An Acoustic Traffic Monitoring System: Design and Implementation

Yueyue Na, Yanmeng Guo, Qiang Fu, *Member, IEEE*, and Yonghong Yan
The Key Laboratory of Speech Acoustics and Content Understanding
Institute of Acoustics, Chinese Academy of Sciences
Beijing, China
{nayueyue, guoyanmeng, qfu, yanyonghong}@hccl.ioa.ac.cn

***Abstract*—Vehicle emits sounds as it travels along the road, which can be used for traffic monitoring. In this paper, an acoustic based traffic monitoring system is designed and implemented. The system utilizes a cross microphone array to collect road-side acoustic signals. Then, lane positions are automatically detected by the built-in lane detection module. Eventually, different measuring indices which reflect the road condition and traffic quality are derived according to the collected signals and the detected lanes. Since acoustic sensor is less expensive than other types of vehicle sensors, and acoustic features are robust against light, weather, and environmental variations, we expect that the proposed acoustic traffic monitoring system will have lower hardware cost, and become a good complement to the existing traffic monitoring techniques.**

Keywords—Intelligent transportation system; traffic monitoring; beamforming; vehicle counting; vehicle speed estimation

I. INTRODUCTION

In modern intelligent transportation system (ITS), many traffic monitors are distributed in the road networks, so that the information which indicates the real-time traffic conditions can be perceived and aggregated to the control center for traffic flow control and dynamitic planning [1]. Since traffic monitors are the indispensable building blocks of ITS, developing more reliable and accurate traffic monitoring techniques will facilitate modern ITS construction, and further improve traffic condition, public safety, and reduce transportation cost [1].

Vehicle emits acoustic sounds when it travels along the road, so, vehicular sound can be used as a kind of feature for traffic monitoring. Compared with other traffic monitoring techniques such as inductive loop [2, 3], radar [1, 4], infrared, magnetic field [5-7], video, etc., acoustic based traffic monitoring has many advantages. First, it is a nonintrusive technique, which will not cause damage to the pavement during the monitor installation and maintenance procedures. Second, it is a passive monitoring technique, which means that no harmful signals will be transmitted to human body. In addition, the hardware cost can be reduced since no signal transmitting device is needed. Meanwhile, the acoustic sensor (microphone) is inexpensive compared with other types of sensors [10, 11]. Third, acoustic features are very robust against light and weather variations, which is helpful for robust traffic monitoring [13].

A lot of researches have been carried out for acoustic based traffic monitoring algorithm design and system construction. E.g., in [8], a microphone array with four subarrays is built for multi-lane traffic monitoring. Channels within subarrays are first added together for signal enhancement purpose, then, cross correlations [9] among different subarrays are calculated to locate vehicles in different lanes. In [10], a single microphone approach is proposed, where the traffic density state is classified into three levels according to the features extracted from the collected road-side acoustic signals. Although this approach cannot perform accurate lane-wise vehicle counting and speed estimation, it is especially suitable for congested city road environment. In [11], an acoustic traffic monitoring system is built with linear microphone arrays mounted on a sign bridge across the road. In the system configuration, each lane is managed by a pair of microphones, the cross correlations are calculated among different microphone pairs for vehicle counting and speed estimation, then, the Mel-frequency cepstrum coefficient (MFCC) feature and hidden Markov model (HMM) are used for vehicle classification, just like the idea of speech recognition [12]. Although this system has high traffic monitoring accuracy, its application is limited by the requirement of a sign bridge for deployment. A commercial acoustic traffic monitoring system is reported in [13]. In this system, up to five lanes can be simultaneously managed, traffic indices such as vehicle count, lane occupancy, and average speed can be derived for each lane. However, a planar microphone array with many array elements is used in this system, which makes the hardware cost still high.

In this paper, an acoustic based traffic monitoring system is introduced. This system first utilizes a cross microphone array to collect road-side vehicular sounds, then, its lane detection module is used to automatically detect lane positions based on the statistical information of the passing by vehicles. Four types of traffic indices, including vehicle count, lane occupancy, vehicle speed, and vehicle type (large vs. small) are derived according to the collected signals and detected lane positions. Our contributions in this paper are: (1) a traffic monitoring

This work is partially supported by the National Natural Science Foundation of China (Nos. 11461141004, 91120001, 61271426), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant Nos. XDA06030100, XDA06030500), the National 863 Program (No. 2012AA012503) and the CAS Priority Deployment Project (No. KGZD-EW-103-2).

978-1-4673-7211-4/15 $31.00 © 2015 IEEE
DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.41

prototype system is designed and developed, (2) an approach for coarse vehicle detection zone construction is introduced, (3) a novel vehicle counting approach is proposed, and (4) a vehicle speed estimation approach is proposed. The rest content of this paper is organized as follows. The basic idea of acoustic traffic monitoring and system overview is introduced in section II. The microphone array and the lane detection strategy are briefly introduced in section III. Then, the coarse detection zone construction algorithm for lane-wise traffic monitoring is depicted in section IV, and the derivation of different traffic indices is depicted in section V. The proposed system is tested in real-world environment, the experimental configuration and result is reported in section VI. At last, we conclude this paper in section VII.

II. Acoustic Traffic Monitoring

A. Acoustic Traffic Monitoring Preliminaries

As shown in Fig. 1, the acoustic traffic monitor is mounted on a road-side structure above the road surface, so, it can be installed without interfering with the traffic, cutting the road surface, or requiring an overhead sign structure. Vehicle emits sounds (consists of engine noise, tire noise, exhaust noise, air turbulence noise, etc. [14]) as it travels along the road. The sound signals are captured and analyzed by the monitor, then, lane positions will be automatically detected, and different traffic statistical indices, such as vehicle count, lane occupancy, vehicle speed and average speed, vehicle type (large vs. small) will be derived.

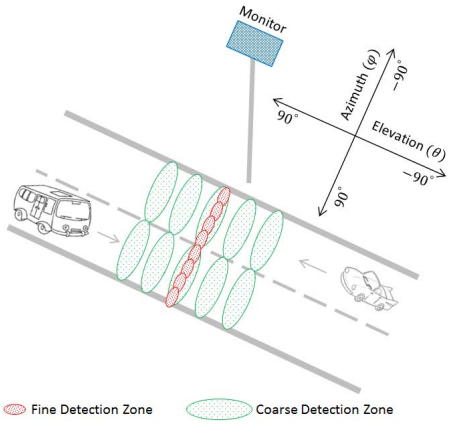


Fig. 1. Acoustic traffic monitoring.

To detect vehicles in different positions, the monitor utilizes microphone array and beamforming [15] techniques to form different kinds of detection zones in different look directions for different purposes. Because of the spatial filtering property of the beamforming, sound signals other than the array look direction will be attenuated by the beamformer, so that vehicle locations can be sensed from the cumulated energy of the beamformer output. There are two kinds of detection zones constructed in Fig. 1: fine detection zones (red ellipses) and coarse detection zones (green ellipses). Fine detection zones are used for lane detection, they have smaller width and fixed positions, multiple fine detections zones are concatenated to form a cross section across the road. Vehicle azimuth will be detected by the system as it passes through this cross section, and the azimuth statistics will be used to detect lane positions. Coarse detection zones are used for lane-wise traffic monitoring, their width and positions will be adaptively updated according to the detected lane positions. Each lane is managed by multiple coarse detection zones with a small elevation apart from each other, different traffic indices will be derived from the energy response curves outputted by these detection zones.

B. System Overview

The proposed acoustic traffic monitoring system comprises of four main modules: microphone array, lane detection module, lane-wise vehicle monitoring module, and user interface. The block diagram of the system is shown in Fig. 2, where blocks indicate modules and submodules, and arrows indicate data flow. The main function of each module is introduced as follows:

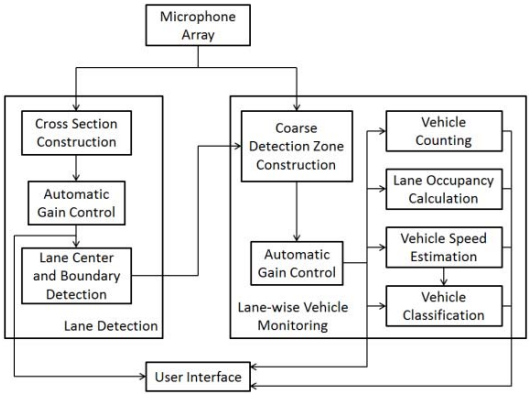


Fig. 2. The acoustic traffic monitoring system block diagram.

Microphone array: used to collect vehicular sounds, perform short time Fourier transform (STFT) [16] and data preprocessing, then, accumulate correlation matrices in the selected working frequency band. The correlation matrices are the output of the microphone array module, which will be used by the following modules for lane detection and traffic monitoring.

Lane detection module: used to detect lane positions, which contains three submodules. First, fine detection zones and the lane detection cross section is constructed by the cross section construction module, the direction-of-arrival (DOA) spectrum of the passing by vehicle is also calculated by this submodule. Then, the automatic gain control (AGC) submodule normalizes the spectrum into the range of [0, 1], which is suitable for data visualization and vehicle appearance decision. At last, the lane center and boundary detection submodule detects the vehicle azimuth from the normalized DOA spectrum, and finds lane centers and boundaries from the azimuth statistics.

After the lanes are detected, the lane positions are passed to the lane-wise vehicle monitoring module, which contains six submodules. First, multiple coarse detection zones are constructed by the coarse detection zone construction submodule according to the detected lane positions, this submodule also outputs the individual zone responses using the microphone array data. Then, the response curves are normalized into the interval of [0, 1] by another AGC submodule, and the normalized curves are used for the calculation of different traffic indices.

Finally, all intermediate results are aggregated to the user interface module for data visualization, such as acoustic traffic imaging (ATI) and coarse detection zone response curves. All calculated traffic indices are also displayed on the user interface for traffic control and dynamic planning, in addition, the user interface is also responsible for the communication between the monitoring system and the user or the host computer.

III. Microphone Array and Lane Detection

The purpose of this paper is to provide a systematic view of the acoustic traffic monitoring system. However, there are still many things to be concerned for microphone array design and lane detection, such as array resolution, aperture size, array topology, number of array elements and element spacing, vehicle DOA estimation, AGC, lane center and boundary detection, etc. Talking too much about these issues will dilute the topic of this paper. So, for completeness, this section only introduces the basic idea of microphone array design and lane detection, for the detailed discussion about these issues, please refer to another of our paper: "Cross Array and Rank-1 MUSIC Algorithm for Acoustic Highway Lane Detection".

A. Microphone Array

The microphone array used in the proposed system is based on the cross array [17-19] structure, Fig. 3 shows the adopted array topology. This array has the aperture size of 20 cm x 32 cm (width x height), and 37 array elements are used. The two axis of the cross array are consist of two uniform linear arrays (ULA), which are called the horizontal and vertical subarrays in this paper. The two subarrays share the common phase center, but have different number of elements and element spacing. Please notice that the array in Fig. 3 is used for experiments, and it is still not the optimal scheme for the actual product. This array uses more elements and denser spacing to leave us enough margins for working frequency band selection. The final array topology will be optimized according to the experimental results to further reduce the required elements.

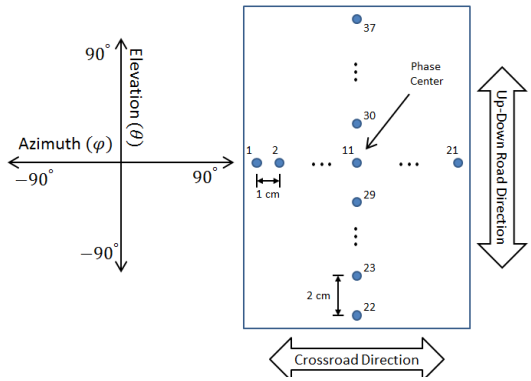


Fig. 3. The microphone array for the prototype system. The array is facing to the reader.

After STFT is performed on each channel to convert the collected time domain sound signal to time-frequency domain, the output of the microphone array module can be denoted as the cross correlation matrices $\boldsymbol{C}$ of the two subarrays, as shown in equation (1), where $\boldsymbol{x}$ stands for the collected multi-channel data, subscripts h and v stand for the horizontal and vertical subarray, f and τ for frequency bin index and STFT frame index, and $\Delta\tau$ is the selected time interval for the correlation ($\Delta\tau$ equals to the number of frames corresponding to 200 ms in our experiments).

$$\boldsymbol{C}(f) = \sum\nolimits_{\tau\epsilon\Delta\tau} \boldsymbol{x}_h(f,\tau)\boldsymbol{x}_v^{\mathrm{H}}(f,\tau) \tag{1}$$

In our system, the energy responses of different detection zones in frequency bin f are calculated according to (2), where $\boldsymbol{w}_h$ and $\boldsymbol{w}_v$ are the beamformers for the horizontal and vertical subarrays.

$$q(f) = \left|\boldsymbol{w}_h^{\mathrm{H}}(f)\boldsymbol{C}(f)\boldsymbol{w}_v(f)\right|^2 \tag{2}$$

Then, the response of the total working frequency band is calculated as the geometric mean of the individual responses [20], as shown in (3), where Δf is the interval of the working frequency band, and $|\Delta f|$ is the number of frequency bins contained.

$$q = \left[\prod\nolimits_{f\in\Delta f} q(f)\right]^{1/|\Delta f|} \tag{3}$$

Since the beamforming procedures in (2) are the same for all frequency bins, the frequency bin index f will be omitted in the following of this paper to make equations more concise.

There are three main benefits to use the cross array scheme and the preprocessing procedure in (1). First, compared with the rectangular array scheme in [13], the number of required array elements is greatly reduced for the array with the same aperture size, so, both hardware cost and system complexity can be reduced. Second, the beamformer design problem is decoupled into two independent sub-problems with smaller size and easier structures, as shown in (2), which further simplifies the calculation. In addition, different beamforming techniques can be separately applied to the horizontal and vertical subarrays to form different vehicle detection zones. In our system, the delay-and-sum (DS) beamforming with Dolph-Chebyshev taper [21] is used to construct the vertical beamformers. The horizontal beamformers for fine and coarse detection zones are different. For fine detection zones, the beamformers are designed with the subspace techniques [22], while the idea for coarse detection zone beamformer design will be depicted in section IV. Third, multiple detection zones can be constructed upon the single correlation matrix $\boldsymbol{C}$ with different horizontal and vertical beamformers. One detection zone can be considered as a "virtual vehicle sensor", the monitoring performance is expected to be improved when more "virtual sensors" are used. However, unlike the inductive loop approach in [2] and the magnetic sensor approach in [5], no additional physical sensors are required to add more detection zones in our system.

B. Lane Detection

Usually, vehicle travels along the lanes and seldom across the lane boundaries, thus, the observed sound energy at the array look directions corresponding to lane centers is high, while the energy corresponding to lane boundaries is low. The lane detection module uses such simple but reasonable assumption to detect lanes. The detected lane positions can be considered as a kind of frequent pattern indicating the vehicle appearance locations in the lane detection cross section, it is unneces-

sary to align the detected lane positions to the actual lane center and boundary look directions of the road [1].

Generally speaking, the lane detection procedure contains three steps. First, the submodule generates fine detection zones and constructs the lane detection cross section. Meanwhile, the DOA estimation algorithms are used to detect the azimuths or the energy spectrums of the passing by vehicles. Second, the azimuth statistics or energy spectrums are accumulated to fit the probability density function (PDF) of vehicle locations in the cross section. The resulted PDF has high values at the positions where vehicle frequently appears. Third, the peaks of the PDF are detected as lane centers, while the valleys at the both sides of a peak are detected as the corresponding lane boundaries.

Fig. 4 gives an example of the estimated vehicle azimuth PDF and the detected lanes in a simulated Chinese four-lane bidirectional highway environment. In the environment of Fig. 4, the monitor is mounted near lane 1, so, the four detected lanes have decreasing widths. The resolution of the beamforming algorithm is a very important problem which should be carefully concerned in the system design. The angular resolution of the beamformer should be high enough to distinguish vehicles travelling in different lanes. From the real-world experiment in section VI we will see that the resolution of the proposed system is adequate for one-side traffic monitoring in a six-lane bidirectional highway system.

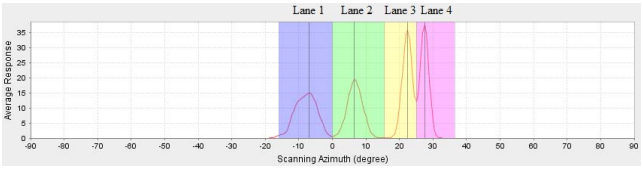


Fig. 4. The demonstration of four detected lanes.

IV. Coarse Detection Zone Construction

A. The Horizontal Beamformer

After lanes are detected, multiple coarse detection zones can be constructed according to the lane positions for lane-wise traffic monitoring. For a single coarse detection zone, since the vertical beamformer is constructed by the fixed beamforming technique, here we mainly concern about the horizontal beamformer construction.

The main difficulty for horizontal beamformer construction is the side-lobe problem. In the ideal case, we expect that when a vehicle passes by, only the zones in the corresponding lane have significant responses, while zones in other lanes have zero responses. However, all beamforming techniques may leave side-lobes [9] at the directions other than the look direction, and they will cause responses from the zones in other lanes. Since the sound pressure levels of different vehicles vary a lot, and there are multiple vehicles simultaneously traveling on the road, it is difficult for a coarse detection zone to distinguish between a real vehicle response and the ghost image caused by its side-lobes. Thus, the constructed horizontal beamformer for a coarse detection zone should not only have high resolution to distinguish vehicles in adjacent lanes, but also have small side-lobes to prevent ghost images.

The acoustic traffic monitoring system in [13] uses the famous minimum variance distortionless response (MVDR) beamforming [23, 24] to construct coarse detect detection zones. This system uses a rectangular microphone array, and the multichannel time domain signals are firstly added together according to columns to form the horizontal subarray data (equivalent to the DS beamforming at $\theta = 0$), as shown in equation (4) and (5), where t is the time domain sample index, i and j are row and column indices of the rectangular array, J is the number of columns.

$$x_j(t) = \sum_i x_{ij}(t) \tag{4}$$

$$\boldsymbol{x}_h(t) = \left[x_1(t), \cdots, x_J(t)\right]^{\mathrm{T}} \tag{5}$$

Then, after STFT, the correlation matrix $\boldsymbol{C}_h$ in frequency domain can be constructed according to (6), and the MVDR beamformer is constructed as (7), where $\mathrm{E}\{\cdot\}$ for expectation, $\boldsymbol{a}_h$ is the horizontal subarray steering vector which indicates the look direction of the monitored lane center.

$$\boldsymbol{C}_h = \mathrm{E}\{\boldsymbol{x}_h \boldsymbol{x}_h^{\mathrm{H}}\} \tag{6}$$

$$\boldsymbol{w}_h = \frac{\boldsymbol{C}_h^{-1}\boldsymbol{a}_h}{\boldsymbol{a}_h^{\mathrm{H}}\boldsymbol{C}_h^{-1}\boldsymbol{a}_h} \tag{7}$$

MVDR beamforming is a kind of adaptive beamforming technique, in which the nulls of the constructed beamformer can adaptively be steered to the positions of the side-lobes according to the observed signal correlation matrix, so that the side-lobes can be well attenuated [9]. However, there are two drawbacks for the preceding approach in the system of [13]. First, the beamformer should be updated every STFT frame, while the matrix inversion operation in (7) has cubic computational complexity, which increases the system burden. Second, according to (4) - (6), only detection zones at $\theta = 0$ can be constructed, which limits the flexibility of the system. Although MVDR beamforming has good side-lobe attenuation ability, since the cross array is used, in the proposed system, it is impossible to construct correlation matrices like in (4) - (6). Therefore, we cannot construct MVDR beamformers directly according to (7). Instead, based on the idea of (7), we propose a simple but effective approach to construct horizontal beamformers as follows.

B. The Lane Model Approach

After lanes are detected, the lane centers and boundaries not only tell us the lane positions, but also tell us the energy distribution of the sound field in the lane detection cross section. From the result of Fig. 4 we can see that, the sound energy in each lane can be approximately modeled by a Gaussian distribution. Thus, the sound field in the cross section can be modeled as a weighted diffuse sound field [25], whose correlation matrix $\boldsymbol{R}$ can be calculated according to (8), which means that the sounds are coming from the directions following the energy distribution of all lanes.

$$\boldsymbol{R} = \int_{\varphi=-90}^{90} g(\varphi)\boldsymbol{a}_h(\varphi)\boldsymbol{a}_h^{\mathrm{H}}(\varphi)d\varphi \tag{8}$$

The weight function $g(\varphi)$ in (8) is generated according to the detected lanes as depicted in (9) and (10), where φ_l is the cen-

ter of lane l, and the standard deviation parameter σ_l can be derived according to the "three sigma law" of the Gaussian function from the detected lane width.

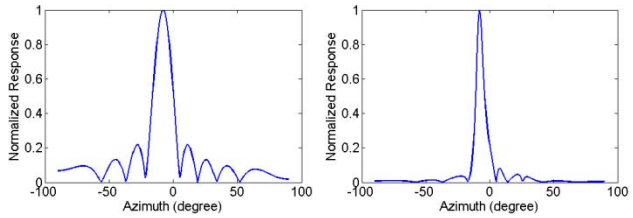
$$g_l(\varphi) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left[-\frac{(\varphi - \varphi_l)^2}{2\sigma_l^2}\right] \tag{9}$$

$$g(\varphi) = \sum_l g_l(\varphi) \tag{10}$$

Once the lane positions are updated, a new correlation matrix based on the lane model can be calculated according to (8) - (10), and like the MVDR approach in (7), the horizontal beamformer can be constructed as (11). This approach is also called the maximum directivity beamforming [26].

$$\boldsymbol{w}_h = \frac{\boldsymbol{R}^{-1}\boldsymbol{a}_h}{\boldsymbol{a}_h^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{a}_h} \tag{11}$$

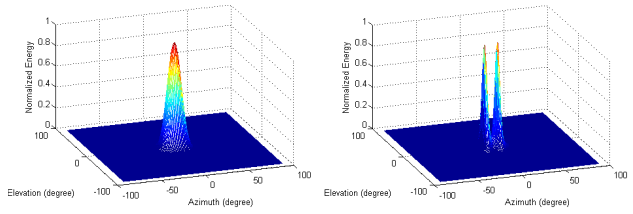
Compared with the MVDR approach in (7) we can find that, the approach in (11) is a sub-optimal approach, since the beamformer cannot adaptively change with the real-time sound field variations. However, this approach has lower computational complexity, since the matrix inversion is only performed when the lane positions are updated.



(a) the DS approach (b) the lane model approach

Fig. 5. The comparison of beampatterns for horizontal beamformers. The working frequency is 7k Hz. The array look direction is -7.9°, which is the first lane center in the simulated road environment in Fig. 4.

To further illustrate the effectiveness of the proposed lane model approach, Fig. 5 gives the beampattern comparison of the DS beamformer and the lane model beamformer. From Fig. 5 (a) we can see that the DS approach has larger main-lobe width and many significant side-lobes, so, it is not suitable for coarse detection zones. On the other hand, in Fig. 5 (b), the beampattern of the lane model approach has smaller main-lobe width and trivial side-lobs, which will effectively suppress the influence from the adjacent lanes.



(a) the DS approach (b) the lane model approach

Fig. 6. The comparison of energy responses. The working frequency is 7k Hz. The DOAs of the two sources are -7.9° and 6.6°, respectively, which are the centers of lane 1 and lane 2 in the simulated road environment in Fig. 4.

Fig. 6 shows the 3D responses when there are two vehicles in lane 1 and lane 2 at $\theta = 0$ in the simulated road environment. The response in Fig. 6 (a) has only one peak, which indicates the resolution of the DS approach is not enough to distinguish the two vehicles in adjacent lanes. However, there are two peaks in Fig. 6 (b), which means that the two vehicles are successfully identified by the lane model approach.

V. Traffic Indices

Four indices that reflect the traffic quality are calculated in our system, including: vehicle count, lane occupancy, vehicle speed, and vehicle type (large vs. small).

A. Vehicle Count

The main problem for vehicle counting is to determine whether there is a vehicle passed by. If a vehicle appears, just increases the vehicle count of the corresponding lane by 1, and initiates the calculation of the other traffic indices.

A naive approach to determine the vehicle appearance is to use the response curve of a single coarse detection zone. We say a zone is activated if its response first rises above then falls below a certain threshold, which indicates a vehicle first enters then leaves the detection zone. The detection threshold in our system is fixed since the response is normalized to [0, 1] by the AGC module. If any one of the detection zones in a lane is activated, the vehicle counting module counts the vehicle. However, there are significant interferences and noise in the real road environment, detection zones may accidentally be activated by unknown interference, which means that such naive approach may cause many false detections.

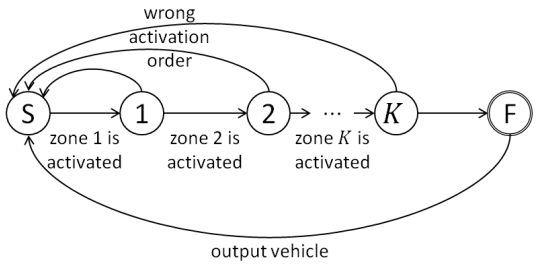


Fig. 7. The finite state machine for vehicle counting.

To robustly detect a passing by vehicle, we use all detection zones in a lane: a vehicle is detected if and only if those detection zones are activated sequentially. The idea of the finite state machine, which is also used for vehicle detection in [6], is extended in our system to record the zone activation order. As shown in Fig. 7, the finite state machine with $K + 2$ states is constructed in a chain structure, S is the initial state and F is the final state, while the other K states indicate the K detection zones. The feedforward arrow points to the expecting state of current state, if all detection zones are activated in order, the state can successfully be transferred from S to F, which indicates a vehicle is passed. Otherwise, if the activation order is wrong, the feedback arrows will reset the machine to its initial state. Although a single detection zone may accidentally be activated in a noisy environment, it is nearly impossible that all zones are activated sequentially by noise. Thus, if the final state of the finite state machine is reached, it can be considered that a vehicle is detected. Then the information such as the

entering and leaving time of the vehicle to each detection zone, as well as the energy responses are collected for further use.

In our system, there are two finite state machines constructed for each lane, one like in Fig. 7, and the other has the reversed order from state 1 to state K. This double chain structure is used for retrograde vehicle detection, which is very important for public safety. If the chain with the reverse order other than current lane's travelling direction is activated, the retrograde alarm will be sounded by the system.

B. Lane Occupancy

Lane occupancy R_t is defined as the percentage of the vehicle passing time to a certain cross section vs. the total observation time, as depicted in equation (12), where t_i is the passing time of the ith vehicle, T is the total observation time.

$$R_t = \frac{\sum_{i=1}^{n} t_i}{T} \times 100 \tag{12}$$

In our system, when a vehicle is reported by the vehicle counting module, the passing time can be calculated as the average time difference of the vehicle enters and leaves individual detection zones. So, the lane occupancy can easily be updated according to (12).

C. Vehicle Speed

For passive vehicle monitoring sensors, such as inductive loop, magnetic sensor, as well as acoustic monitor, at least two detection zones should be established to estimate the vehicle speed. Since the inter-sensor distance can be accurately measured in the installation procedure for loops and magnetic sensors, vehicle speed in these monitoring approaches can be estimated as the ratio of inter-sensor distance and detection zone activation time difference [5, 7]. However, in our system, since multiple detection zones are constructed for each lane, it is possible to perform more accurate speed estimation with the redundant information provided by these detection zones. In this subsection, we propose a new speed estimation strategy based on the activation time of multiple detection zones. Here the zone activation time means the time average of vehicle enters and leaves a detection zone.

Fig. 8 is the illustration of the proposed speed estimation approach. After a vehicle is detected, the distance r from the monitor to the corresponding lane center can be calculated as (13), where h is monitor height, which is available in the monitor installation, and b is the distance from the road-side base structure to the corresponding lane center, which can be calculated from the road parameters.

$$r = \sqrt{h^2 + b^2} \tag{13}$$

Since the elevation angle of each detection zone is known, the distance d in Fig. 8 for each detection zone can be calculated by (14), and the activation time t_k of each detection zone satisfies the relation in (15), where v is the vehicle speed to be estimated.

$$d_k = r\tan(\theta_k) \tag{14}$$

$$t_k = \frac{d_k}{v} = \frac{r}{v}\tan(\theta_k) \tag{15}$$

Letting $u = r/v$ as the slope of the linear equation in (15), then, for all detection zones, an over-determined linear system can be established as (16) and (17), where $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_K]^{\mathrm{T}}$, $\boldsymbol{t} = [t_1, \cdots, t_K]^{\mathrm{T}}$, $\mathbf{1}$ is a K dimensional vector with all ones, and c is the constant term of the linear equation.

$$\boldsymbol{\theta}[u, c]^{\mathrm{T}} = \boldsymbol{t} \tag{16}$$

$$\boldsymbol{\Theta} = [\tan(\boldsymbol{\theta}), \mathbf{1}] \tag{17}$$

The least square approach [27] can be used to solve this over-determined linear system, as depicted in (18), where the subscript ls stands for the least square solution. Then, vehicle speed can be estimated by (19).

$$[u, c]_{ls}^{\mathrm{T}} = (\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{t} \tag{18}$$

$$v = r/u \tag{19}$$

Considering the relationship between linear speed and angular speed, from equation (19) we can also find that, the physical interpretation of $1/u$ is the vehicle angular speed (rad/s).

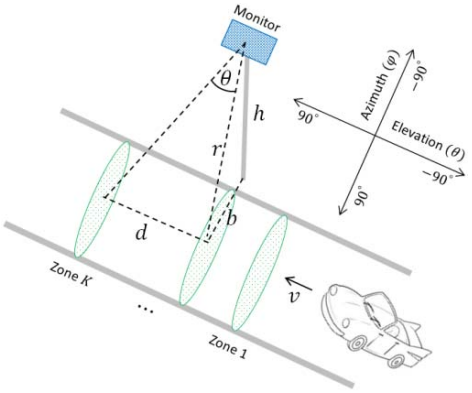


Fig. 8. The illustration of vehicle speed estimation.

D. Vehicle Type

Like inductive loop and magnetic sensor, a response curve will be generated when a vehicle passes through a detection zone. This curve may be considered as the signature of a certain vehicle type, which may be used for vehicle classification. Vehicle classification methods are investigated in many literatures, e.g., the support vector machine (SVM) based approach in [5], the neural network approach in [2], and the decision tree approach in [6]. However, for large/small vehicle classification, we use a simple approach which utilizes the definition of "long vehicle" in the China standard GB/T 26771-2011 in [28]. A vehicle is classified as "long vehicle" by microwave traffic monitor if its passing time exceeds 2.5 times of the average passing time. Thus, in our system, a simple classifier is built which records the passing time of individual vehicles and calculates the average value as the classification boundary, then, vehicle classification is performed according to the definition of the "long vehicle".

VI. Experiment

To test the proposed traffic monitoring system, the real-world experiment was carried out in a six-lane bidirectional highway environment, as shown in Fig. 9. The height of the

microphone array is 5.5 m, the lane width is 3.75 m, and the distance from the road-side base structure to the first lane is 3.25 m. Sound signals were collected via a National Instruments (NI) PXIe system with three PIXe 4499 sound and vibration data acquisition cards (48 channels in total). The sampling rate of 32 kHz was used, meanwhile, the developed prototype system was also deployed in the NI system for real-time traffic monitoring.

(a) the microphone array (b) video data example

Fig. 9. The real-world experimental environment.

Fig. 10 shows the GUI of the developed prototype system. This system is developed in Java, there are four main parts in the system GUI. The table on the top shows the traffic indices for each lane. The panel on the left visualizes the ATI of vehicles, the bottom of the ATI is the lane detection cross section, and bright spot means high sound energy, which indicates the appearance of a vehicle. The panel on the middle right visualizes the vehicle azimuth PDF and the detected lanes, the lanes are also painted on the ATI with the same colors. The panel on the lower right visualizes the responses of individual coarse detection zones.

Fig. 10. The system GUI.

In the experiment, four lanes were detected by the monitoring system. From the traveling directions of the observed vehicles, it can be inferred that, from left to right, the first three lanes are for the up-road direction, while the fourth lane (or lanes, because of the resolution problem as shown in Fig. 6 (a)) is for the down-road direction. We deduce that there may be several reasons why only four of six lanes were detected. First, the array resolution is not high enough. Second, the sound energy from the farthest two lanes is too small to be detected. Third, the sound field is jammed by the sound barrier at the opposite side of the road, as shown in Fig. 9. However, since the three lanes for the up-road direction are successfully detected, the proposed system is adequate for one-side traffic monitoring in the environment of Fig. 9.

To evaluate the system performance, 30 minutes of the audio data are used. The corresponding video data collected by a road-side camera were labeled by human, which are used as the ground truth. For video based lane occupancy and speed estimation, the "six-meter lines" on the road are used as the reference to estimate the vehicle passing time, as shown in Fig. 9 (b). There are 25 video frames in one second, so, the accuracy of the video data is 0.04 second.

The traffic monitoring results are listed in Table I, the system performance is evaluated by the error rate, which is calculated according to equation (20). Comparing the acoustic based results with the ground truth, we can find that the proposed system can achieve acceptable accuracy for vehicle counting and speed estimation.

$$error\ rate = \frac{|video\ result - audio\ result|}{video\ result} \times 100 \qquad (20)$$

The error rate of the lane occupancy is high since the ways for vehicle passing time estimation between video and audio are different. For video based results, vehicle passing time is estimated as the time which a vehicle passes the front of the "six-meter lines", so, the thickness of the corresponding cross section can be considered as zero. However, for audio based results, the passing time is estimated as the time difference of a vehicle enters and leaves a coarse detection zone, so, the thickness of the resulted cross section is nonzero, which is related to the beam width of the vertical beamformer. Although the values of the two lane occupancies are different, they are both derived according to the definition in (12), so, we believed that the audio based lane occupancy can also reflect the traffic quality.

TABLE I. TRAFFIC MONITORING EVALUATION RESULTS

		Video Result	Audio Result	Error Rate (%)
Vehicle Count	Lane 1	92	90	2.00
	Lane 2	435	368	15.00
	Lane 3	566	482	14.00
	Total	1093	940	13.00
Lane Occupancy (%)	Lane 1	1.65	1.44	12.70
	Lane 2	6.55	8.26	26.10
	Lane 3	6.29	9.94	58.11
	Total	14.49	19.64	35.56
Average Speed (km/h)	Lane 1	76.58	53.55	30.07
	Lane 2	92.76	74.77	19.40
	Lane 3	111.32	97.56	12.36
	Total	101.01	84.43	16.42
Vehicle Type (Large/Small)	Lane 1	37/55	3/87	91.00/58.00
	Lane 2	142/293	31/337	78.00/15.00
	Lane 3	25/541	30/452	20.00/16.00
	Total	204/889	64/876	68.00/1.00

For vehicle classification, only vehicle passing time is used as the classification feature in current implementation. The high classification error rates in Table I tell us that such simple implementation is still not enough for real application. From the video data, we can find that the lengths of some large vehicles, e.g., trucks, are not always longer than 2.5 times of small vehicles. Thus, only considering passing time as the single feature may not distinguish large vehicles from small ones. To improve the classification performance, more features may be added in, like average energy, the shape of the response curve, etc., and more advanced classifiers may be used, like decision tree and SVM [5, 6].

VII. Conclusion and Future Work

In this paper, an acoustic traffic monitoring system is introduced, which uses the emitting sound of the passing by vehicles to perform multi-lane traffic surveillance. Since acoustic features are robust against light and weather variations, it is possible to utilize acoustic features to perform robust traffic monitoring. First, the main modules of the proposed system are introduced to provide an overview of the system structure. Then, the construction of coarse detection zones and the calculation of different traffic indices are investigated. At last, the real-world experiment illustrates the availability of the proposed system.

In future work, we will perform more real-world experiments to verify and improve the accuracy of the derived traffic indices. E.g., GPS and radar gun can be used for speed comparison and calibration. Other traffic monitoring policies, like microwave and video based traffic monitors can also be incorporated in the experiments for comparison. In addition, we would like to investigate more advanced vehicle classification methods which are suitable for acoustic features.

Acknowledgment

The authors would like to thank our senior engineer Jian Wang for microphone array machining, and real-world vehicular sounds acquisition.

References

[1] H. Zhang, W. Yu, X. Sun, "Adaptive Traffic Lane Detection Based on Normalized Power Accumulation," IEEE Internation Conference on Intelligent Transportation Systems, pp. 968-973, 2008.

[2] S. Meta, M. G. Cinsdikici, "Vehicle-Classification Algorithm Based on Component Analysis for Single-Loop Inductive Detector," IEEE Transactions on Vehicular Technology, vol. 59, no. 6, pp. 2795-2805, 2010.

[3] S. Park, S. G. Ritchie, "Innovative Single-Loop Speed Estimation Model with Advanced Loop Data," IET Intelligent Transport Systems, vol. 4, no. 4, pp. 232-243, 2010.

[4] P. Wang, C. Li, C. Wu, H. Li, "A Channel Awareness Vehicle Detector," IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 2, pp. 339-347, 2010.

[5] S. Taghvaeeyan, R. Rajamani, "Portable Roadside Sensors for Vehicle Counting, Classification, and Speed Measurement," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 1, pp. 73-83, 2014.

[6] B. Yang, Y. Lei, "Vehicle Detection and Classification for Low-Speed Congested Traffic With Anisotropic Magnetroresistive Sensor," IEEE Sensors Journal, vol. 15, no. 2, pp. 1132-1138, 2015.

[7] H. Li, H. Dong, L. Jia, D. Xu, Y, Qin, "Some Practical Vehicle Speed Estimation Methods by a Single Traffic Magnetic Sensor," IEEE International Conference on Intelligent Transportation Systems, pp. 1566-1573, 2011.

[8] S. Chen, Z. Sun, B. Bridge, "Traffic Monitoring Using Digital Sound Field Mapping," IEEE Transactions on Vehicular Technology, vol. 50, no. 6, pp. 1582-1589, 2001.

[9] J. Benesty, J. Chen, Y. Huang, "Microphone Array Signal Processing," vol. 1, Springer Science & Business Media, 2008.

[10] V. Tyagi, S. Kalyanaraman, R. Krishnapuram, "Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics," IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 3, pp. 1156-1166, 2012.

[11] X. Wu, Z. Zhang, R. Peng, Y. Fu, W. He, K. Xie, G. Song, "A Means and Corresponding Methods for Acoustic Highway Traffic Monitoring," China Patent CN 102682765 B, 2013 (in Chinese).

[12] L. 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.

[13] J. P. Kuhn, B. C. Bui, G. J. Pieper, "Acoustic Sensor System for Vehicle Detection and Multi-Lane Highway Monitoring," U.S. Patent No. 5,798,983. 25 Aug. 1998.

[14] V. Cevher, R. Chellappa, J. H. McClellan, "Vehicle Speed Estimation Using Acoustic Wave Patterns," IEEE Transactions on Signal Processing, vol. 57, no. 1, pp. 30-47, 2009.

[15] B. D. V. Veen, K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," IEEE ASSP Magazine, vol. 5, no. 2, pp. 4-24, 1988.

[16] J. G. Proakis, D. G. Manolakis, "Digital Signal Processing Principles, Algorithms, and Applications," Fourth Edition, Publishing House of Electronics Industry, Beijing, 2010.

[17] R. H. MacPhie, L. Yuan, "A Modified Mills Cross With Elements Spaced One Wavelength Apart," General Assembly and Scientific Symposium, pp. 1-4, 2011.

[18] H. M. Aumann, "A Pattern Synthesis Technique for Multiplicative Arrays," PIERS Proceedings, Cambridge, USA, pp. 864-867, 2010.

[19] R. H. MacPhie, "A Mills Cross Multiplicative Array with the Power Pattern of a Conventional Planar Array," Antennas and Propagation Society International Symposium, pp. 5961-5964, 2007.

[20] M. Wax, T. Shan, T. Kailath, "Spatio-Temporal Spectral Analysis by Eigenstructure Methods," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 32, no. 4, pp. 817-827, 1984.

[21] P. Lynch, "The Dolph-Chebyshev Window: A Simple Optimal Filter," Monthly Weather Review, vol. 125, no. 4, pp. 655-660, 1997.

[22] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, 1986.

[23] J. Capon, "High-Resolution Frequency-Wavenumber Spectrum Analysis," Proceedings of the IEEE, vol. 57, no. 8, pp. 1408-1418, 1969.

[24] T. L. Marzetta, S. H. Simon, H. Ren, "Capon-MVDR Spectral Estimation From Singular Data Covariance Matrix, with no Diagonal Loading," Proc. Fourteenth Annual Workshop on Adaptive Sensor Array Processing, MIT Lincoln Laboratory, 2006.

[25] B. Rafaely, "Spatial-Temporal Correlation of a Diffuse Sound Field," The Journal of the Acoustical Society of America, vol. 107, no. 6, pp. 3254-3258, 2000.

[26] V. Tourbabin, M. Agmon, B. Rafaely, J. Tabrikian, "Optimal Real-Weighted Beamforming with Application to Linear and Spherical Arrays," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 9, pp. 2575-2585, 2012.

[27] S. J. Leon, "Linear Algebra with Applications," Seventh Edition, China Machine Press, Beijing, 2007.

[28] Y. Zhu, W. Hou, "The Setting Specification for Microwave Traffic Detector," China Standard GB/T 26771-2011, 2011 (in Chinese).