RISP
NCSP'16
International Workshop on
Nonlinear Circuits, Communications and Signal Processing
Honolulu, Hawaii, USA, March 6-9, 2016

Vehicle Counting and Lane Estimation with ad-hoc Microphone Array in Real Road Environments

Takuya Toyoda[1], Nobutaka Ono[2], Shigeki Miyabe[1], Takeshi Yamada[1], and Shoji Makino[1]

[1] University of Tsukuba
1–1–1 Tennodai, Tsukuba, Ibaraki, 305-8577 Japan
Phone/FAX:+81-29-853-6432/+81-29-853-7387
E-mail: toyoda@mmlab.cs.tsukuba.ac.jp,
{miyabe, maki}@tara.tsukuba.ac.jp,
takeshi@cs.tsukuba.ac.jp

[2] National Institute of Informatics /
SOKENDAI (The Graduate University for Advanced Studies)
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan
Phone/FAX:+81-03-4212-2827/+81-03-4212-2699
E-mail: onono@nii.ac.jp

Abstract

In this paper, we present a new traffic lane estimation method that involves analyzing the difference between the power peaks of channels in an acoustic traffic monitoring technique based on an ad-hoc microphone array. Then, we apply traffic monitoring systems involving two traffic lane estimation methods to data recorded in various real road environments. With our proposed traffic monitoring systems, we perform channel synchronization based on single source activities, estimate the number of vehicles by analyzing the power envelope of each channel and detecting the peaks, and estimate the traffic lane in which a vehicle is moving by analyzing the time differences or power differences between the power peaks of the channel. We demonstrate the effectiveness of our proposed method by comparing the experimental results obtained with our proposed traffic monitoring systems in various real road environments.

1. Introduction

Monitoring traffic is important in relation to easing traffic congestion. Fixed-point observation techniques employ loop type sensors, ultrasonic sensors, infrared sensors, seismic sensors and movie cameras.However, these sensing systems suffer from high installation and maintenance costs. In addition, the accuracy of vehicle detection with these sensing systems depends heavily on the sensing conditions. For example, with image processing, the quantity of data becomes enormous and accuracy deteriorates at night and in bad weather. As a result, there has been a need for an easy method of monitoring traffic.

An effective monitoring approach should be inexpensive to install and maintain and have little dependence on the sensing environment. Several studies based on acoustic sensing have already attempted to meet these requirements [1, 2, 3, 4].

With the aim of realizing an easier and simpler sensing system, we proposed an acoustic traffic monitoring technique [5] based on an ad-hoc microphone array [6, 7] that combines independent recording devices for multichannel recording. The advantage of this approach is that its combination of small recording devices makes it easy to install. In this paper, to confirm the effectiveness of our proposed acoustic traffic monitoring technique, we apply our two proposed traffic monitoring systems to extensive data. The two systems differ from the traffic lane estimation methods, and the other processing techniques are the same in these systems. To synchronize the asynchronous recording channels [8, 9, 10], our proposed system first compensates for sampling mismatches by using single source activity [10]. The system then analyzes the power envelope of each channel and detects the peaks to estimate vehicle travel. Finally, one of the systems analyzes the time differences between the power peaks of the channels, as reported in [5], and the other system analyzes the power differences between the peaks of the channels. We compare the two proposed traffic monitoring systems by performing experiments in various real road environments. Our experimental results confirm the effectiveness of our proposed traffic monitoring system.

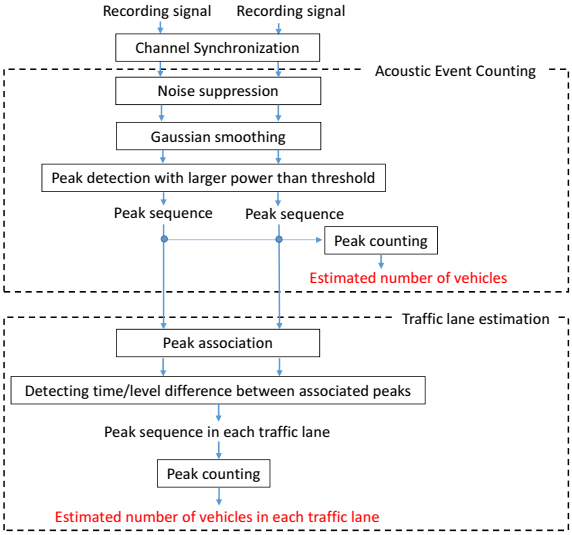

Figure 1: Our traffic monitoring system

2. Our Proposed Traffic Monitoring System

Fig. 1 shows our proposed traffic monitoring system. To make it possible to install and maintain recording devices inexpensively, our traffic monitoring system records the sounds made by moving vehicles using asynchronous recording devices. So, channel synchronization is necessary with our systems.

2.1 Channel synchronization

First, we perform channel synchronization by using two intervals of a recording of a single source activity as proposed in [10]. At the beginning and end of the recording, we generate chirp signals and use these signals as single source activities. Although the times at which the single source activities were recorded vary in each channel, the times at which they were generated were the same. So, we perform channel synchronization by compensating for the sampling frequency mismatch and the difference between the beginning and end of the recording based on the time relation between the single source activities in each channel.

2.2 Acoustic event counting

This subsection describes a method for estimating the number of vehicles according to peak detection of the power envelopes. Since the captured sound power is expected to reach its maximum level when vehicle is passing in front of microphone, we try to count the number of vehicles by detecting the peaks of the captured sound. However, due to the influence of background noise and fluctuations in the observations, many spurious peaks appear in the raw data. Therefore, it is necessary to suppress the background noise and apply smoothing to the power before peak detection.

2.2.1 Noise suppression

Since the sound of a moving vehicle and background noise have different spectra generally, we try to suppress the noise by using a linear time-invariant Wiener filter. We manually find the frames where the vehicle is moving and not moving, and let $S_V(\omega)$ and $S_N(\omega)$ be the average moving sound of a vehicle and the ambient noise power spectra, respectively. Here $\omega = 1, \cdots, \Omega$ denotes the discrete frequency index. We designed the following Wiener filter $W(\omega)$, assuming the stationarity of the power of the moving sound of a vehicle and ambient noise, to enhance the bands with a high SNR.

$$W(\omega) = \frac{S_V(\omega)}{S_V(\omega) + S_N(\omega)}. \tag{1}$$

Then, let $X_i(\omega, k)$ be the observed signal at the i-th microphone and an angular frequency ω in the k-th frame, and the noise suppressed signal $\hat{X}_i(\omega, k)$ is given as

$$\hat{X}_i(\omega, k) = W(\omega) X_i(\omega, k). \tag{2}$$

2.2.2 Gaussian smoothing

To count acoustic events, we consider signal power time series $Y_i(k)$ given by

$$Y_i(k) = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \hat{X}_i(\omega, k) \hat{X}_i^*(\omega, k), \tag{3}$$

where $\{\cdot\}^*$ denotes a complex conjugation. Even though a straightforward way is to count acoustic events directly from $Y_i(k)$, minute frequent fluctuations appear over $Y_i(k)$, which disturb the count. Hence, to reduce such frequent fluctuations, we calculate the power envelopes by smoothing. We perform smoothing with a Gaussian-window-shaped filter $g(m)$ given by

$$g(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{m^2}{2\sigma^2}\right), \tag{4}$$

where m is the time frame, and σ is the standard deviation of the Gaussian window. The signal smoothed by the Gaussian-window-shaped filter $\hat{Y}_i(k)$ is given by

$$\hat{Y}_i(k) = \sum_{m=\frac{-G}{2}}^{\frac{G}{2}} Y_i(k-m) g(m), \tag{5}$$

where G is the width of the Gaussian window.

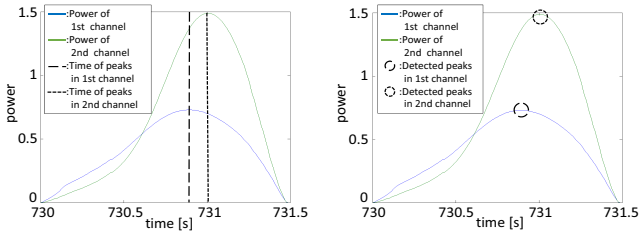

Figure 2: Time difference (left) and level difference (right) of power peaks between channels

2.2.3 Peak detection with larger power than threshold

Even with such smoothing, many peaks with small values remain. To ignore such small peaks, we discard those with powers smaller than a certain threshold. If $\hat{Y}_i(k)$ is the maximum value and exceeds h, let the time of peak k be $p_i(e)$, where h stands for the threshold, $p_i(e)$ for $e = 0, \cdots, E-1$ stands for the time of the detected peaks at the i-th microphone, and E is the total number of detected peaks. Then we estimated the number of vehicles by counting the number of peaks.

2.3 Traffic lane estimation

This subsection describes a method for estimating in which traffic lane the vehicle is moving. We use signals recorded by microphones that are installed in approximately the same position on either side of two-lane roads.

2.3.1 Peak association

As a first step, we associate two detected peaks at each microphone derived from the same vehicle. If $|p_1(e) - p_2(e)|$ is less than d, let $p_1(e)$ be $\hat{p}_1(e)$, and $p_2(e)$ be $\hat{p}_2(e)$, where $p_i(e)$ stands for the powers of detected peaks that are larger than the threshold in each channel, $\hat{p}_i(e)$ represents the power of associated peaks. Range d is an arbitrary constant that is determined in consideration of the speed of the moving vehicle, the width of the road, the distance between microphones, the sampling frequency, and the frame length of STFT. When there are many peaks meeting those conditions, we choose peaks with the smallest time difference.

2.3.2 Detecting time/level difference between associated peaks

Since the distance to each microphone varies depending on the traffic lanes, there are the differences between the propagation times and powers of the two microphones, according to the traffic lane in which the vehicle is moving as shown in Fig. 2. So, we try to classify the traffic lanes in which the vehicle is moving using two methods.

The first method is to classify the traffic lanes by using the differences between the propagation times of the channels (Method 1). Then, we classify the traffic lane in which the vehicle is moving with the following procedure. If $(\hat{p}_1(e) - \hat{p}_2(e))$ is less (more) than 0, we classify the lane as the left (right) traffic lane.

The second method is to classify the traffic lanes by using the differences between the power peaks of the channels (Method 2). Then, we classify the traffic lane in which the vehicle is moving with the following procedure. If $\frac{\hat{p}_1(e)}{\hat{p}_2(e)}$ is less (more) than 1, we classify the lane as the left (right) traffic lane.

Figure 3: Photograph of recording setup in Tsukuba (left) and Tokyo (right)

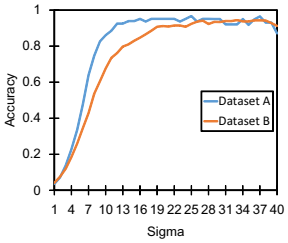

Figure 4: Relationship between accuracy of traffic lane estimation (Method 2) and smoothing parameter

3. Experiment in Various Real Road Environments

To confirm the effectiveness of our proposed traffic monitoring systems, we evaluated the estimation accuracy of our proposed methods utilizing the sounds of vehicles moving recorded in various real road environments.

3.1 Recording in various environments

When recording, we took account of traffic and vehicles running at a constant speed, because the estimation accuracy may be dependent on the influence of these factors. Therefore, we recorded the sounds of moving vehicles with ad-hoc microphone pairs and a video to obtain the correct road traffic information for two-lane roads in two locations (Tsukuba and Tokyo), which have different traffic as shown in Fig. 3. Then, recording devices were installed between two signals so that the vehicles would run at a constant speed, and in approximately the same position on either side of two-lane roads. Table 1 shows the recording conditions.

In our experiments, in Tsukuba, we clapped our hands at the center of the road and utilized these sounds as single source activities. However, in Tokyo, clapping hands at the center of the road was difficult, because the traffic was busy and there was insufficient time to stand at the center of road. So, at the beginning and end of recording, we gathered all the microphones and generated a chirp signal with a speaker, and utilized these sounds as single source activities. Then, we utilized devices that fixed the positions of the microphones and speakers in relation to each other.

Table 1: Experimental datasets recorded in Tsukuba and Tokyo

Dataset	A	B
Recording location	Tsukuba	Tokyo
Road width [m]	7.1	10.7
Distance between microphones [m]	9.5	10.7
Length of data [sec]	660	728
Number of vehicles in 1 minute	6	18
Number of vehicles that passed each other	8	94
Total number of vehicles	64	220
Recording devices	SANYO ICR-PS603RM	
Video camera	SONY HDR-CX420	

Table 2: Experimental conditions of datasets A and B

Sampling frequency [kHz]	48
Frame lengths of STFT [samples]	2048
Frame shift widths of STFT [samples]	512
Standard deviation σ of Gaussian window	25
Width of Gaussian window	6σ+1
Threshold h	0.07 (Dataset A) or 0.01 (Dataset B)
Range d of peak association [frames]	38

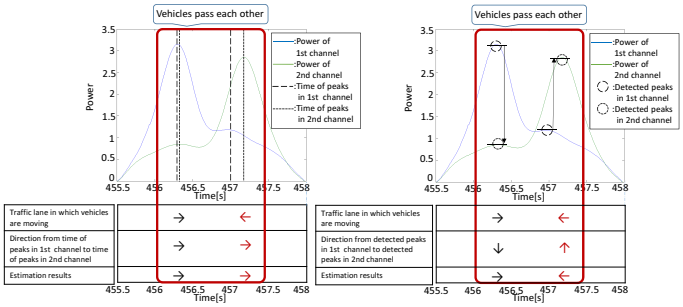

Figure 5: Experimental result for Method 1 (left) and Method 2 (right) using dataset A

3.2 Experimental conditions

We conducted an experiment to compare estimations obtained with our proposed acoustic sensing techniques and the correct road traffic information regarding the number of moving vehicles and the traffic lane in which they were moving.

We made ground-truth data based on the video and the temporal waveform of the sound power. After omitting intervals containing the sounds of motorcycles or bicycles, we estimated the number of vehicles and their traffic lanes from the recorded sounds. Table 1 shows the experimental data. Datasets A and B indicate the sounds recorded in Tsukuba and Tokyo, respectively, they were used in the experiment.

We experimentally determined parameter h of datasets A and B, because h is influenced by ambient noise in each recording environment. Then we experimentally determined parameter σ using dataset A, and we used the same value in the experiment using dataset B, because σ is a parameter related to smoothing the sounds of moving vehicles and is not strongly influenced by the recording environment. Fig. 4 shows the relationship between the accuracy of the traffic lane estimation (Method 2) and the smoothing parameter of datasets A and B. In the experiments using dataset B, σ is not optimal. So, if we determine σ experimentally in each recording environment, the estimation accuracy can be increased. The other experimental conditions are shown in Table 2.

For the evaluation, we used a one-second interval as a tolerance, which means that detection within ± 0.5 second from the ground truth was counted as true detection. We evaluated our proposed acoustic sensing technique using the F-measure of acoustic event counting and precision of traffic lane estimation obtained from our comparison of the estimation and the correct traffic information provided by the reference video.

3.3 Evaluation of acoustic event counting

F-measure is a measure used for the comprehensive evaluation of accuracy and completeness. So, we use F-measure to evaluate the accuracy of acoustic event counting.

$$\text{precision} = N_c/N_e \tag{6}$$

$$\text{recall} = N_c/N_r \tag{7}$$

Table 3: Numbers of vehicles in the experimental acoustic event counting results

(a) Dataset A

		Ground truth	
		T	F
Detected vehicles	T	58	3
	F	6	

(b) Dataset B

		Ground truth	
		T	F
Detected vehicles	T	146	8
	F	74	

Table 4: Estimated numbers in the experimental traffic lane estimation results

	Correct estimation	Total estimation
Method 1 using dataset A	56	61
Method 2 using dataset A	59	61
Method 1 using dataset B	92	149
Method 2 using dataset B	134	149

Table 5: Evaluation of accuracy of acoustic event counting and traffic lane estimation using datasets A and B with F-measure and Accuracy

	F-measure of acoustic event counting	Precision of traffic lane estimation
Method 1 using dataset A	0.9280	0.9180
Method 2 using dataset A		0.9672
Method 1 using dataset B	0.7480	0.6174
Method 2 using dataset B		0.8993

$$\text{F-measure} = \frac{\text{precision} \cdot \text{recall}}{\frac{1}{2}(\text{precision} + \text{recall})} \quad (8)$$

where N_c is the correctly estimated number, N_e is the total estimated number, and N_r is the total true number. F-measure has a value range of 0 to 1, and a higher value means better accuracy.

3.4 Evaluation of traffic lane estimation

We use only precision to evaluate the traffic lane estimation, because the accuracy of that have nothing to do with completeness. In this evaluation, we use the experimental results that were estimated in each traffic lane.

3.5 Experimental result and consideration

Fig. 5 shows estimation results obtained with the proposed acoustic sensing technique at certain time intervals using dataset A. The results show that Method 2 could estimate the traffic lane correctly when vehicles pass each other, on the other hand Method 1 estimated the traffic lane mistakenly when vehicles passed each other, because we could not obtain the time differences between two channels calculated from the positional relationships between vehicles and microphones when the vehicles passed each other in front of the microphones. It is considered that the sounds generated by moving vehicles are a combination of engine sound and the sound produced by the vehicle's tires, and these are moving sounds.

Table 3 shows the total true number of vehicles, the number of estimated vehicles and the number of correctly estimated vehicles counted for the two criteria with our proposed acoustic event counting method using datasets A and B. Table 4 shows the number of correct estimations and the number of total estimations for our proposed traffic lane estimation method using datasets A and B. We evaluated the results of acoustic event counting and traffic lane estimation results with F-measure and precision, respectively. Table 5 shows the F-measure for acoustic event counting and precision for traffic lane estimation.

We compared the results of the evaluation acquired experimentally using datasets A and B. In Methods 1 and 2, F-measure calculated using experimental results obtained with dataset B was less than that calculated using experimental results obtained with dataset A, because the number of vehicles that pass each other in dataset B was greater than that in dataset A, and the ratio of the number of vehicles that pass each other of the total number of vehicles was about 22%. In Method 1, precision acquired experimentally using dataset A achieved high values, however precision acquired experimentally using dataset B was much lower than that acquired experimentally using dataset A, because the number of vehicles that pass each other in dataset B greatly exceeded that of dataset A. In Method 2, precision acquired experimentally using dataset A achieved high values, and precision acquired experimentally using dataset B achieved high values comparable to that acquired experimentally using dataset A. These results confirmed the effectiveness of our proposed acoustic sensing technique (Method 2), based on peak detection and an analysis of the differences between the power peaks of the channels, because high accuracy was obtained.

4. Conclusion

In this paper, we proposed two traffic monitoring system with an ad-hoc microphone array. These systems are different from the traffic lane estimation method. One of the methods was based on the differences between the propagation time of the channels, the other method was based on the differences between the power peaks of the channels. We applied our two proposed systems to extensive data, and evaluated the experimental results. The high accuracy of the experimental results confirmed the versatility of our proposed traffic monitoring system.

5. Acknowledgment

This work was supported by a Grant-in-Aid for Scientific Research (B) (Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25280069).

References

[1] N. Shimada *et al.*, "A study on an approaching vehicle detection using a linear microphone array-based acoustic sensing," *IEICE Technical Report SIS2009-71*, pp. 125–128, 2010. (in Japanese)

[2] Y. Nooralahiyan *et al.*, "A field trial of acoustic signature analysis for vehicle classification," *Transpn Res-C*, vol. 5, no. 3/4, pp. 165–177, 1997.

[3] M. Sobreira *et al.*, "Automatic classification of traffic noise," *Proc. IOA*, pp. 6221–6226, 2008.

[4] N. Evans *et al.*, "Automated acoustic identification of vehicles," *Proc. IOA*, pp. 238–245, 2008.

[5] T. Toyoda *et al.*, "Traffic monitoring with ad-hoc microphone array, " *Proc. IWAENC*, pp. 319–323, 2014.

[6] E. Robledo-Arnuncio *et al.*, "On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation," *Proc. WASPAA*, pp. 34-37, Oct. 2007.

[7] Z. Liu "Sound source separation with distributed microphone arrays in the presence of clock synchronization errors," *Proc. IWAENC*, 2008.

[8] S. Markovich-Golan *et al.*, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," *Proc. IWAENC*, 2012.

[9] S. Miyabe *et al.*, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," *Proc. ICASSP*, pp. 674–678, 2013.

[10] R. Sakanashi *et al.*, "Speech enhancement with ad-hoc microphone array using single source activity, " *Proc. APSIPA*, pp. 1–6, 2013.