if ideal learning is performed by SOM algorithm, as shown in Fig.4, it is expected that several groups which reflected the current state will be made on the map. Therefore, the system can recognize the current state by the winner's position on the map that learning process was completed.

V. METHOD

Fig.5 shows the overview of the method. The method consists of a learning phase and a recognition phase. At the learning phase, a map is created by SOM. At the recognition phase, the state of the current frame is classified into the passing each other state or the other state by the map created at the learning phase.

A. Learning Phase

1) Features for learning: The first step in SOM is to determine the features for learning. It is required that these features have a correlation which can influence the vehicles that are passing each other. Any error in determining these features will result in the system's inability to perform effective learning. Therefore, the following 31-D vectors are observed as features, using the tendency shown in Fig.1 and Fig.2.

- The number of peaks less than or equal to 12kHz.
- The frequencies of 30 peaks divided by sampling frequency.

If the number of peaks is less than 30, 0 is padded. For example, in the case of sampling frequency $sr = 48000$[Hz], $p = 3$, $f_0 = 1000$[Hz], $f_1 = 4000$[Hz], and $f_2 = 8000$[Hz], the input to SOM becomes $\{$3, 1000/sr, 4000/sr, 8000/sr, 0, $\ldots$, 0$\}$.

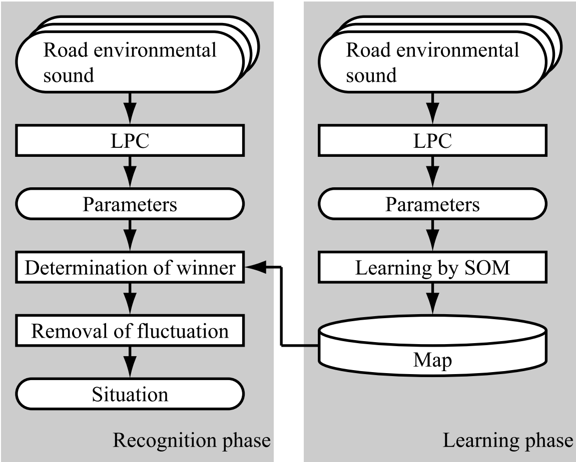


Fig. 5. The overview of the method.

2) *Learning process and determination of the domain of the passing each other state:* In the beginning, $\boldsymbol{m}_i$ of node i is initialized with the random value. $\boldsymbol{m}_i$ learns by using features shown in V-A.1. The near nodes become similar with the progress of learning, and several groups were made on the map when the learning is converged. Although these groups, for example, include the domain of the passing each other state, the domain of the transition from the passing each other state to the other state, and the domain of noise state, we are more concerned with the domain of the passing each other state in this paper. It is necessary to select the domain of the passing each other state in order to recognize the vehicles which pass each other. This process is performed by viewing. After the learning, we input the road environmental sound, and select the domain of the passing each other state by observing the winner's track on the map.

B. Recognition phase

1) *Determination of winner:* The winner is determined by comparing features with $\boldsymbol{m}_i$. In this phase, the map is not changed but only determines the winner.

2) *Removal of fluctuation:* If ideal learning is performed, the winner will move if any vehicle approaches in the opposite side, as explained further. At first, the winner is in the domain of the other state. Then the winner moves towards the domain of the passing each other state. Finally, the winner goes back to the domain of the other state. Sometimes, the winner fluctuates between the passing each other state and other state intermittently. Therefore, it is necessary to remove this fluctuation. We solved this problem by using two values, t_{on} and t_{off}. t_{on} is the threshold value of the time that the winner holds when the winner moves from the passing each other state to other state, while t_{off} is the threshold value of the time that the winner holds when the winner moves from the other state to the passing each other state. These fluctuations can be removed by not changing the current state while the current frame is in t_{on} or t_{off}. This process is shown in Fig.6.

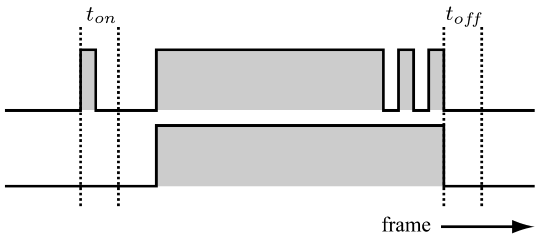


Fig. 6. Removal of the fluctuation.

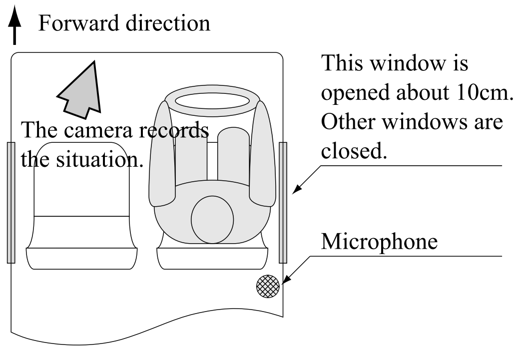


Fig. 7. The position of the microphone.

VI. EXPERIMENT

A. Experimental method

In order to evaluate our proposed method, we run experiments on the road. Normally, the sensor should be mounted outside the vehicle, e.g. the side mirror base, to enable the driver and passengers to behave freely. However, as this research is still in the laboratory stage, the microphone is mounted inside the vehicle, shown in Fig.7.

The microphone was mounted on the inside wall of the opposite lane side. The microphone's height was adjusted to the side mirror's height, and the microphone was protected by a microphone wind screen. In order to make the same situation as if the microphone is mounted on the outside of the vehicle, we open the window about 10cm. In this case, we take care that the wind does not hit the microphone directly. Other windows are closed. In addition, the air-conditioner and audio system were not operated because these are not the noise source if the microphone is being mounted on the outside of the car. A video camera was set on the passenger's seat in order to record the actual situation.

The sounds were recorded using a SONY ECM-23FII electret condenser microphone, and a Digital Audio Tape (DAT) recorder on the fine days in November and January. These sounds were recorded in 16-bits monaural at 48kHz, and were captured in a PC via optical S/PDIF connection. Although this process is off line, future samples are not used in the method.

Recording was performed by driving at the road that fulfilled conditions of TableI, and a total of 2573 seconds of sounds were obtained. A total of 55000 frames (about half of the total time) were used at the learning phase, while others were used at the recognition phase.

TableIII shows the parameters used for the learning of SOM.

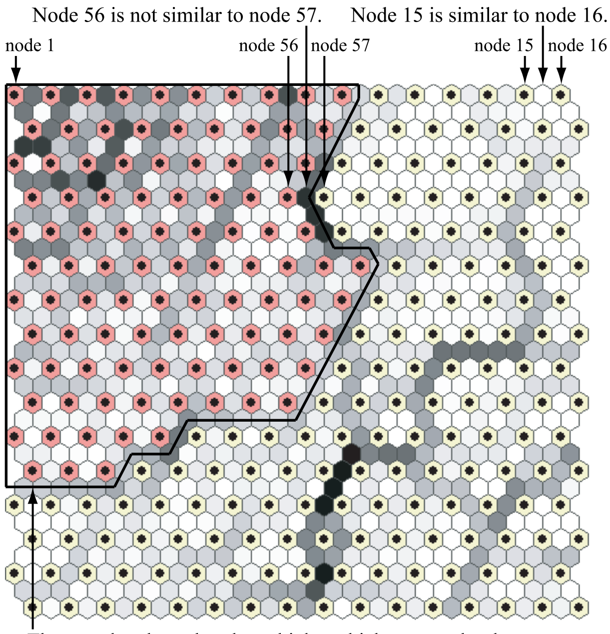


Fig. 8. The created map.

The size of the map grows from experience. The map is square. If N_0 is too small, some nodes are not used for learning. The radius of N_0 is half at the one side of the map. At the learning phase, the order of the inputted features set is given randomly. If a uniform random number is generated, the frame is used twice for learning. α_0 is generally used value. $\alpha(t)$ and $N_c(t)$ decrease monotonically in time.

B. Experimental result

1) Learning Phase: Fig.8 shows the example of the map created using the parameters of TableIII at the learning phase.

Fig.8 is displayed using U-Matrix[9]. U-Matrix shows the distance between each node of the output layer by shades in a gray scale, and indicates that the distance between the two nodes is separate if the shade is dark. The nodes surrounded by the line in Fig.8 are the nodes determined to be in the passing each other state by viewing. Fig.8 was separated into two or more domains. Experimental results reveal that the left-hand side domain showed the passing each other state. Moreover, if the vehicle on the opposite lane existed nearby, the winner was located in the left-hand side of the passing each other state. During the transition from the passing each other state to the other state, the winner was located near the boundaries of the domains.

TABLE III
THE PARAMETERS USED FOR THE SOM LEARNING.

Parameters	Set value
Size of map	16×16
Times of learning T	110000
Initial neighborhood N_0	8
Neighborhood function $N_c(t)$	$N_0 \times (1 - \frac{t}{T})$
Initial learning-rate factor α_0	0.2
Learning-rate factor function $\alpha(t)$	$\alpha_0 \times (1 - \frac{t}{T})$

2) Recognition phase: Fig.9 shows the working of this system at the recognition phase. In Fig.9, the top shows the actual situation of the current frame. The second from the top of Fig.9 shows the result of the spectral estimation. The middle shows the position of the winner on the map. This map and Fig.8 are the same. The second from the bottom is the history of the winner's position. If the winner is in the passing each other state, the diamond is plotted at the "Pass" position. When the winner is in the other state, the diamond is plotted at the "Other" position to the contrary. The bottom is the result after removing the fluctuation, shown in the second graph from the bottom.

Fig.9(a) shows the result at $t = 17.79$ [sec.]. The vehicle on the opposite lane still exists far away. Consequently, the influence of the vehicle on the opposite lane on road environmental sound is small. Thus the peaks do not appear by LPC. The winner of this frame exists in the position that the middle of Fig.9(a) shows. This position is not in the passing each other state.

Fig.9(b) is a result at $t = 18.90$ [sec.]. The vehicle on the opposite lane exists nearby. We noted several peaks appear by LPC and the winner exists in the passing each other state.

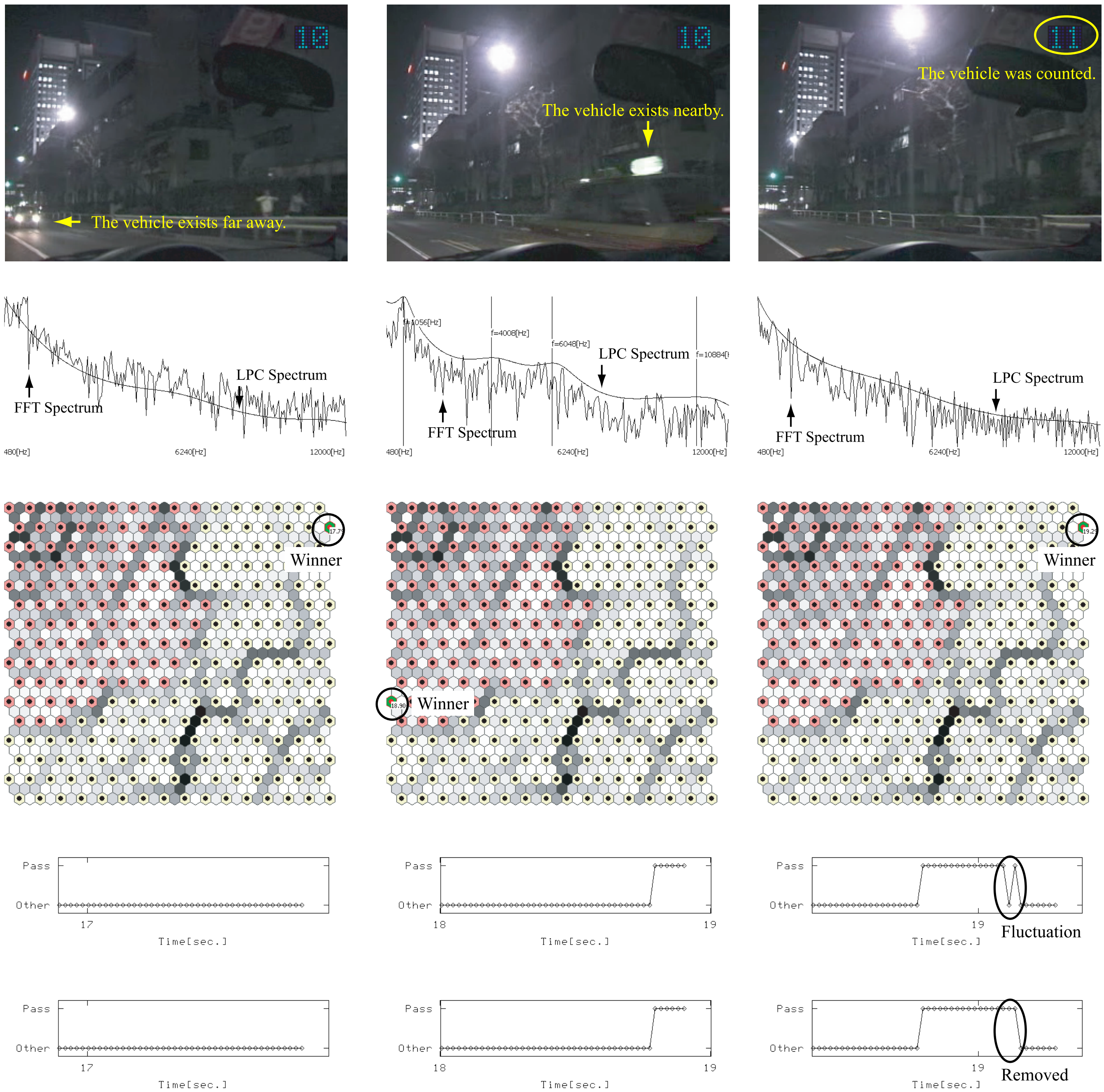
Fig.9(c) is a result in $t = 19.29$ [sec.]. In this frame, the probe-car has passed each other. The peaks do not appear by LPC and the winner exists in the other state. Moreover, the second from the bottom of Fig.9(c) shows the frame which is not in the passing each other state exist. This frame, however, was corrected after the fluctuation removal process. Therefore, the number of vehicles that passed each other was increased by only one.

VII. Discussion

A number of factors, for example, the SOM parameters, the features that are used at the learning phase, and the frames that are used at the learning process, influence the results at the learning phase. At a preliminary experiment, SOM learning was performed by using selected frames. Some frames are (1) the frames that vehicles are passing each other and (2) the frames that vehicles do not exist on opposite lane. In the case of (1), winners were located in the passing each other state if vehicles were passing each other. Winners, however, were located randomly if the frames that the vehicles did not exist on the opposite lane were inputted. In this case, some of the winners were located in the passing each other state. In the case of (2), except that the passing each other state and the other state were interchanged, the same tendency as (1) was shown found.

VIII. Conclusion

This paper proposed a method for traffic measurement using sound sensor which acts as mobile sensor on the vehicle. Traffic measurement was performed by counting the vehicles which were passing each other on the opposite lane. This method used the spectral estimation by LPC analysis and the learning by SOM algorithm. This method recognizes the states by comparing the winner's position with the map produced by learning phase.



(a) $t = 17.79$ [sec.] (b) $t = 18.90$ [sec.] (c) $t = 19.29$ [sec.]

Fig. 9. The workings of this system at the recognition phase.

There is room for improvement in terms of accuracy because the selection of parameters that are used for SOM learning is not robust. From now on, the detailed observation of the influence of the parameter's change is needed to raise the accuracy of traffic measurement.

References

[1] H.Kuboyama, S.Ozawa. Measurement of Heavy Traffic in a Tunnel from Image Sequences, *IEICE Trans.*, Vol.J85–DII, No.2, pp.210–218, 2002.

[2] H.Taniguchi, A.Seki, H.Furusawa, S.Kuroda, S.Ikehata. A Method of Motion Analysis Using Spatio–Temporal Image —Directional Temporal Plane Transform—, *IEICE Trans.*, Vol.J77–DII, No.10, pp.2019–2026, 1994.

[3] I.Naka, Y.Noro, K.Kuno. Relation between road traffic volume and percentile noise Level L_α —A study of L_α based on Q–V curve—, *J. Acoust. Soc. Jpn.*, Vol.56, No.5, pp.301–307, 2000.

[4] M.Komatsu, T.Kato, S.Kuwano, S.Namba, Y.Hayashi. The masking effect of road traffic noise by using of wind–induced foliage sounds, *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp.657–658, 2000.

[5] K.Taguchi, H.Obata, S.Minamisawa. Discrimination of vehicle types based on time–varying spectral patterns of traffic noise, *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp.777–778, 2002.

[6] S.Furui. *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker Inc, New York, 1989.

[7] K.Akizuki, Y.Matsuyama, O.Yoshie. *C Language Digital Signal Processing*, Baifukan, Tokyo, 1997.

[8] T.Kohonen. *Self-organizing maps*, Springer–Verlag, New York, 1995.

[9] A.Ultsh. Self organized feature maps for monitoring and knowledge aquisition of a chemical process, *Proc. ICANN'93 Int. Conf. on Artificial Neural Networks*, pp.864–867, 1993.