Cross Array and Rank-1 MUSIC Algorithm for Acoustic Highway Lane Detection

Yueyue Na, Yanmeng Guo, Qiang Fu, *Member, IEEE*, and Yonghong Yan

***Abstract*—A vehicle emits sound as it travels along the road, which can be used as a kind of robust feature for traffic monitoring. In this paper, an acoustic-based lane detection approach is introduced for a multilane traffic monitoring system. First, a microphone array is designed according to a typical Chinese highway configuration. The design is based on the cross-array structure, and the cross-correlation matrix from the two subarrays in the selected working frequency band is calculated for the subsequent traffic monitoring operations. Then, a cross section across the road is constructed by beamforming, in which the single-source assumption can be applied, and the passing vehicle azimuth is detected by the proposed rank-1 Multiple Signal Classification (MUSIC) algorithm. Finally, a Parzen-window-based technique is proposed to estimate the vehicle azimuth probability density function (pdf) from the individual azimuth observations. Lane centers and boundaries can be revealed from the peak and valley patterns of the estimated pdf. A prototype traffic monitoring system is developed, and several lane detection approaches are compared in both simulated and real-world environments in the developed system framework. The experimental results exhibit the efficiency of the proposed approach.**

***Index Terms*—Traffic monitoring, microphone array, beamforming, direction-of-arrival estimation, lane detection.**

I. Introduction

In modern intelligent transportation system (ITS), a large amount of traffic monitors are distributed in the road network, so that real-time traffic statistics can be collected by the control center, which will be further used for flow control and dynamic planning. As the "eyes" and the "ears" of the control center, these monitoring devices are the essential building blocks of the ITS. Developing more reliable and accurate traffic monitoring techniques will facilitate road condition, improve public safety, and reduce transportation cost [1].

Various kinds of traffic monitoring techniques are developed for different road conditions and environments. Among them, the induction loop has been used for more than fifty years [1], however, this in-pavement device suffers from high installation/maintenance cost [2]. Modern non-invasive traffic monitoring techniques can be roughly classified into two categories: active and passive [2]. Active approaches mainly based on laser, infrared, radar [1], ultrasound, and other techniques. This type of devices detects vehicles by transmitting a signal and detecting its reflection. On the other hand, passive approaches, such as video, passive infrared, magnetic sensors [3], as well as acoustic based monitors [2], [4], perform vehicle surveillance by the reflected (visible light), or the emitted signal (infrared, sound, etc.) from vehicles.

This paper describes an acoustic based lane detection approach, which is the first step for constructing a multi-lane acoustic traffic monitoring system. Compared with other approaches, acoustic based systems have many advantages: First, acoustic sensor (microphone) is much less expensive than other types of sensors [4], so, the total hardware cost can be reduced. Second, acoustic features are very robust against whether, light, and environmental variations [2], which guarantees the reliability of the traffic data. However, there also have some difficulties to build an acoustic traffic monitoring system: Since the sound wavelength used for vehicle detection has the magnitude of centimeters, the resolution of acoustic based devices is limited under the restriction of acceptable aperture size [5]. Moreover, in addition to the vehicular sound,[1] there are still many other sounds and noise in an open-air environment, i.e., multi-source and noisy signals should be processed. How to robustly detect vehicles in a complex environment is a challenging task for acoustic traffic monitoring.

There have many researches been conducted for acoustic vehicle detection and traffic monitoring. For example, in [6], a uniform linear array (ULA) with four microphones is used for vehicle approaching detection in T-intersection roads to prevent traffic accidents. A cross correlation based method is used for sound source localization in this paper. In [7], two microphones with optimized inter-sensor distance are used to collect vehicular sounds, then, the generalized cross correlation functions and particle filters are used to estimate vehicle speed

Manuscript received November 15, 2015; accepted January 12, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 11461141004, Grant 61271426, Grant 11504406, Grant 11590770, Grant 11590771, Grant 11590772, Grant 11590773, and Grant 11590774; by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06030100, Grant XDA06030500, and Grant XDA06040603; by the National 863 Program under Grant 2015AA016306; by the National 973 Program under Grant 2013CB329302; and by the Key Science and Technology Project of the Xinjiang Uygur Autonomous Region under Grant 201230118-3. The Associate Editor for this paper was F.-Y. Wang.

Y. Na, Y. Guo, and Q. Fu are with the Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China (e-mail: nayueyue@hccl.ioa.ac.cn; guoyanmeng@hccl.ioa.ac.cn; qfu@hccl.ioa.ac.cn).

Y. Yan is with the Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China, and also with Xinjiang Laboratory of Minority Speech and Language Information Processing, Ürümqi 830011, China (e-mail: yanyonghong@hccl.ioa.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2521661

[1] In this paper, since the vehicle emitting noise signal is of interest, here we use "vehicular sound" to distinguish it from the common use of "noise" for undesired interference.

and wheelbase length. In [8], a microphone array is designed with four subarrays located in its up-down road and cross road directions. Signals within subarrays are first summed together as zero-degree delay-and-sum (DS) beamforming [9] for enhancement purpose. Then, cross correlations between the enhanced up-down road and cross road subarray pairs are calculated for vehicle detection. Since two perpendicular scanning directions are adopted, vehicles in different lanes can be distinguished in this system. Cross correlation based methods detect vehicles via the time difference of arrival (TDOA) of sounds among different microphones. Although high performance can be achieved by this class of methods in single sound source cases, the cross correlation spectrum will get more confused as the number of sources increases [9]. Therefore, the three preceding approaches are suitable in light and moderate traffic flow environments.

In addition to the TDOA based approaches, a single microphone approach is proposed in [4] for traffic density state estimation, which classifies the road condition into free, medium, and jammed traffic flows according to the collected vehicular sounds. Although this approach is suitable for crowded city roads, it cannot perform line-wise traffic monitoring, accurate vehicle counting, and speed estimation. The work in [10] is another single microphone approach for vehicle monitoring. In this work, different mathematical and physical models are established to estimate different vehicular parameters, such as speed, wheelbase length, and tire track length. However, the complexity of these models prevents its application in multi-vehicle conditions. A commercial acoustic multi-lane highway monitoring system can be found in [2]. This system is based on the beamforming [11] technique, up to five lanes can be simultaneously managed, traffic quality measuring indices such as vehicle count, average speed, and lane occupancy, can also be derived. However, this system uses a rectangular array, which contains many array elements, so the total hardware cost is still high.

Lane detection is very important for multi-lane traffic monitoring, since the accuracy and robustness of the detected lanes will directly affect the following traffic indices calculation. In this paper, an acoustic highway lane detection approach is proposed. This approach is based on the famous MUSIC (MUltiple SIgnal Classification) algorithm [12]–[15], which first detects the direction of arrival (DOA) of passing vehicles, then, estimates lane positions by the cumulated DOA statistics. Our contributions in this paper are: (1) a microphone array is designed for traffic monitoring; (2) the rank-1 MUSIC algorithm is derived for real-time vehicle detection, which has higher angular resolution and lower computational complexity; (3) a lane detection algorithm which detects lanes from vehicle DOA statistics is presented. The rest of this paper is organized as follows: In Section II we briefly introduce the basic concept of acoustic traffic monitoring and the idea of microphone array design. In Section III the rank-1 MUSIC algorithm for vehicle detection is depicted, then, the method for lane center and boundary detection is given in Section IV. Experiments and comparisons are conducted in Section V to show the effectiveness and performance improvement of the proposed approach. At last, we conclude this paper in Section VI.

The frequently used notations in this paper are listed below for easy reference.

1. Italic lowercase letters denote scalars, boldface italic lowercase letters denote column vectors, and boldface italic uppercase letters denote matrices, e.g., a, $\boldsymbol{a}$, and $\boldsymbol{A}$.
2. Superscripts *, ^{T}, and ^{H} denote complex conjugate, matrix and vector transpose, and conjugate transpose, respectively, $\boldsymbol{A}^H = (\boldsymbol{A}^T)^* = (\boldsymbol{A}^*)^T$.
3. Commas separate values within rows, e.g., $\boldsymbol{a} = [a_1, a_2]^T$.
4. The source signal, the array collected signal, and the beamforming output are denoted by the letters s, x, and y, respectively.
5. Indices m and n denote array element index and source index, respectively. There are M array elements and N sources in the model.
6. Subscripts h and v denote the horizontal and the vertical subarray in the cross array, which have M_h and M_v elements, respectively.
7. Subspaces are denoted as boldface italic uppercase handwritten letters, e.g., $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{E}}$ for signal and noise subspaces.

II. Microphone Array Design

A. Fundamentals of Acoustic Highway Traffic Monitoring

As depicted in Fig. 1, the acoustic monitor is installed on the existing roadside structure with its normal direction pointing to the road center. Vehicle emits sound (consists of engine noise, tire noise, exhaust noise, air turbulence noise, etc. [4], [10]) when it travels along the road. The sound is captured by the acoustic monitor, and then analyzed for vehicle detection and traffic monitoring. It is unnecessary to calibrate the monitor's normal direction carefully in the installation, as the relative lane positions can be automatically detected by the lane detection algorithm.

The basic idea of acoustic traffic monitoring is utilizing beamforming techniques [5], [11], [16] to form multiple "detection zones" in different look directions. Because of the spatial filtering property [11] of the beamforming, incoming vehicular sounds other than the array look direction will be attenuated by the algorithm. Then, the resulted signal energy can be accumulated to determine the presents or absences of a vehicle in a detection zone, which can be further used for traffic monitoring. Different detection zones with different shapes, widths, and positions are formed for different purposes. Fig. 2 is the top view of Fig. 1, the monitor operates from a "sidefire" position. There are two kinds of detection zones constructed: fine detection zones (red ellipses) and coarse detection zones (green ellipses). Fine detection zones are used to detect lanes, they have smaller widths and fixed positions. Multiple fine detection zones are concatenated to form a cross section across the road. Once a vehicle passes through this cross section, its azimuth will be detected by the algorithm in Section III, and lane positions will be estimated by the algorithm in Section IV. On the other hand, coarse detection zones are used for lane-wise traffic monitoring. The width of a coarse detection zone covers an entire lane, and each lane is managed by multiple

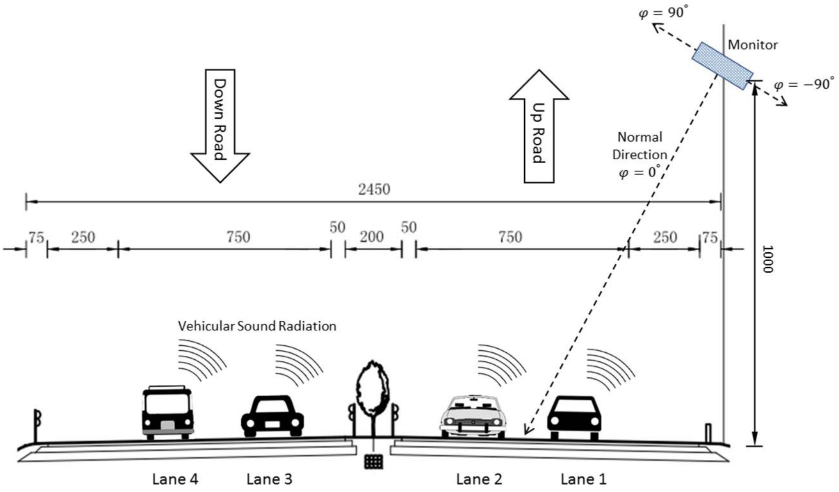


Fig. 1. Acoustic highway traffic monitoring. The lengths are in centimeters, and the angles are in degrees. The lane width is 375 cm. We suppose that the monitor is used for one-side traffic monitoring, so, the monitor's normal direction is pointing to the road center of the up-road direction.

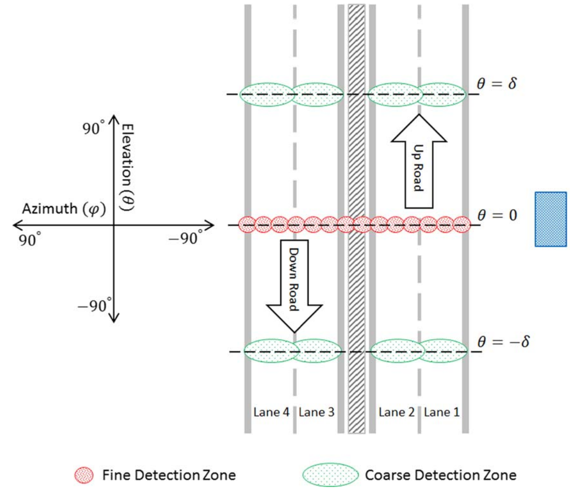


Fig. 2. Coordinate system and detection zones. The angles are in degrees.

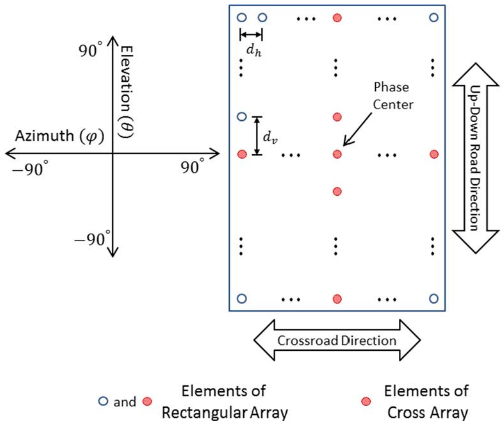


Fig. 3. Rectangular array and corresponding cross array. In this figure, the array is facing to the reader, so, the azimuth is reversed compared with Fig. 2.

(at least two) coarse detection zones. Basic measuring indices which reflect the road condition and traffic quality, such as vehicle count and lane occupancy can easily be derived by counting the number and the detection time of passing vehicles. In addition, vehicle speed and vehicle size (small vs. large) can be estimated from the vehicle detection time differences among coarse detection zones in the same lane [3]. Since the observed lane positions may change with weather, temperature, and traffic conditions, the positions and widths of coarse detection zones are adaptively updated in the monitoring procedure.

A two dimensional coordinate system is required for multi-lane traffic monitoring. Here we suppose the monitor is mounted far enough from the sound sources, so that the far-field model [16] can be applied. The DOA of a sound source in far-field model is described by two parameters: azimuth (φ) and elevation (θ) in spherical coordinate system. The coordinate system used for vehicle detection is also depicted in Figs. 1 and 2, where the monitor is facing down to the road. Since directions behind the monitor are not required, the azimuth here ranges from -90° to 90°, with 0° at the normal direction. Please note that the four-lane highway architecture used in this paper is only for demonstration purpose, the proposed algorithm can easily be configured to manage more than four lanes.

B. *Rectangular Array vs. Cross Array*

A planar array is required for two dimensional source localization. Instead of the rectangular array topology used in the system of [2], the cross array structure is used in this paper, as shown in Fig. 3. Both the rectangular and the cross array use the uniform element spacing scheme, however, the horizontal and the vertical spacing d_h and d_v are different.

The signal model in (1) [13] is used to compare the two array topologies, where $\boldsymbol{s} = [s_1, \ldots, x_N]^T$ and $\boldsymbol{x} = [x_1, \ldots, x_M]^T$ are the source and the collected narrow band signals with the central frequency equals to f, $\boldsymbol{e}$ is the noise term, $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$ is the mixing matrix, and $\boldsymbol{a}_n = [a_{1n}, \ldots, a_{\mathrm{Mn}}]^T$ is called the steering vector of source n. In far-field model, a_{mn} can be given by equation (2), where $j = \sqrt{-1}$, c is the wave propagation speed (340 m/s for sound in the air), $\boldsymbol{d}_n$ is an unit vector indicating the look direction of source n, and $\boldsymbol{r}_m$ is the mth array element position relative to the phase center. Both

$\boldsymbol{d}_n$ and $\boldsymbol{r}_m$ are in the three dimensional Cartesian coordinate system [5], [11], [16].

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} \tag{1}$$

$$a_{\mathrm{mn}} = \exp\left(j2\pi f \frac{\boldsymbol{d}_n^T \boldsymbol{r}_m}{c}\right) \tag{2}$$

For the rectangular array, the beamforming output y_r can be calculated according to (3), where $\boldsymbol{w}$ is called the beamformer, which can be designed by different beamforming algorithms [11], [17] for different signal enhancement requirements.

$$y_r = \boldsymbol{w}^H \boldsymbol{x} \tag{3}$$

In acoustic traffic monitoring, signal energy of the array output should be accumulated for vehicle detection. Substituting the signal model (1) into (3), and supposing different sources, sources and noise are uncorrelated, the output energy can be depicted in (4)–(6), where $E\{\cdot\}$ for expectation, $\boldsymbol{C}_{\mathrm{rx}}$ and $\boldsymbol{C}_{\mathrm{re}}$ are the signal and noise correlation matrix of the rectangular array, $\sigma_n^2 = E\{s_n s_n^*\}$ is the power of source n.

$$\boldsymbol{C}_{\mathrm{rx}} = E\{\boldsymbol{x}\boldsymbol{x}^H\} \tag{4}$$

$$\boldsymbol{C}_{\mathrm{re}} = E\{\boldsymbol{e}\boldsymbol{e}^H\} \tag{5}$$

$$\begin{aligned} E\{y_r y_r^*\} &= \boldsymbol{w}^H \boldsymbol{C}_{\mathrm{rx}} \boldsymbol{w} \\ &= \left|\boldsymbol{w}^H \boldsymbol{a}_1\right|^2 \sigma_1^2 + \cdots + \left|\boldsymbol{w}^H \boldsymbol{a}_N\right|^2 \sigma_N^2 + \boldsymbol{w}^H \boldsymbol{C}_{\mathrm{re}} \boldsymbol{w} \end{aligned} \tag{6}$$

Supposing the DOAs of different sources are different, when the beamformer $\boldsymbol{w}$ is steered to the look direction of source n, only the $\left|\boldsymbol{w}^H \boldsymbol{a}_n\right|^2 \sigma_n^2$ term in (6) has significant value, so, source n is enhanced. In an open-air environment, the noise vector $\boldsymbol{e}$ in (1) consists of sensor noise, environmental noise, and many other vehicular sounds far away in different directions, i.e., the diffuse noise model [18], [19] should be considered. The result is that the noise correlation matrix $\boldsymbol{C}_{\mathrm{re}}$ usually has full rank, which increases the difficulty of the vehicle detection problem [13].

On the other hand, as shown in Fig. 3, cross array [20]–[22] is made up of two perpendicular ULAs, which are referred to the horizontal and the vertical subarray in this paper. The two subarrays can share the same phase center if they both have odd number of elements. Two different beamformers can be used to manipulate the two subarrays, as shown in equation (7) and (8).

$$y_h = \boldsymbol{w}_h^H \boldsymbol{x}_h \tag{7}$$

$$y_v = \boldsymbol{w}_v^H \boldsymbol{x}_v \tag{8}$$

Cross array is also called multiplicative array [22], since the energy accumulation of the array is just the cross correlation of its two subarray's outputs. Here we omit the tedious mathematical proof, and give the conclusion that, in far-field model, the output of the cross array can be depicted in the form of (9)–(11) [23], where $\boldsymbol{C}_{\mathrm{cx}}$ and $\boldsymbol{C}_{\mathrm{ce}}$ are the signal and the noise correlation matrix of the cross array.

$$\boldsymbol{C}_{\mathrm{cx}} = E\{\boldsymbol{x}_h \boldsymbol{x}_v^H\} \tag{9}$$

$$\boldsymbol{C}_{\mathrm{ce}} = E\{\boldsymbol{e}_h \boldsymbol{e}_v^H\} \tag{10}$$

$$\begin{aligned} E\{y_h y_v^*\} &= \boldsymbol{w}_h^H \boldsymbol{C}_{\mathrm{cx}} \boldsymbol{w}_v \\ &= \boldsymbol{w}^H \boldsymbol{a}_1 \sigma_1^2 + \cdots + \boldsymbol{w}^H \boldsymbol{a}_N \sigma_N^2 + \boldsymbol{w}_h^H \boldsymbol{C}_{\mathrm{ce}} \boldsymbol{w}_v \end{aligned} \tag{11}$$

Comparing equation (11) with equation (6), we can find that they have very similar form: Both arrays utilize the inner product between the beamformer and the steering vector of the original rectangular array for signal enhancement. This means that for signal power estimation tasks, the cross array can approximate the performance of the corresponding rectangular array in theory, however, with much fewer array elements [22]!

To reduce the hardware cost, meanwhile maintaining the array performance, the cross array scheme is adopted in our system. After the collected time domain acoustic signals are transformed to frequency domain by short-time Fourier transform (STFT) [24], the correlation matrix can be approximated according to (12), where $\Delta\tau$ is the selected STFT frame index interval, $[f_1, f_2]$ is the working frequency band interval, which should be chosen $f_{\min} < f_1 < f_2 < f_{\max}$. The lower and upper bound $f_{\min}$ and $f_{\max}$ will be determined in the following two subsections. When $f_2 - f_1$ is not too large, the data can still be considered as narrow band signals with the central frequency $f = (f_1 + f_2)/2$.

$$\boldsymbol{C}_{\mathrm{cx}} \approx \sum_{f \in [f_1, f_2]} \sum_{\tau \in \Delta\tau} \boldsymbol{x}_h(f, \tau) \boldsymbol{x}_v^H(f, \tau) \tag{12}$$

Finally, the energy q of the cross array can be calculated according to (13). In addition to the benefit of cost reduction, there still have other advantages of this approach. First, the beamformer designing problem is decoupled into two independent and smaller problems in (13), which gives the system additional flexibility. Second, once the elevation of a cross section is chosen, the vertical beamformer $\boldsymbol{w}_v$ is fixed. Then, different horizontal beamformers $\boldsymbol{w}_h(\varphi)$ for different φ can be designed directly from the vertically preprocessed data $\boldsymbol{z}$ in (14), which reduces the complexity of cross section construction. To overcome the enormous variations of the environmental noise, the DS beamformer with Dolph-Chebyshev taper [25] is used as the vertical beamformer in our system. Third, multiple coarse detection zones within $\varphi \in [-90°, 90°]$ and $\theta \in [-\delta, \delta]$ can be established upon the single correlation matrix $\boldsymbol{C}_{\mathrm{cx}}$. One coarse detection zone can be considered as one "virtual sensor" on the road. Constructing multiple (possibly more than two) coarse detection zones for each lane may improve the monitoring accuracy, while no additional physical sensors are required.

$$q(\varphi) = \left|\boldsymbol{w}_h^H(\varphi) \boldsymbol{C}_{\mathrm{cx}} \boldsymbol{w}_v(\theta)\right|^2 = \left|\boldsymbol{w}_h^H(\varphi) \boldsymbol{z}\right|^2 \tag{13}$$

$$\boldsymbol{z} = \boldsymbol{C}_{\mathrm{cx}} \boldsymbol{w}_v(\theta) \tag{14}$$

Although the rectangular array is more robust than the corresponding cross array since more array elements are used, it has higher cost and lower flexibility. In the rectangular

TABLE I
LANE LOOK DIRECTIONS (DEGREE)

Lane Number	Lower Boundary	Upper Boundary	Lane Center	Lane Width
1	18.0	35.0	27.1	17.0
2	35.0	47.1	41.6	12.1
3	54.0	60.3	57.4	6.3
4	60.3	64.8	62.7	4.5

array scheme of [2], time domain signals are summed together according to the array's columns before STFT to reduce the complexity of the horizontal beamformer design. Column-wise summation is equivalent to the DS beamforming at $\theta = 0°$, which also means that the system can only generate detection zones at $\theta = 0°$. To construct two different detection zones per lane for vehicle speed estimation, the system in [2] accumulates two correlation matrices in different frequency bands to form two nested detection zones. Comparing the approach in [2] with the cross array approach in (13), vehicle traveling directions (up-road vs. down-road) cannot be distinguished by two nested detection zones, and an additional correlation matrix is required to form a new detection zone. Moreover, it cannot construct detection zones different from $\theta = 0°$.

C. Array Resolution and Working Frequency Lower Bound

A typical Chinese four-lane highway structure is given in Fig. 1. Supposing the height of the monitor is 1000 cm, then, the look directions of lane centers and boundaries can easily be derived, which are listed in Table I. To perform multi-lane traffic monitoring, vehicles in adjacent lanes should be distinguished by the system, so, the angular resolution requirement can approximately be estimated as the look direction differences between adjacent lane centers, which are 14.5°, 15.8°, and 5.3° between lane 1–2, lane 2–3, and lane 3–4, respectively. Since the system is used in highway environment, it can be assumed that the distance between adjacent vehicles in the same lane is large enough when the traffic is not congested. Therefore, the array resolution in the up-down road direction is automatically fulfilled once the resolution requirement in the crossroad direction is reached.

When the wave propagation speed is fixed, the beam width[2] (main lobe width) of beamforming algorithms is related to the array aperture size and the signal frequency [5], [16]. Larger aperture size and higher frequency result in sharper beam. For practical usage, the monitor size should keep small enough for easy installation, so, it is fixed at about 20 cm (crossroad direction) × 30 cm (up-down road direction) in the array designing procedure. Although the aperture size is limited, the frequency range of vehicular sound extends from near 0 Hz to 16 kHz, with significant energy at the lower frequencies [2], which enable us to select a proper frequency band to meet the angular resolution requirement.

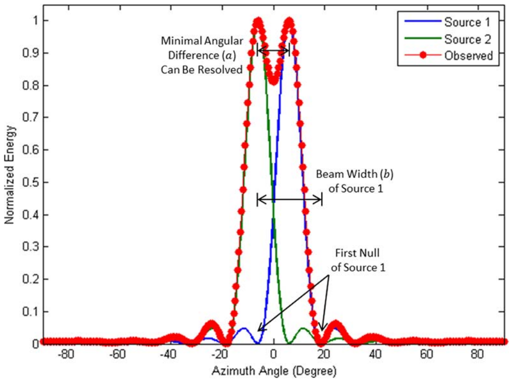


Fig. 4. Resolution of the DS beamforming.

TABLE II
WORKING FREQUENCY LOWER BOUNDS ($l = 20$ cm)

Lane Pairs	Lane Center Difference (a) (Degree)	Frequency Lower Bound (f_{min}) (Hz)
1-2	14.5	6790
2-3	15.8	6244
3-4	5.3	18404

The DS beamforming is a classical algorithm, in which signals from different array elements are time delayed and summed together to enhance the signal from a certain direction [9]. Although the DS algorithm doesn't have the sharpest beam, its beam with b can analytically be expressed in equation (15) [9], which can be used to theoretically estimate the working frequency lower bound in our problem. Since the cross-road direction needs to be considered, the horizontal subarray aperture size l is needed in (15).

$$b = 2\arcsin\left(\frac{c}{lf}\right) \tag{15}$$

The resolution of a beamformer describes its ability to distinguish signals coming from two directions close to each other [16]. The Rayleigh criterion [26] states that two directions can be just exactly resolved from the observed energy spectrum when the peak of source 1 falls on the first null of source 2, as depicted in Fig. 4. According to the preceding formulation, for the highway structure in Fig. 1, the working frequency lower bound $f_{\min}$ to distinguish vehicles in different lanes can be solved from (15), as given in (16), where $a = b/2$ is the angular resolution, i.e., look direction difference of two adjacent lane centers.

$$f_{\min} = \frac{c}{l\sin(a)} \tag{16}$$

According to (16), the $f_{\min}s$ to distinguish each pair of adjacent lanes are given in Table II. These parameters will be used in the next subsection to guide the microphone array design.

D. Array Element Spacing and Working Frequency Upper Bound

Another issue should be addressed in sensor array design is the spatial aliasing problem. The spatial sampling theorem in equation (17) [5] tells us that the horizontal subarray element

[2] In some literatures, beam width refers to the angle between the half power (−3 dB) points of the main lobe. However, in this paper, we follow the definition in [9], which refers beam width to the region between the first zero-crosses on either side of the main lobe (Fig. 4).

TABLE III
CROSS ARRAY DESIGNS ($l = 20$ cm FOR RESOLUTION CALCULATION)

No.	f_{max} (kHz)	λ_{min} (cm)	Resolution (a) (Degree)	Aperture Size (width x height) (cm)	d_h (cm)	d_v (cm)	M_h	M_v	M
1	8	4.25	12.3	20 x 32	2.0	4.0	11	9	19
2	10	3.40	9.8	21 x 30	1.5	3.0	15	11	25
3	12	2.83	8.1	21 x 30.8	1.4	2.8	16	12	28
4	14	2.43	7.0	20.4 x 31.2	1.2	2.4	18	14	32
5	16	2.13	6.1	20 x 32	1.0	2.0	21	17	37
6	18	1.89	5.4	20.7 x 30.6	0.9	1.8	24	18	42
7	20	1.70	4.9	20 x 30.4	0.8	1.6	26	20	46

spacing d_h should smaller than the half of the minimum wavelength $\lambda_{\min}$ in the signal of interest to perform the full scan in $\varphi \in [-90^\circ, 90^\circ]$. Equation (17) also gives us the basic relationship between d_h and the working frequency upper bound $f_{\max}$.

$$d_h < \frac{\lambda_{\min}}{2} = \frac{c}{2 f_{\max}} \tag{17}$$

However, in the highway monitoring problem in Fig. 2, vertical beamformers only need to be steered within a small elevation range $\theta \in [-\delta, \delta]$ ($\delta = 3^\circ$ in our experiment) to form vertically separated detection zones. According to the basic array signal processing theory, the first pair of grating lobes (spatial aliasing) appears at $\theta_g = \pm 90^\circ$ when the look elevation $\theta = 0^\circ$ and $d_v = \lambda_{\min}$. However, signal energy is negligible at $\theta_g \approx \pm 90^\circ$ since vehicles are far away from the monitor. Therefore, the spatial sampling theorem in (17) can be relaxed to (18) for the vertical subarray, which further reduces the number of required array elements.

$$d_v < \lambda_{\min} = \frac{c}{f_{\max}} \tag{18}$$

Under the constraints in Table II, equation (17) and (18), several cross array schemes are derived in Table III, where M_h, M_v, and M are the element number of horizontal subarray, vertical subarray, and the total array, respectively.

In Table III, the array resolution increases with $f_{\max}$, however, the number of required array elements also increases. Compared with Table II, if the DS beamforming is used, all schemes in Table III can distinguish lane 1, 2, and 3, which is enough for one-side monitoring, however, only the last one can distinguish all four lanes in Fig. 1.

Finally, the scheme no. 5 in Table III is chosen for our prototype system, as it has moderate cost and sufficient resolution margin, which enable us to further select the proper working frequency band and optimize the array topology in experiments. In the next section, we will introduce the rank-1 MUSIC algorithm, which can also distinguish all four lanes in the simulation because of its super directivity property.

III. THE RANK-1 MUSIC ALGORITHM

A. MUSIC Algorithm Revisited

Supposing there are fewer sources than sensors $(N < M)$, the signal model in (1) implies that, the source component $\boldsymbol{As}$ resides in the N dimensional subspace $\boldsymbol{\mathcal{S}}$ of the M dimensional observation space. Subspace $\boldsymbol{\mathcal{S}}$ is spanned by all source steering vectors, as depicted in (19) [13].

$$\boldsymbol{\mathcal{S}} = \text{span}\{\boldsymbol{A}\} = \text{span}\left\{[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]\right\} \tag{19}$$

Let us denote the noise subspace $\boldsymbol{\mathcal{E}}$ as the orthogonal complement (denoted by the superscript $\perp$) of $\boldsymbol{\mathcal{S}}$, which is spanned by some basis vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M-N}$:

$$\boldsymbol{\mathcal{E}} = \boldsymbol{\mathcal{S}}^\perp = \text{span}\{\boldsymbol{U}\} = \text{span}\left\{[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M-N}]\right\} \tag{20}$$

According to the orthogonality, the null directivity patterns can be formed when the steering vector $\boldsymbol{a}(\varphi)$ is steered to the source DOAs, as shown in (21), where φ_n represents the look direction of $\boldsymbol{a}_n$.

$$\boldsymbol{U}^H \boldsymbol{a}(\varphi) = \boldsymbol{0}, \quad \varphi \in \{\varphi_1, \ldots, \varphi_N\} \tag{21}$$

Equivalently, the MUSIC spectrum is defined in (22), which has peaks at source DOAs [13]. Although only the azimuth is used in the derivations in (21) and (22), the algorithm can easily be extended to the two dimensional case which contains both azimuth and elevation.

$$q(\varphi) = \frac{1}{\left|\boldsymbol{U}^H \boldsymbol{a}(\varphi)\right|^2} \tag{22}$$

It is essential for MUSIC to accurately identify $\boldsymbol{\mathcal{S}}^\perp = \text{span}\{\boldsymbol{U}\}$. In the signal model of (1), when the noise term $\boldsymbol{e}$ is not salient, $\boldsymbol{U}$ is usually obtained via the eigenvalue decomposition (EVD) of $\boldsymbol{C}_x = E\{\boldsymbol{x}\boldsymbol{x}^H\}$, by assigning $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M-N}$ to the eigenvectors corresponding to $\boldsymbol{C}_x$'s $M - N$ smallest eigenvalues [13].

There still have some practical issues to be addressed for the usage of the classical MUSIC algorithm. First, the number of sources N, which is usually unknown in real-world applications, should be estimated correctly in order to select the proper number of basis vectors to span the noise subspace. Second, the EVD operation has the time complexity of $O(M^3)$, which is not suitable for real-time DOA estimation.

B. The Rank-1 MUSIC Algorithm for Vehicle Detection

The effect of the vertical beamformer in equation (14) is forming a cross section across the road, so that the sound outside this cross section will be attenuated. This effect can be revealed by equation (23), which is derived by substituting the signal model in (1) into the horizontal and vertical subarray data

$\boldsymbol{x}_h$ and $\boldsymbol{x}_v$, then, substituting the results into (9) and (14), where $\boldsymbol{a}_{\mathrm{hn}}$ and $\boldsymbol{a}_{\mathrm{vn}}$ are the horizontal and the vertical steering vectors of source n.

$$\boldsymbol{z} = \sigma_1^2 \boldsymbol{a}_{h1} \boldsymbol{a}_{v1}^H \boldsymbol{w}_v + \cdots + \sigma_N^2 \boldsymbol{a}_{\mathrm{hN}} \boldsymbol{a}_{\mathrm{vN}}^H \boldsymbol{w}_v + \boldsymbol{C}_{\mathrm{ce}} \boldsymbol{w}_v \tag{23}$$

For highway environment, there is a great chance that a cross section is occupied by only a single vehicle at a time when the traffic is not congested. Without loss of generality, supposing source 1 is in the cross section, which means that only the first term on the right hand side of (23) has significant value. So, the cross section data $\boldsymbol{z}$ can be reformulated to (24), where $\boldsymbol{r}$ is a residual term.

$$\boldsymbol{z} = \sigma_1^2 \boldsymbol{a}_{h1} \boldsymbol{a}_{v1}^H \boldsymbol{w}_v + \boldsymbol{r} \tag{24}$$

$$\boldsymbol{r} = \sigma_2^2 \boldsymbol{a}_{h2} \boldsymbol{a}_{v2}^H \boldsymbol{w}_v + \cdots + \sigma_N^2 \boldsymbol{a}_{\mathrm{hN}} \boldsymbol{a}_{\mathrm{vN}}^H \boldsymbol{w}_v + \boldsymbol{C}_{\mathrm{ce}} \boldsymbol{w}_v \tag{25}$$

Under the single source assumption, the vehicle detection problem in our application can be simplified to a single source azimuth estimation problem from the cross section data $\boldsymbol{z}$ in (24).

From the subspaces perspective introduced in the preceding subsection, it is easy to see that the signal subspace in (24) is spanned by only one vector $\boldsymbol{a}_{h1}$. According to the fundamental subspaces theorem [27] in linear algebra, the noise subspace $\boldsymbol{\mathcal{E}}$, which is the orthogonal complement of the signal subspace $\boldsymbol{\mathcal{S}}$, can be found as the null space of $\boldsymbol{z}^H$ if the diffuse noise is not salient, as depicted in (26), where $N(\cdot)$ for null space.

$$\boldsymbol{\mathcal{E}} = \mathrm{span}^{\perp}(\boldsymbol{z}) = N(\boldsymbol{z}^H) \tag{26}$$

According to the definition of null space [27], the basis vectors $\boldsymbol{u}$ which span the subspace $N(\boldsymbol{z}^H)$ can be found by solving the linear equation in (27).

$$\boldsymbol{z}^H \boldsymbol{u} = 0 \tag{27}$$

Since there are $M_h - 1$ free variables in (27), the subspace $N(\boldsymbol{z}^H)$ can be spanned by $M_h - 1$ vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M_h-1}$ with the format in (28). And u_{1m}, $m \in \{1, \ldots, M_h - 1\}$ can easily be derived by the closed-form solution in (29) and (30), where $\mathrm{re}(\cdot)$ and $\mathrm{im}(\cdot)$ for the real and the imaginary part of a complex number.

$$\begin{aligned} \boldsymbol{u}_1 &= [u_{11}, -1, 0, \ldots, 0]^T \\ \boldsymbol{u}_2 &= [u_{12}, 0, -1, \ldots, 0]^T \\ &\vdots \\ \boldsymbol{u}_{M_h-1} &= [u_{1\ M_h-1}, 0, 0, \ldots, -1]^T \end{aligned} \tag{28}$$

$$\mathrm{re}(u_{1m}) = \frac{\mathrm{re}(z_1)\mathrm{re}(z_{m+1}) + \mathrm{im}(z_1)\mathrm{im}(z_{m+1})}{\mathrm{re}(z_1)^2 + \mathrm{im}(z_1)^2} \tag{29}$$

$$\mathrm{im}(u_{1m}) = \frac{\mathrm{im}(z_1)\mathrm{re}(z_{m+1}) - \mathrm{re}(z_1)\mathrm{im}(z_{m+1})}{\mathrm{re}(z_1)^2 + \mathrm{im}(z_1)^2} \tag{30}$$

Finally, the proposed rank-1 MUSIC spectrum can be derived according to (31), where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M_h-1}]$ is derived from (28)–(30), and each scanning direction φ forms a fine detection zone in Fig. 2. In addition to the vehicle azimuth, vehicle approaching and leaving status are also required for vehicle appearance decision. So, the numerator $|\boldsymbol{z}|$, which represents the energy in the principal subspace direction, is introduced to control the spectrum amplitude in (31).

$$q(\varphi) = \frac{|\boldsymbol{z}|}{\left|\boldsymbol{U}^H \boldsymbol{a}_h(\varphi)\right|^2} \tag{31}$$

C. Theoretical Analysis

Comparing the rank-1 MUSIC algorithm in (14), (28)–(31) with the original MUSIC algorithm in (22), we can find that the rank-1 approach directly works on the cross section data in (14), and the EVD procedure is not required. To further explain the proposed algorithm, let's reformulate the correlation matrix of the original MUSIC algorithm in the single source case, which yields equation (32), where $\boldsymbol{x} = \boldsymbol{a}_1 s_1 + \boldsymbol{e}$.

$$\boldsymbol{C}_x = E\{\boldsymbol{x}\boldsymbol{x}^H\} = \sigma_1^2 \boldsymbol{a}_1 \boldsymbol{a}_1^H + \boldsymbol{C}_e \tag{32}$$

In (32), the signal term $\sigma_1^2 \boldsymbol{a}_1 \boldsymbol{a}_1^H$ is a rank-1 matrix. When the diffuse noise term $\boldsymbol{C}_e$ is not salient, we will find by EVD that the leading eigenvector $\boldsymbol{v}_1 = \boldsymbol{a}_1$, and other eigenvectors $\boldsymbol{v}_2, \ldots, \boldsymbol{v}_M$ are orthogonal to $\boldsymbol{a}_1$. Then, $\boldsymbol{U} = [\boldsymbol{v}_2, \ldots, \boldsymbol{v}_M]$ can be used by the original MUSIC algorithm in (22).

On the other hand, the correlation of the cross section data in (24) gives us equations (33)–(35), where $\boldsymbol{R}$ is a residual term.

$$\boldsymbol{Z} = \boldsymbol{z}\boldsymbol{z}^H = \sigma_1^4 \varepsilon^2 \boldsymbol{a}_{h1} \boldsymbol{a}_{h1}^H + \boldsymbol{R} \tag{33}$$

$$\varepsilon^2 = \boldsymbol{a}_{v1}^H \boldsymbol{w}_v \boldsymbol{w}_v^H \boldsymbol{a}_{v1} \tag{34}$$

$$\boldsymbol{R} = \sigma_1^2 \boldsymbol{a}_{h1} \boldsymbol{a}_{v1}^H \boldsymbol{w}_{v1} \boldsymbol{r} + \boldsymbol{r} \boldsymbol{w}_{v1}^H \boldsymbol{a}_{v1} \boldsymbol{a}_{h1}^H \sigma_1^2 + \boldsymbol{r}\boldsymbol{r}^H \tag{35}$$

Comparing (32) with (33), if the EVD is performed on $\boldsymbol{Z}$, and the 2nd to the Mth eigenvectors are used to span the noise subspace, then, we can find that it is equivalent to apply the original MUSIC algorithm on the horizontal subarray data. That's the reason why the proposed approach can estimate the source DOA in the cross section. However, the noise subspace can directly be estimated from the cross section data $\boldsymbol{z}$ by (28)–(30), and no explicit EVD is required. In (33), the matrix $\boldsymbol{Z}$, which is used to theoretically estimate the noise basis, is a rank-1 matrix, so, we name the proposed approach as the rank-1 MUSIC algorithm.

When there are multiple sources in the cross section, the vector $\boldsymbol{z}$ can no longer be derived from (23) to (24), which means that the signal and the noise subspaces cannot easily be derived. However, if we still calculate the noise subspace according to (28)–(30), the resulted subspace $\mathrm{span}(\boldsymbol{U})$ will overlap with the signal subspace. This is because the dimension of $\mathrm{span}(\boldsymbol{U})$ is $M_h - 1$, and the dimension of the signal subspace is greater than 1 in the multisource cases. Once the two subspaces are overlapped, the orthogonality between $\boldsymbol{U}$ and $\boldsymbol{a}_h$ is destroyed, so, the spectrum in (31) yields negligible response. Although the rank-1 MUSIC algorithm fails to detect multi-sources, traffic monitoring will not be affected. This is because lane positions are estimated from the vehicle DOA statistics, a few of misses will not affect the lane detection result. In addition, lane-wise traffic monitoring is carried out by coarse detection zones in Fig. 2, instead of the lane detection algorithm. Multi-vehicles in different lanes will separately be handled by the coarse detection zones in each lane.

As the conclusion to this section, we derive the computational complexity of the rank-1 MUSIC algorithm, and compare it with several other DOA estimation approaches. To perform fairly comparison, we suppose the correlation matrices are already available. For the planar array approach in [2], since the data in the same array columns are add together to form the cross-road cross section, the resulted correlation matrix $\boldsymbol{C}_x$ has the size of $M_h \times M_h$. While for the cross array approaches, the correlation matrix $\boldsymbol{C}_{\mathrm{cx}}$ is calculated according to (12), which has the size of $M_h \times M_v$.

First, the DS approach in (36), which is used in [2] for lane detection, takes the time complexity of $O(M_h^2)$ to calculate the energy spectrum at the look direction φ. It's cross array version, which is described in (13), takes the time complexity of $O(M_h M_v)$. Since $M_h > M_v$ holds in our cross array designs in Table III, the complexity of (36) is slightly higher than (13).

$$q(\varphi) = \boldsymbol{a}_h^H(\varphi)\boldsymbol{C}_x\boldsymbol{a}_h(\varphi) \tag{36}$$

The Capon-MVDR in (37) [28] is another frequently used DOA estimation approach, which is also used in the lane-wise vehicle detection module in [2]. To calculate the energy spectrum in (37), the inverse of the correlation matrix should be calculated first, which has the time complexity of $O(M_h^3)$, so, the final time complexity of (37) is also $O(M_h^3)$.

$$q(\varphi) = \frac{1}{\boldsymbol{a}_h^H(\varphi)\boldsymbol{C}_x^{-1}\boldsymbol{a}_h(\varphi)} \tag{37}$$

For the original MUSIC algorithm in (22), the noise subspace is derived from the EVD of the correlation matrix $\boldsymbol{C}_x$, so, the time complexity is $O(M_h^3)$. While in the rank-1 MUSIC approach, the cross section data $\boldsymbol{z}$ should be calculated first by (14), with the time complexity of $O(M_h M_v)$. Then, $M_h - 1$ basis vectors are calculated by (28)–(30), with the time complexity of $O(M_h)$. At last, the energy spectrum is calculated by (31), with the time complexity of $O(M_h^2)$. So, the total time complexity of the rank-1 MUSIC algorithm is $O(M_h^2)$, which is just the same level as the DS approach used in the lane detection module in [2]. However, from the experiments in Section V, we will see that the rank-1 MUSIC algorithm has higher resolution than the DS approach.

IV. Probabilistic Models for Automatic Gain Control and Lane Detection

A. Automatic Gain Control (AGC)

Before extracting the vehicle azimuth from (31), an AGC module is required to normalize the energy spectrum into the interval of [0, 1], which is convenient for vehicle appearance decision and data visualization (see Fig. 8). The amplitude variation of the azimuth spectrum can be roughly classified into two categories: If there is a vehicle in the lane detection cross section, the spectrum will exhibit high amplitude, with the peak reveals the vehicle azimuth. However, if there are no vehicles passed by, the spectrum will stay in a relatively low level. From the preceding analysis, the azimuth spectrum can be normalized by (38), where ← means assignment, β is the background threshold, which is used remove the background noise response, and α is the foreground threshold, which is used for the amplitude normalization, $\hat{q}$ is current local maximum which is given in (39). After the energy is normalized, a fixed threshold γ ($\gamma = 0.5$ in our system) can be used to determine the vehicle appearance in the lane detection cross section.

$$q \leftarrow \begin{cases} 0, & q < \beta \\ \frac{(q-\beta)}{(\hat{q}-\beta)}, & q > \alpha \\ \frac{(q-\beta)}{(\alpha-\beta)}, & \text{otherwise} \end{cases} \tag{38}$$

To automatically choose α and β, the amplitude variation of the azimuth spectrum is modeled by its probability density function (PDF), which is estimated by the Parzen window technique [29] as follows. First, a local maximum $\hat{q}$ of the spectrum is selected by (39), where $\Delta\tau_{\mathrm{agc}}$ is a local time interval (e.g., no. of STFT frames corresponds to 2 seconds in our system) before current time. The usage of local maxes is learned from the AGC module in [2], which uses the average of the local maxes as the normalization threshold. Since the local maximum can roughly generalize the max response caused by a single vehicle or caused by background noise, it is a good approximation in the energy pdf estimation.

$$\hat{q} = \max_{\varphi \in [-90^\circ, 90^\circ], \tau \in \Delta\tau_{\mathrm{agc}}} q(\varphi, \tau) \tag{39}$$

As the time goes on, after a new local time interval is reached, the Gaussian kernel used for Parzen window estimation is constructed by (40), where σ_{agc} ($\sigma_{\mathrm{agc}} = 2$ in our AGC module according to the "three-sigma rule") controls the window width. The pdf of the azimuth spectrum is updated according to (41), where ω ($\omega = 0.999$ in our AGC module) is the forgetting factor, which makes the pdf, as well as α and β, can be adaptively adjusted with time.

$$g(q) = \frac{1}{\sqrt{2\pi}\sigma_{\mathrm{agc}}} \exp\left[-\frac{(q-\hat{q})^2}{2\sigma_{\mathrm{agc}}^2}\right] \tag{40}$$

$$p(q) \leftarrow \omega p(q) + (1-\omega) g(q) \tag{41}$$

Typical pdf estimated by (31), (39)–(41) has the bimodal structure as shown in the example of Fig. 5, where the first and the second modal stands for the energy distribution of the background noise and the passing vehicles, respectively. Obviously, after this pdf is estimated, α can be set to the value corresponding to the second modal, and β can be set to the value corresponding to the valley between the two peaks, as shown in Fig. 5. Sometimes, when the traffic load is very heavy, the bimodal pdf in Fig. 5 may degenerate to the unimodal structure since there may always have vehicles in the lane detection cross section. In this degenerated case, which is equivalent to the local maxes averaging approach in [2], α can be set to the value corresponding to the single peak of the curve, and β can be set several dBs (6 dB is adequate) below α [2].

B. Lane Detection

Since vehicle usually travels along lane centers and seldom crosses lane boundaries, the observed vehicle azimuths can

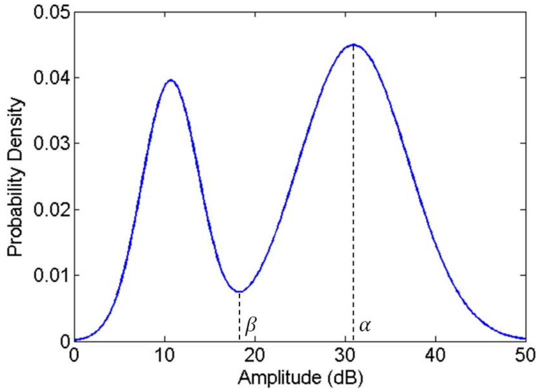


Fig. 5. Typical pdf of the amplitude local maxes. This curve is derived from real vehicular sound.

also be modeled by a probabilistic based approach. After the corresponding pdf is estimated, lane centers can be detected as the peaks of the pdf, and lane boundaries can be detected as the valleys adjacent to the peak [1], [2]. Please note that the detected lane positions do not necessarily have to be the same as the physical lane look directions as illustrated in Fig. 1 and Table I, also, they may vary with time due to the changes of weather and traffic conditions [1].

A normalized power accumulation (NPA) based approach is introduced in [1] and [2] to estimate the pdf from the observed azimuth spectrums. The basic idea of this approach is shown in (42), which has the similar idea as (41). In (42), $p(\varphi)$ indicates the azimuth pdf, $q(\varphi)$ is the azimuth spectrum from (36), then, normalized by (38). Meanwhile, $\max q(\varphi) > \gamma$ should be fulfilled, which indicates the presents of a vehicle

$$p(\varphi) \leftarrow \omega p(\varphi) + (1-\omega) q(\varphi). \tag{42}$$

Although this approach works well in [1] and [2], it is not suitable for the spectrum derived from (31). Because of the super resolution property of the MUSIC algorithm, the resulted spectrum usually contains sharp spikes at vehicle DOAs, instead of bell-shaped peaks in [1] and [2]. It means that the resulted pdf by (42) will not smooth enough for peak and valley detection, as shown in Fig. 6(a).

Reconsider the output of the vehicle detection algorithm in the previous section, the detected vehicle azimuths can be considered as examples sampled from the underlying unknown vehicle azimuth pdf. So, the Parzen window technique [29] can also be used here to estimate the pdf from a finite number of observed vehicle azimuths. The detailed steps are depicted in (43)–(45), which are similar to (39)–(41). First, in (43), the peak angle $\hat{\varphi}$ is selected from the result of (31) as the vehicle DOA (when $\max q(\varphi) \geq \gamma$ is satisfied), then, the Gaussian window is constructed according to (44), at last, the pdf is updated according to (45)

$$\hat{\varphi} = \underset{\varphi \in [-90^\circ, 90^\circ]}{\arg\max}\ q(\varphi) \tag{43}$$

$$g(\varphi) = \frac{1}{\sqrt{2\pi}\sigma_{\text{lane}}} \exp\left[-\frac{(\varphi - \hat{\varphi})^2}{2\sigma_{\text{lane}}^2}\right] \tag{44}$$

$$p(\varphi) \leftarrow \omega p(\varphi) + (1-\omega) g(\varphi). \tag{45}$$

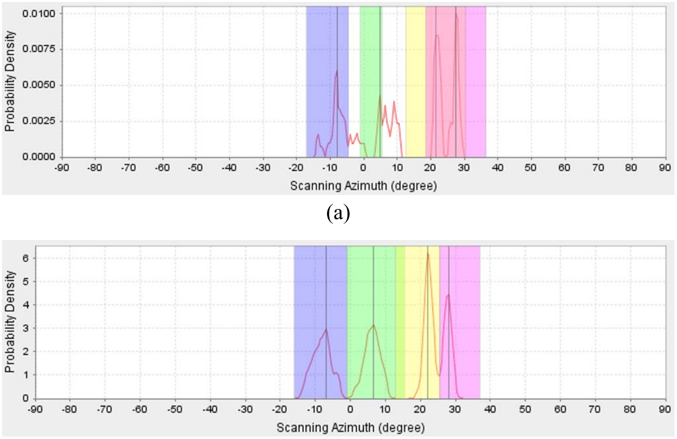


Fig. 6. Comparison of the NPA based and the Parzen window based lane detection approaches. The red curve is the estimated pdf, and different lanes are detected and marked in different colors. Both approaches use the result of (31) as the input, 150 vehicles are simulated in the environment of Fig. 1 for the pdf estimation. (a) The NPA based approach. (b) The Parzen window based approach.

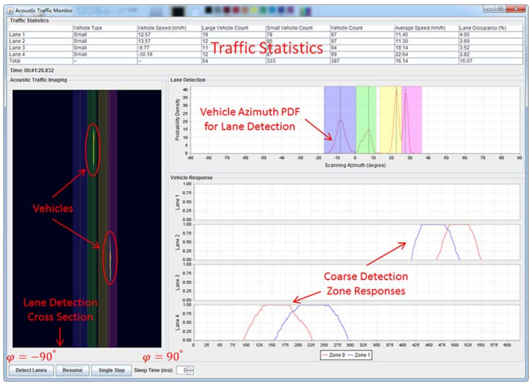
In (44), the value of σ_{lane} can be derived from the minimal lane width of the road environment. From Table I, $\sigma_{\text{lane}} = 4.5/6 = 0.75$ is used in our system according to the "three-sigma rule" of the Gaussian function.

An example of detected lane positions by the Parzen window based approach is given in Fig. 6(b). Compared with Fig. 6(a), all four lanes in the simulated environment of Fig. 1 are clearly detected in Fig. 6(b). After lanes are detected, the coarse detection zones in Fig. 2 can be constructed by existing beamforming techniques. Please refer to [30] for the full acoustic traffic monitoring system construction.

V. Experiment

All experiments are carried out on a prototype system developed in Java, as shown in Fig. 7. There are four main parts in the system GUI: The table on the top shows the lane-wise statistical information of the traffic flow, including vehicle type (large vs. small), vehicle speed, vehicle count, average speed, and lane occupancy. The acoustic traffic imaging (ATI) panel on the left visualizes the response of the lane detection cross section, i.e., the normalized result from (31), (36), or (37). As the system is running, the ATI is scrolling upward, with red spots entering the ATI from the bottom, which reveal the high energy parts caused by the passing vehicles. In all experiments, the scanning azimuth φ ranges from -90° to 90°, with $\Delta\varphi = 0.5^\circ$. The lane detection panel on the middle right shows the cumulated energy curve or pdf curve used for lane detection. Lane centers are marked as gray vertical lines, and different lanes are marked in different colors. The detected lanes are also painted on the ATI with the same colors. The lower right part of the GUI is the vehicle response panel, which visualizes the responses from different coarse detection zones. These curves are used for lane-wise traffic monitoring.

Three DOA estimation approaches and two lane detection approaches are mentioned in this paper, including the DS



Acoustic Traffic Monitor

Traffic Statistics

	Vehicle Type	Vehicle Speed (km/h)	Large Vehicle Count	Small Vehicle Count	Vehicle Count	Average Speed (km/h)	Lane Occupancy (%)
Lane 1	Small	12.57	19	78	97	11.45	4.05
Lane 2	Small	13.57	12	85	97	11.30	3.69
Lane 3	Small	-9.77	11	83	94	19.14	3.52
Lane 4	Small	-30.19	12	87	99	22.64	3.82
Total	--	--	54	333	387	16.14	15.07

Time: 00:41:28.832

Fig. 7. The prototype system GUI.

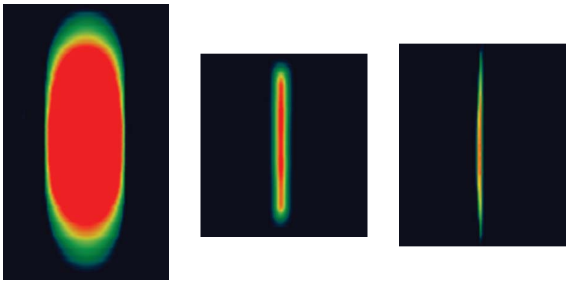


Fig. 8. ATI of different DOA estimation approaches. (a) The DS approach. (b) The MVDR approach. (c) The rank-1 MUSIC approach.

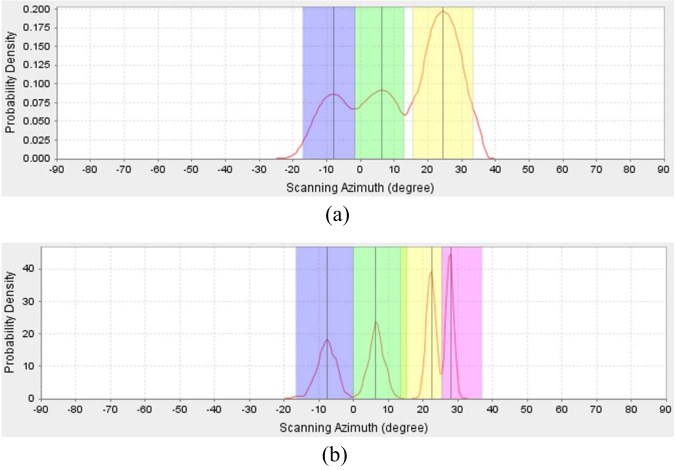


Fig. 9. Lane detection results of the DS + NPA approach and the MVDR+ PARZEN approach in the simulated environment. (a) The DS + NPA approach. (b) The MVDR + PARZEN approach.

approach in (36) [2], the MVDR approach in (37) [28], the proposed rank-1 MUSIC algorithm in (31), the NPA based approach in (42) [1], [2], and the proposed Parzen window based approach (PARZEN) in (43)–(45). In the following experiments, three combinations of these techniques are compared, including DS + NPA, MVDR + PARZEN, and MUSIC + PARZEN.

A. *Simulated Experiment*

To perform comparison, a simulated highway environment is established according to Fig. 1, and the array scheme no. 1 in Table III is used. One moving vehicle is simulated by the Gaussian white noise with 20 kHz sampling rate. The incident azimuth of the sound source is randomly sampled from four Gaussians which are used to model the four lanes. The parameters of the four Gaussians are derived according to Table I. After the vehicle azimuth is chosen, the corresponding elevation is varying from -90° to 90° for lane 1 and lane 2, and from 90° to -90° for lane 3 and lane 4 during the simulation, which simulates the up and down-road traveling vehicles. In all simulations, two vehicles with different lanes and different speed (elevation changing rate) are simulated traveling simultaneously in the environment of Fig. 1. According to the environmental configuration, the look directions of the four lane centers are: -7.9°, 6.6°, 22.4°, and 27.7°, respectively.

First, Fig. 8 shows the single-vehicle ATI results from the compared DOA estimation algorithms. In Fig. 8, we can see that a simulated vehicle is visualized as a “blob” in the DS approach, however, as a “line” in the MVDR and the MUSIC approaches, which means that the latter two approaches have much higher resolution than the first one. From the width of the vehicle visualization, we can see that the proposed approach has the highest resolution.

The high resolution property of the proposed approach can be further demonstrated in Fig. 9. Fig. 9(a) gives the lane detection result of the DS + NPA approach, the three peaks are detected as: -8°, 6.5°, and 24.5°. Comparing this result with the true lane centers, we can find that lane 1 and lane 2 are correctly detected within the error tolerance of $\Delta\varphi = 0.5^\circ$. However, due to the insufficient of the resolution, the algorithm cannot distinguish lane 3 and lane 4. Instead, the farthest two lanes are incorrectly detected as one lane, with the center as the average of them. Fig. 9(b) gives the lane detection result of the MVDR + PARZEN approach. The four detected lane centers are -7.5°, 6.5°, 22.5°, and 28.0°, respectively, which are in the error tolerance of the true lane centers. The result of the MUSIC + PARZEN approach is shown in Fig. 6(b), and the same lane centers are detected as the MVDR + PARZEN approach.

B. *Real-World Experiment*

The simulated experiment shows that the array scheme no. 1 in Table III, which has the lowest cost, is adequate for basic traffic monitoring. However, the array scheme no. 5 is used in the real-world experiment, in order to leave us enough resolution margins for algorithm and array topology optimization. In real-world experiment, the environmental parameters are almost the same as Fig. 1, except that the road is six-lane bidirectional highway, and the microphone array is mounted at the height of 5.5 m, as shown in Fig. 10. 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

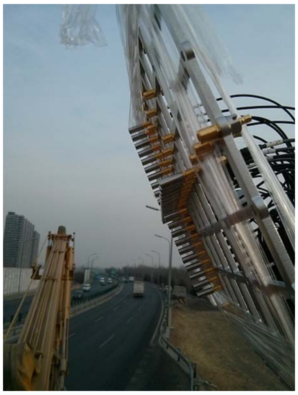
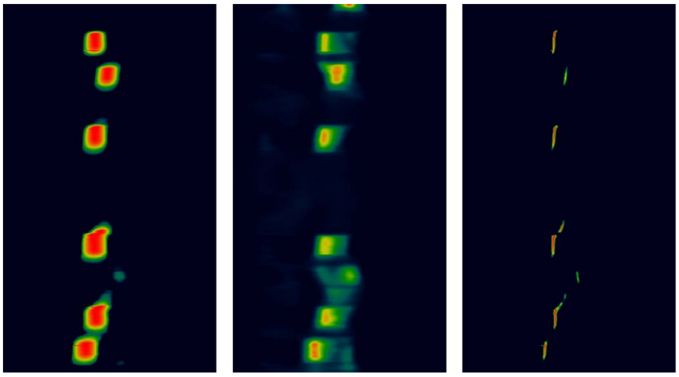
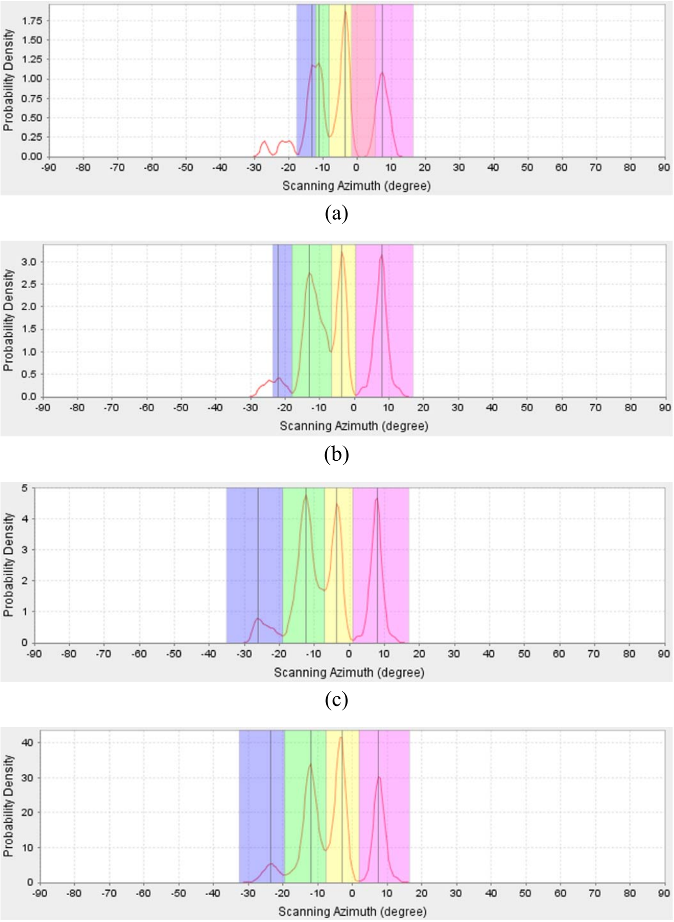


Fig. 10. Real-world experimental environment. This photograph was taken in the microphone array installation procedure.



(a) (b) (c)

Fig. 11. ATIs generated from real data. The signal length is 10 seconds, and the same time window is used for all images. (a) The DS approach. (b) The MVDR approach. (c) The rank-1 MUSIC approach.



(a)

(b)

(c)

(d)

Fig. 12. The lane detection procedure in real-world experiment. (a) 50 vehicles (3 minutes). (b) 100 vehicles (5 minutes). (c) 150 vehicles (7 minutes). (d) (45 minutes).

developed prototype system was also deployed in the NI system for real-time traffic monitoring.

Fig. 11 shows the ATI examples of the three compared DOA estimation algorithms. The working frequency band of 6.5 to 7 kHz was used. Comparing the practical ATIs in Fig. 11 with the simulated results in Fig. 8 we can find that, the visualizations are consistent for both the DS and the MUSIC approaches. However, for the MVDR approach, the visualized vehicle spots from the real data are much thicker than the simulation, which means that the algorithm resolution degrades a lot. We deduce that the reason is the lack of the vertical beamformer: In the MVDR approach, DOA spectrum is calculated from the general form of the correlation matrix in (4), the special topology of the cross array cannot be utilized. Without the help of the vertical beamformer in (14), sounds outside the lane detection cross section cannot be attenuated before MVDR. Thus, the enormous diffuse noise in the real highway environment will degrade the algorithm performance [31], which can be understood as using an array with finite degree of freedom to attenuate infinite number of interferences.

The lane detection result of the proposed MUSIC + PARZEN approach is shown in Fig. 12. We can see from Fig. 12 that after about 150 vehicles passed by, the resulted vehicle azimuth pdf almost converges to the structure of four peaks, i.e., four lanes are detected. According to the highway configuration and the detected vehicle traveling directions, it can be inferred that lane 1, 2, and 3 (blue, green, and yellow) are the three up-road lanes, and lane 4 (pink) is a down-road lane (or lanes, because of the phenomenon in Fig. 9(a)). We deduce that there may be several reasons why only four of six lanes are detected. First, the array resolution may still not high enough. Second, sound energy from the farthest two lanes may be too small to be detected. Third, the sound field may be jammed by the sound barrier at the opposite side of the road, as shown in Fig. 10. However, since the three up-road lanes are successfully detected, the proposed system is adequate for one-side traffic monitoring in the environment of Fig. 10.

In the developed prototype system, four types of traffic quality measuring indices are derived, including vehicle count, lane occupancy, vehicle speed, and vehicle type (large vs. small). Real-world experiments have also been conducted to compare the traffic monitoring accuracy between the audio based approach and the video based ground truth. Please see [30] for the detailed description.

VI. Conclusion and Future Work

Lane detection is the first step of multi-lane traffic monitoring. This paper has proposed an acoustic based lane detection approach, which can automatically detect lane positions and widths from the vehicle emitting sounds. Compared with other traffic monitoring techniques, acoustic based feature is robust against light and weather variations, which is helpful for robust traffic monitoring. In addition, acoustic sensor (microphone) is relatively inexpensive than other traffic monitoring sensors like camera and radar.

In order to perform acoustic based traffic monitoring, a microphone array is designed according to a typical Chinese highway configuration. The adopted array topology is based on the cross array structure, which is suitable to form different vehicle detection zones via beamforming algorithms. The cross correlation matrix from the horizontal and the vertical subarrays in the selected working frequency band is calculated for the following lane detection and vehicle monitoring operations.

The proposed lane detection approach comprises of two steps: First, forming a lane detection cross section, and detecting the azimuths of passing vehicles; second, utilizing azimuth statistics to find lane centers and widths. In the first step, with the help of the vertical beamformer, the single sound source assumption can be applied to the resulted array data. So that the MUSIC based sound source DOA estimation algorithm is simplified to its rank-1 version for vehicle detection. The new algorithm doesn't need to perform EVD, so that it has lower computational complexity. In the second step, vehicle azimuth pdf is estimated by the proposed Parzen window based technique via an online updating manner. Since vehicle usually travels along lanes and seldom crosses lane boundaries, the positions of lane centers and boundaries can be revealed from the peak and valley patterns of the estimated pdf. Both simulated and real-word experiments are conducted on a traffic monitoring prototype system developed in Java, and the efficiency of the proposed approach is shown from the resulted ATI and vehicle azimuth pdf comparisons.

In future work, we will mainly focus on the acoustic based lane-wise traffic monitoring approaches, e.g., how to construct high resolution coarse detection zones according to the detected lanes. In addition, the way to derive and improve different traffic quality measuring indices will also be investigated.

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, and X. Sun, "Adaptive traffic lane detection based on normalized power accumulation," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 2008, pp. 968–973.

[2] J. P. Kuhn, B. C. Bui, and G. J. Pieper, "Acoustic Sensor System for Vehicle Detection and Multi-Lane Highway Monitoring," U.S. Patent 5 798 983, Aug. 25, 1998.

[3] S. Taghvaeeyan and R. Rajamani, "Portable roadside sensors for vehicle counting, classification, and speed measurement," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 73–83, Feb. 2014.

[4] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, "Vehicular traffic density state estimation based on cumulative road acoustics," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1156–1166, Sep. 2012.

[5] I. McCowan, *Microphone Arrays: A Tutorial*. Brisbane, Qld, Australia: Queensland Univ., 2001, pp. 1–38.

[6] K. Kodera, A. Itai, and H. Yasukawa, "Approaching vehicle detection using linear microphone array," in *Proc. IEEE Int. Symp. Inf. Theory Appl.*, 2008, pp. 1–6.

[7] P. Marmaroli, M. Carmona, J. Odobez, X. Falourd, and H. Lissek, "Observation of vehicle axles through pass-by noise: A strategy of microphone array design," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1654–1664, Dec. 2013.

[8] S. Chen, Z. Sun, and B. Bridge, "Traffic monitoring using digital sound field mapping," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1582–1589, Nov. 2001.

[9] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Berlin, Germany: Springer-Verlag, 2008.

[10] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave pattern," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 30–47, Jan. 2009.

[11] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[12] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[13] N. Ito, E. Vincent, N. Ono, and S. Sagayama, "Robust Estimation of Directions-of-Arrival in Diffuse Noise Based on Matrix-Space Sparsity," Institut national de recherche en informatique et en automatique, (INRIA) Rocquencourt, France, Tech. Rep.RR-8120, 2012.

[14] L. Kumar, A. Tripathy, and R. M. Hegde, "Robust multi-source localization over planar arrays using MUSIC-group delay spectrum," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4627–4636, Sep. 2014.

[15] F. Yan, M. Jin, S. Liu, and X. Qiao, "Real-valued MUSIC for efficient direction estimation with arbitrary array geometries," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1548–1560, Mar. 2014.

[16] J. J. Christensen and J. Hald, *Beamforming Technical Review*, H. K. Zaveri, Ed. Naerum, Denmark: Bruel & Kjaer, 2004.

[17] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Process.*, vol. 93, no. 12, pp. 3264–3277, 2013.

[18] B. Rafaely, "Spatial-temporal correlation of a diffuse sound field," *Acoust. Soc. Amer.*, vol. 107, no. 6, pp. 3254–3258, Jun. 2000.

[19] N. Ito, N. Ono, E. Vincent, and S. Sagayama, "Designing the wiener postfilter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 2818–2821.

[20] B. R. Slattery, "Use of mills cross receiving arrays in radar systems," *Proc. Inst. Electr. Eng.*, vol. 113, no. 11, pp. 1712–1722, Nov. 1966.

[21] E. F. Berliner, J. P. Kuhn, S. A. Rawson, and A. D. Whalen, "Acoustic highway monitor," U.S. Patent No. 6,195,608, Feb. 27, 2001.

[22] R. H. MacPhie, "A mills cross multiplicative array with the power pattern of a conventional planar array," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, 2007, pp. 5961–5964.

[23] X. Guo, S. Yang, and H. Zhang, "A low-frequency superdirective acoustic vector sensor array," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2015, pp. 2734–2738.

[24] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing Principles, Algorithms, And Applications*, 4th ed. Beijing, China: House Electron. Ind., 2010.

[25] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[26] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

[27] S. J. Leon, *Linear Algebra With Applications*, 7th ed. Beijing, China: China Machine Press, 2007.

[28] T. L. Marzetta, S. H. Simon, and H. Ren, "Capon-MVDR spectral estimation from singular data covariance matrix, with no diagonal loading," *Proc. 14th Annu. Workshop ASAP, MIT Lincoln Laboratory*, 2006, pp. 1–6.

[29] E. Parzen, "On estimation of a probability density function and mode" *Ann. Math. Statist.*, pp. 1065–1076, Sep. 1962.

[30] Y. Na, Y. Guo, Q. Fu, and Y. Yan, "An acoustic traffic monitoring system: Design and implementation," *Proc. 12th IEEE Int. Conf. Ubiquitous Intell. Comput.*, 2015, pp. 119–126.

[31] C. Pan, J. Chen, and J. Benesty, "On the noise reduction performance of the MVDR beamformer in noisy and reverberant environments," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 815–819.

Yueyue Na received the B.S., M.S., and Ph. D. degrees in computer science and technology from Beijing Jiaotong University, Beijing, China, in 2005, 2008, and 2014, respectively. He is currently a Postdoctoral Associate with the Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China. His research interests include microphone array signal processing, blind source separation, and machine learning.

Yanmeng Guo received the B.E. degree in electronic information engineering from Southeast University, Nanjing, China, in 1999, and the M.S. and Ph.D. degrees in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, in 2002 and 2007, respectively. She is currently an Associate Professor with the Institute of Acoustics, Chinese Academy of Sciences. Her current research interests include front-end processing for speech and audio applications, such as sound source localization, target speech detection, and speech enhancement.

Qiang Fu received the B.E. degree from Xi'an Technological University, Xi'an, China, in 1994, the M.S. degree in electronic engineering from Chongqing University of Posts and Telecommunications, Chongqing, China, in 1997, and the Ph.D. degree in electronic engineering from Xidian University, Xi'an, in 2000. In 2000, he was working as a Researcher in Motorola China Research Center, Shanghai, China. From 2001 to 2002, he was working as a Senior Research Associate with the Center for Spoken Language Understanding, OGI School of Science and Engineering, Oregon Health & Science University, Oregon, USA. From 2002 to 2004, he was working as a Senior Postdoctoral Research Fellow with the Department of Electric and Computer Engineering, University of Limerick, Ireland. He is currently a Professor with the Institute of Acoustics, Chinese Academy of Sciences, China. His research interests include speech analysis, microphone array processing, far-field speech recognition, audio–visual signal processing, and machine learning for signal processing. Dr. Fu is a member of the IEEE Signal Processing Society. He received the Outstanding Science and Technology China Academy Award in 2014.

Yonghong Yan received the B.E. degree with the Department of Electronic Engineering, Tsinghua University, in 1990, and the Ph.D. degree in computer science and engineering from Oregon Graduate Institute of Science and Engineering (OGI) in 1995. From 1995 to 1998, he worked with OGI as an Assistant Professor, an Associate Director, and an Associate Professor of the Center for Spoken Language Understanding. From 1998 to 2001, he worked as the Principal Engineer of Intel Microprocessors Research Laboratory, Director and Chief Scientist of Intel China Research Center. In 2002, he returned to China to work for Chinese Academy of Sciences. He is a Professor and the Director of Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences. His research interests include large vocabulary speech recognition, speaker/language recognition, and audio signal processing. He is the author of more than 100 papers and is a holder of 40 patents.