# Vehicle Speed Estimation using Acoustic Wave Patterns

Volkan Cevher, *Member, IEEE,* Rama Chellappa, *Fellow, IEEE*
James H. McClellan, *Fellow, IEEE*

*Abstract*— **We estimate a vehicle's speed, its wheelbase length, and tire track length by jointly estimating its acoustic wave pattern with a single passive acoustic sensor that records the vehicle's drive-by noise. The acoustic wave pattern is determined using the vehicle's speed, the Doppler shift factor, the sensor's distance to the vehicle's closest-point-of-approach, and three envelope shape (ES) components, which approximate the shape variations of the received signal's power envelope. We incorporate the parameters of the ES components along with estimates of the vehicle engine RPM, the number of cylinders, and the vehicle's initial bearing, loudness and speed to form a vehicle profile vector. This vector provides a *fingerprint* that can be used for vehicle identification and classification. We also provide possible reasons why some of the existing methods are unable to provide unbiased vehicle speed estimates using the same framework. The approach is illustrated using vehicle speed estimation and classification results obtained with field data.**

## I. INTRODUCTION

Estimation of vehicle motion parameters using signals received at passive sensors is a classical signal processing problem [1]–[6]. When a single passive acoustic sensor is used, wave propagation effects are used to determine the source movements based on the following assumptions that the vehicle $\mathcal{A}$) is a point source [1], [2], $\mathcal{B}$) has stationary signal characteristics that admit a model such as an autoregressive moving average (ARMA) model [2], and $\mathcal{C}$) produces a pure tone [1]. These assumptions are only partially satisfied by vehicles; hence estimation algorithms based on these assumptions do not perform as expected when they are applied to field data.

When an array of passive acoustic sensors is used, existing approaches in the literature concentrate on the correlation among the multiple microphone signals. Forren and Jaarsma [4] aim to classify vehicles based on their axle detections by exploiting the tire noise generated by vehicles. They use signal correlations among three known microphones under assumption $\mathcal{B}$. However, they do not model any interference effects of the tires. Valcarce *et al.* [3] exploit the differential time delays to estimate the speed by assuming $\mathcal{A}$ and $\mathcal{B}$. They use additive Gaussian noise models and obtain biased speed estimates as in [2]. They also explain the bias of their estimates using a delay error term that models the movement of the vehicle during the acoustic signal propagation [7]. Lo and Ferguson [5] develop a nonlinear least squares method for speed estimation using multiple microphones with a quasi-Newton method for computational efficiency. The estimated speed is based on time-delay-of-arrival estimates, obtained using generalized cross correlation method, under assumptions $\mathcal{A}$ and $\mathcal{B}$.

In this paper, we provide a power based algorithm for vehicle speed estimation using a single microphone. We describe the spectral

V. Cevher is with the Electrical and Computer Engineering Department, Rice University, Houston, TX 77005. R. Chellappa is with the Electrical Engineering and Computer Science Departments, University of Maryland, and the Center for Automation Research, University of Maryland College Park, MD 20742. J. H. McClellan is with the Center for Signal and Image Processing, School of ECE, Georgia Institute of Technology, Atlanta GA 30332.

and spatial content of vehicle signals and recast the speed estimation problem as a spatial acoustic pattern recognition problem. We calculate the received signal's power envelope and approximate it using three envelope shape (ES) components. The ES components spatially decompose the total vehicle noise into parts that also account for tire interference effects, tire horn effects, and air turbulence effects, which are not considered in the current literature. For estimation, we introduce a vehicle profile vector that characterizes the ES components and also includes classifying vehicle information such as engine revolutions per minute (RPM) and the number of cylinders. The vehicle profile vector can be regarded as a *fingerprint* of the vehicle.

A motivation for the vehicle profile vector is the following acoustic correspondence problem: given recorded measurements of two vehicles (calibration recordings), we would like to determine, with high confidence, the label of the vehicle when it drives by another control microphone. This problem has applications in distributed sensor networks [8], [9]. The problem becomes complicated when 1) the control microphone has a different distance to the closest point of approach (CPA) of the vehicle, 2) the vehicle is moving with a different speed or moving on a different medium (e.g., gravel as opposed to asphalt), 3) whether or not it is raining (has rained), 4) the vehicle is or was significantly loaded. In this paper, we comment on how we can tackle the correspondence problem using the vehicle profile vector.

The organization of the paper is as follows. Section II elaborates on the spectral and spatial content of the vehicle acoustic signals. Section III casts the acoustic speed estimation problem as a spatio-temporal sampling problem and discusses the scaling ambiguity issue. The signal interference phenomenon of the vehicle tires is introduced in Sect. IV. Then, Sects. V and VI detail the ES components and the method used for vehicle profile vector estimation. Finally, the field experiment results are given in Sect. VII.

## II. VEHICLE SIGNAL'S SPECTRAL AND SPATIAL CONTENT

A vehicle's acoustic signal consists of a combination of various noise signals generated by the engine, the tires, the exhaust system, aerodynamic effects, and mechanical effects (e.g., axle rotation, brake pads, and suspension). Hence, the spectral content of a vehicle's signal includes wideband processes as well as harmonic components. It also has a spatial distribution because the noise sources are at different locations on the vehicle. The mixture weighting of these spectral components at any given location is dependent on the vehicle's speed, whether the vehicle is accelerating, decelerating, turning, and whether the vehicle is in good mechanical condition. In general, one can approximate a well maintained vehicle's signal as consisting of four noise components:

*a) Engine Noise:* The noise from an internal combustion engine contains a deterministic harmonic train and a stochastic component so it can be modeled by the same methods used for human speech [10], [11]. The stochastic component of the engine noise is largely due to the turbulent air flow in the air intake (or intercooler), the engine cooling systems, and the alternator fans. This stochastic component is wideband in nature. The deterministic component is caused by the fuel combustion in the engine cylinders and has more power than

the stochastic component. The lowest deterministic tone is called the cylinder fire rate $f_0$, defined as the firing rate of any one cylinder in the engine. Since each cylinder fires once every two engine revolutions in a four-stroke engine, there is a simple relationship between $f_0$ and the RPM $\chi$ of a vehicle:

$$f_0 = \frac{\chi}{60 \times 2} \text{ Hz}. \tag{1}$$

The strongest tone in the engine noise is called the engine fire rate $F_0$, and it is related to $f_0$ in a simple manner: $F_0 = f_0 \times p$, where $p$ is the number of cylinders in the engine. One can think of $F_0$ and its integer multiples as analogous to the formant frequencies in human speech. The expressions for $f_0$ and $F_0$ model reality quite well; however, small deviations do occur. For example, in modern cars, each cylinder is individually controlled by an engine management system, which might fluctuate $f_0$ and $F_0$ to optimize fuel consumption or torque. Hence, in some cases, the locations of the $f_0$- and $F_0$-harmonics might provide a fingerprint for the specific engine [10].

Car manufacturers try to suppress the engine noise as much as possible for the passengers' comfort inside the vehicle cabin in frequency ranges where human ears are most sensitive to (1kHz to 4kHz) [11]. In addition, the manufacturers try to suppress the noise levels outside the car as mandated by the federal standards for highway noise (e.g., in the US, see [12], [13]). They design quieter engines and also exploit the body of the vehicle to filter the engine noise. To achieve this, the interior of the engine compartment is usually treated with material for acoustical attenuation (the metallic shell also acts as a filter). Hence, in some cases, the engine noise might be stronger on the side and at the very front of the car than other directions, because sound propagation through the axle, the front grill, and the bottom of the engine block cannot be filtered effectively.

*b) Tire Noise:* The term tire noise is defined as the noise emitted from a rolling tire as a result of its interaction with the road surface. The tire noise is the main source of a vehicle's total noise at speeds higher than 50km/h [14]. It consists of two components: vibrational noise and air noise [15], [16]. The vibrational component is caused by the contact between the tire threads and the pavement texture. Its spectrum is most dominant between $100 - 1000$Hz. The air noise is generated by the air being sucked-in or forced out of the rubber blocks of a tire and is dominant in the frequency ranges between 1000 and 3000Hz. The actual frequency calculations are complicated by the tread geometry [17].

In the driving direction of the car, the road and the tire forms a geometrical structure that amplifies the noise generated by the tire-road interaction [16], [18], [19]. This effect is called the *horn effect* and has a directional pattern [18]. This amplification results in a strong vehicle tire noise component in the frequency range $600 - 2000$Hz ( [16]: Chapter 7.1.25). The directivity of the horn effect depends on the tire width and radius, the tire shoulders, the tire tread geometry as well as the weight and torque on the tire. Analytical calculations based on these factors are rather difficult, and hence, numerical approaches such as boundary element methods are used to simulate the horn effect for a given tire configuration [18], [19]. Notably, most of the total tire noise power including the horn effect lies in the frequency range $700 - 1300$Hz with a multi coincidence peak around 1000Hz [16].

*c) Exhaust Noise:* The exhaust system consists of the exhaust manifold, catalytic converter, resonator, exhaust pipe, muffler, and the tail pipe. The system goes from the engine compartment to the back of the car generating the exhaust noise. Due to the system's spatial distribution, this noise is less prominent in the front of a vehicle.

Unlike the engine block noise, the exhaust system noise increases significantly with the engine load. The exhaust noise is also affected by engine turbo/super chargers and after-coolers [20], [21].

Manufacturers use a combination of reactive and absorptive silencers to keep the exhaust noise level down. The exhaust noise has broadband characteristics with most of its power concentrated around low frequencies. It has the same harmonic frequency structure as the engine and has additional tail pipe resonances that occur at the fundamental frequency $f_e = c/(2l)$, where $l$ is the tail pipe length and $c$ is the speed of sound [20], [21].

*d) Air Turbulence Noise:* Vehicle induced turbulence can become an important factor in the overall perceived loudness of a vehicle as the vehicle speed increases. This noise is due to air flow generated by the boundary layer of the vehicle and is prominent immediately after the vehicle passes by the sensor (as a distinctive *whoosh* sound). The turbulence noise depends on the aerodynamics of the vehicle as well as the ambient wind speed and its orientation [22], [23]. In our problem, we only consider the case when the wind speed is much less than the vehicle speed. For this case, perturbation analysis methods can yield analytical expressions for the mean and the variance of the turbulent velocity components [24]. These expressions may be used to further improve the results presented in this paper.

## III. Speed Estimation as a Spatio-Temporal Sampling Problem

To understand the speed estimation problem, it is instructive to consider a monopole source moving with speed $v$ at a constant heading, as illustrated in Fig. 1. Let $z(t) = z(\mathbf{u}_m, t)$ be the continuous acoustic signal observed at the microphone position $\mathbf{u}_m = (x_m, y_m)$ in the $2D$ plane. In this geometry, $y_m$ is defined as the microphone distance to the CPA, and is assumed known. The speed estimation problem then reduces to recovering $v/y_m$ and the time at CPA $t_{\text{cpa}}$, using the microphone signal samples $z[n]$[1], where the sampling frequency is appropriately chosen to avoid frequency aliasing. Because the monopole signal is not known *a priori*, the information needed to determine the speed $v$ is extracted from the physics of the acoustic wave propagation.

### A. Information in Acoustic Wave Phase and Power

Let $s(t)$ be a zero-mean stationary random process that models the acoustic signal emitted by the monopole source. To simplify the results in this section, we concentrate on the following case, where the Fourier transform of the source signal is assumed to be bandlimited as follows:

$$|S(\Omega)| = \begin{cases} S, & \Omega_1 \leq \Omega \leq \Omega_2, \, 2\mathcal{W} \triangleq \Omega_2 - \Omega_1; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The model (2) is quite general because superpositions of signals represented by (2) can be used to model the tire noise and the engine harmonics with proper choices of $\Omega_1$, $\Omega_2$, and $S$ [25].

The source signal at the microphone is related to the monopole signal as follows. When the acoustic propagation speed is much larger than the source speed ($c \gg v$), the *omnidirectional* microphone signal at time $t = n/F_s$ can be written as [26]:

$$z[n] \approx \frac{1}{r^a(t)} s\left(\beta(t)t - \frac{r(t)}{c}\right)\Bigg|_{t=\frac{n}{F_s}} = \frac{1}{r^a[n]} s\left(\beta[n]\frac{n}{F_s} - \frac{r[n]}{c}\right), \tag{3}$$

---

[1] The bracket signal notation indicates sampled versions of continuous signals, e.g., $x[n] = x(t)\big|_{t=n/F_s}$, as defined in [25]. In this paper, a constant sampling frequency $F_s$ is used for time signals unless stated otherwise.
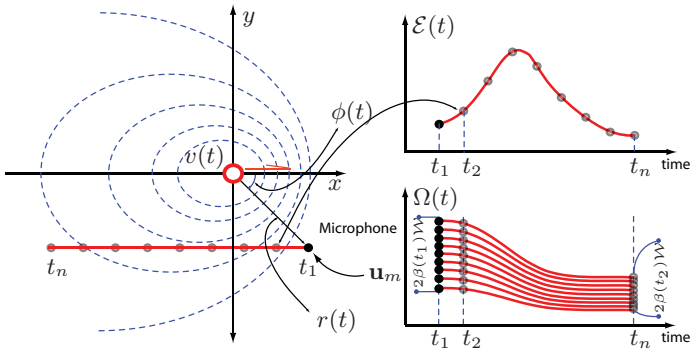
Fig. 1. (*Left*) Geometry of the speed estimation problem. Our objective is to estimate the monopole speed $v$ given the acoustic signal measurements $z(\mathbf{u}_m, t)$ at the stationary microphone $M$ situated at $\mathbf{u}_m = (x_m, y_m)$. As the source moves, the microphone samples the acoustic wave pattern between times $t_1$ and $t_n$. (*Right*) For a constant amplitude monopole source, the envelope of the received signal $\mathcal{E}$ consists of the amplitude samples of the wave pattern generated by the monopole. In this case, the monopole time-frequency bandwidth is modulated by a function $\beta(t)$, called the Doppler shift factor, due to the vehicle movement. As the source approaches and moves away from the microphone, $\beta(t)$ decreases from greater than one to less than one.

where $r(t)$ is the microphone distance to the target, $a$ is the acoustic attenuation constant, $\beta(t) = 1 + v \cos(\phi(t))/c$ is called the Doppler shift factor, and $\phi(t)$ is the microphone bearing in the moving monopole frame with respect to the monopole heading as shown in Fig. 1. For acoustic sources, the attenuation factor $a$ is equal to 1 or $1/2$ for spherical or cylindrical wave propagation, respectively [6]. From now on, we assume spherical propagation and use $a = 1$.

The frequency modulation effect of the Doppler factor on the observed signal $z(t)$ in (3) carries information about the monopole motion (see Fig. 1) [1], [2]. To show the effect, we first derive the $\tau$-sample discrete Fourier transform (DFT) coefficients $Z_n[k]$ of the signal $z[m]$ ($m = n, \ldots, n + \tau - 1$) below ($F_s > 2\beta[n]\mathcal{W}$ for all $n$) [25]:

$$Z_n[k] = F_s Z_n(\Omega)|_{\Omega = \frac{2\pi k}{\tau} F_s}, \tag{4}$$

where $Z_n(\Omega)$ is the Fourier transform of the continuous signal as modeled in (3). Assume that the observation period $\tau$ is small enough that the target is approximately stationary: $\tau/F_s \ll 1/v$. Then, we obtain

$$Z_n(\Omega) = \frac{e^{-j\frac{r[n]}{\beta[n]c}\Omega}}{\beta[n]r[n]} S\left(\frac{\Omega}{\beta[n]}\right). \tag{5}$$

Since the bandwidth of the source signal $S(\Omega)$ is constant $\mathcal{W}$, the bandwidth of the observed signal is modulated by $\beta[n]$ as illustrated in Fig. 1.

Apart from the bandwidth modulation, the wave propagation also affects the received signal power. For an observation period of $\tau$ samples much larger than the inverse bandwidth of the source signal $s(t)$ ($\tau/F_s \times \mathcal{W}/\pi \gg 1$)[2] but short enough so that the target is approximately stationary, the signal power can be determined using Parseval's relation [25]:

$$P_z[n] = \frac{1}{\tau} \sum_{m=n}^{n+\tau-1} |z[m]|^2 = \frac{1}{\tau^2} \sum_{k=0}^{\tau-1} |Z_n[k]|^2. \tag{6}$$
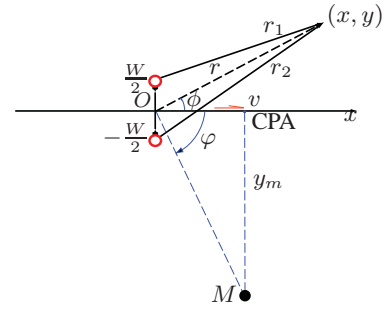


Fig. 2. Dipole geometry. When the dipole sources are correlated, the resulting wave propagation effect on the received signal power is not a superposition of individual monopole effects.

Applying the Euler-MacLaurin summation formula[3] and using (2) and (3), one can approximate (6) as follows ($F_s > 2\beta[n]\mathcal{W}, \forall n$):

$$P_z[n] \approx \frac{F_s}{2\pi\tau} \int_{\beta[n]\Omega_1}^{\beta[n]\Omega_2} |Z_n(\Omega')|^2 d\Omega'$$

$$= \frac{F_s}{2\pi\tau} \int_{\beta[n]\Omega_1}^{\beta[n]\Omega_2} \left| \frac{e^{-j\frac{r[n]}{\beta[n]c}\Omega'}}{\beta[n]r[n]} S\left(\frac{\Omega'}{\beta[n]}\right) \right|^2 d\Omega'$$

$$= \frac{F_s}{2\pi\tau} \int_{\Omega_1}^{\Omega_2} \frac{S^2}{\beta^2[n]r^2[n]} \beta[n]d\Omega = \frac{S^2 F_s \mathcal{W}}{\pi\tau} \times \frac{1}{\beta[n]r^2[n]}. \tag{7}$$

### B. Scaling Ambiguity

The monopole speed estimation using the power modulation of the received signal is prone to a global scaling ambiguity when the Doppler shift factor is approximately 1, $\beta(t) \approx 1$: if the monopole signal amplitude, range, and speed are scaled by the same constant $\eta$, the wave propagation effects on the power of the received signal (7) do not change. This approximation holds typically up to moderate vehicle speeds, e.g., $v \leq 20$m/s. For higher vehicle speeds, for example, typical highway speeds $v = 20 - 40$m/s, the scaling ambiguity is local around the true speed because the variation of the Doppler shift factor due to the scaling is bounded by $2|\eta - 1|v/c \ll 1$. Hence, to determine the absolute speed of the vehicle using a single microphone, a reference distance to the target is needed to resolve the ambiguity. This issue is similar to the world scaling issue commonly encountered in structure from motion algorithms using a single camera [29].

### IV. INTERFERENCE PHENOMENA

When the source signal has a spatial extent, it is crucial to consider the interference effects while estimating the speed. To demonstrate the interference effects, consider a dipole source (e.g., the two front tires of a vehicle) moving along the $x$-axis as illustrated in Fig. 2. In this case, the source signal at the microphone is the sum of the two source signals that are assumed to be coherent:

$$z[n] = \sum_{i=1,2} \frac{1}{r_i[n]} s\left(\beta_i[n]\frac{n}{F_s} - \frac{r_i[n]}{c}\right), \tag{8}$$

where $\beta_i(t)$ ($i = 1, 2$) is the Doppler shift factor of each monopole source in the dipole. It is possible to show that if the following

---

[2]If this asymptotic condition is satisfied, the discrete Fourier coefficients, and hence, $P_z[n]$ and $P_z[n + \tau]$ as defined in (6), are also statistically uncorrelated [27].

[3]Euler-MacLaurin formula: $\sum_{m=1}^{M-1} f[m] \approx \int_0^M f(x)dx - \frac{1}{2}(f(0) + f(M))$. This rule is the dual of the trapezoidal rule of the Newton-Cotes formulas, which is extremely accurate for periodic functions integrated over their periods [28].

condition holds

$$r \gg \frac{W}{2}\left(\sqrt{1+\sin^2\phi}+|\sin\phi|\right),\qquad(9)$$

then one can approximate $r_i \approx r$ and $\beta_i \approx \beta$ as defined in the monopole source case (Fig. 2). Derivation of (9) uses the Taylor series expansion of the range terms and is the dual of the problem in [30] for determining the near field of an array for a point source. The condition (9) defines an approximate boundary of the near field of the dipole source, after which the individual monopole ranges and Doppler shift factors can be approximated.

In the far-field, with the same assumptions for the monopole source, the Fourier transform $Z(\Omega)$ of the signal $z(t)$ can be written as

$$Z(\Omega) \approx \frac{1}{\beta[n]r[n]}S\left(\frac{\Omega}{\beta[n]}\right)e^{-j\frac{\Omega}{\beta[n]c}r_1[n]}\left[1+e^{-j\frac{\Omega}{\beta[n]c}(r_2[n]-r_1[n])}\right].\qquad(10)$$

Hence, the received signal bandwidth is modulated as in the monopole case. However, note that the additional term in the brackets in (10) plays a crucial role when we look at the average received signal power:

$$P_z[n] = \frac{S^2 F_s \mathcal{W}}{\pi\tau} \times \frac{1}{\beta[n]r^2[n]} \times \rho[n],\qquad(11)$$

where $\rho[n]$ is called the interference term:

$$\rho[n] = 2 + \frac{c}{\mathcal{W}(r_2[n]-r_1[n])}\sum_{i=1,2}(-1)^i\sin\left(\frac{\Omega_i}{c}(r_2[n]-r_1[n])\right)$$
$$= 2 + 2\mathrm{sinc}\left(\frac{\mathcal{W}}{c}(r_2[n]-r_1[n])\right) \times$$
$$\cos\left(\frac{(\Omega_1+\Omega_2)}{2c}(r_2[n]-r_1[n])\right).\qquad(12)$$

In this example, the interference term for the dipole has a maximum value of 4 because the summation of two coherent sources results in four times the power as the monopole considered in the previous section. When the dipole source signal is baseband, i.e, $\Omega_1 = -\Omega_2$ and $\mathcal{W} = \Omega_2$, the interference term has the following simpler form:

$$\rho[n] = 2 + 2\mathrm{sinc}\left(\frac{\mathcal{W}}{c}(r_2[n]-r_1[n])\right).\qquad(13)$$

The interference terms in (12) and (13) have a special hyperbolic pattern (Fig. 3). In the far field of the dipole, the interference term is constant along the asymptotes of the hyperbolas defined as $r_2 - r_1 = 2\alpha$ for $\alpha = \frac{W}{2}\sin\phi$. Moreover, it is well-known that the local extremes of the sinc function correspond to its intersections with the cosine function. Hence, a minima and a maxima of the sinc function are on the average half the cosine period away from each other. Thus, for a baseband source, the consecutive local maxima and minima of $\rho[n]$ with respect to its asymptotic angles can be approximately related by

$$\phi_{\max} \approx \sin^{-1}\left[\frac{\pi c}{\mathcal{W}W} + \sin\phi_{\min}\right].\qquad(14)$$

## V. JOINT ESTIMATION OF SPEED AND SPATIAL ACOUSTIC PATTERNS

### A. Envelope Shape Components

To determine a vehicle's speed using acoustic observations from a single microphone, we jointly estimate the vehicle's spatial acoustic pattern. In the previous section, we introduced an interference effect that contributes to the total spatial acoustic pattern, e.g., due to the front tires. We denote any such component that makes up a vehicle's
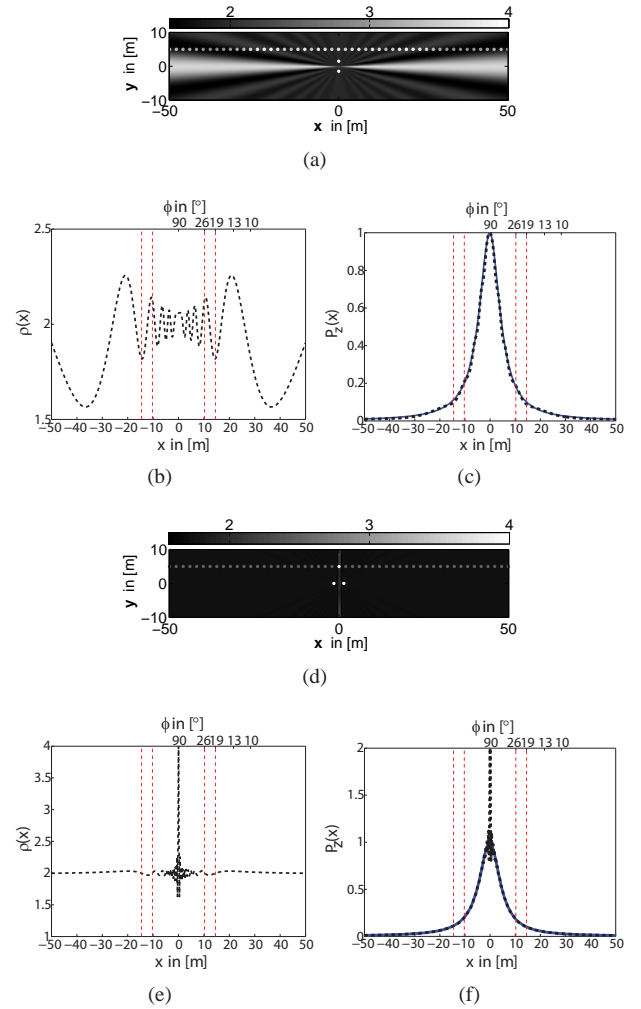


(a)



(b)      (c)



(d)



(e)      (f)

Fig. 3. Interference patterns ((a), (b), (d), (e)) and power functions ((c), (f)) along the $x$ and $y$ directions for a dipole source at the origin with the following parameters: $v = 20$m/s, $\Omega_1 = -2\pi \times 1200$, and $\Omega_2 = 2\pi \times 1200$. For (a)-(c), the main axis of the dipole is the $y$-axis, where $W = 1.5$m, whereas for (d)-(f), the main axis of the dipole is oriented along the $x$-axis, where $W = 3$m. The geometrical configuration of the dipoles are also shown in (a) and (d) along with the path of the target where $y_m = 5$m. In (b) and (d), the interference term $\rho$ (13) is plotted with respect to the distance corresponding to the samples taken along the dots in (a) and (d), respectively. In (c) and (f), the power functions are plotted with (dashed line) and without (solid line) the interference term, corresponding to the samples taken along the dots in (a) and (c), respectively. The power functions are normalized so that their value at $x = 0$m is 1.

spatial acoustic pattern as an *envelope shape* (ES) component. Earlier, we derived the interference effect on the observed acoustic power of the microphone signal *with respect to the source position*. However, in this section, we use the reciprocity theorem and change the reference frame from the moving vehicle to the microphone to derive the ES components [31]. For simplicity, we model the ES components using three piecewise constant functions in dB scale with respect to the microphone bearing $\varphi$ as illustrated in Fig. 4. More components may be used but are likely to lead to overfitting [32]. We make the connection between the ES components and the received signal power in the next subsection.

*The first ES component* $\rho_\gamma(\varphi)$ in Fig. 4 models the signal interference due to the front and rear tires as dipole sources (also refer to Fig. 5). This component explains the perturbation in the envelope function of the vehicle acoustic drive-by signals between
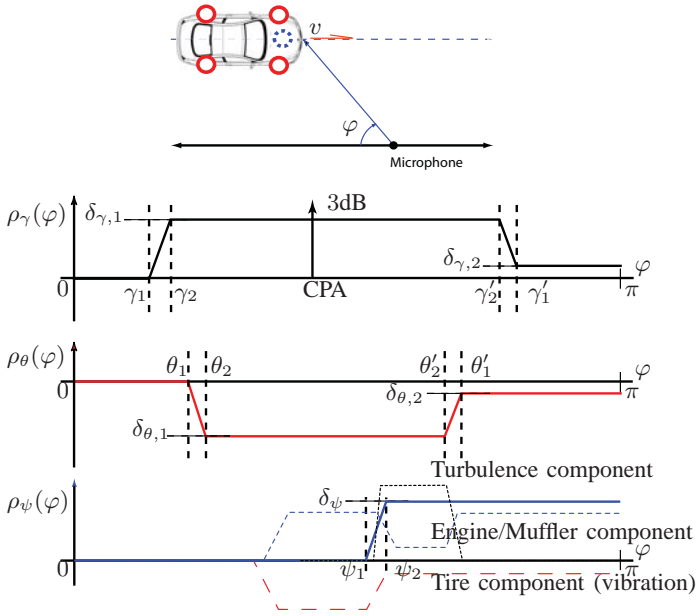
Fig. 4. The microphone bearing reference orientation is defined as the moving direction of the vehicle. Then, a vehicle's spatial acoustic pattern can be approximated by three main components in $\varphi$. The first component $\rho_\gamma$ is due to signal interferences from the front and rear tires. The second component $\rho_\theta$ explains the variation as the microphone comes out of the horn-effect area of the tires. The third component $\rho_\psi$ is an approximate component that accounts for a composite engine/exhaust/tire/turbulence noise effect around the vehicle CPA.
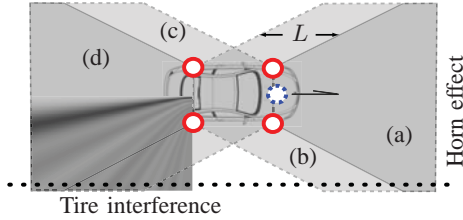


Fig. 5. The components $\rho_\gamma$ and $\rho_\theta$ approximately capture the horn and tire interference effects. Horn effect areas are shown for the front tires (a) and the rear tires (b) in the driving direction; and for the front tires (c) and the rear tires (d) in the opposite direction. The horn effect amplification is the greatest in the overlapping areas. Therefore, the tire noise interference effects are also most prominent in those areas due to the amplification. As the microphone samples the vehicle acoustic pattern (black dots), it leaves and then re-enters the horn effect area. This allows us to estimate the vehicle wheelbase length.

the microphone bearings of $\varphi = 19°-26°$ in Fig. 3(b), where the horn effect amplifications are present. At the microphone bearings less than $\varphi = 19°$, the microphone noise dominates the interference effects to the increase in target distance. For example, at $\varphi = 13°$, the distance of the vehicle to the microphone is almost 1.5 times the distance at $\varphi = 19°$. Hence, the vehicle noise power at $\varphi = 13°$ is less than half the vehicle noise power at $\varphi = 19°$ (see Fig. 3(c)). After $\varphi = 26°$, the horn effect amplifications also begin to diminish in addition to the decreased swing amplitude in the interference.

In this ES component, the tire interference decreases between the bearings $\gamma_1$ and $\gamma_2$, increasing the first component to $\delta_{\gamma,1}$. The angles $\gamma_i$ are related to the width of the car (dipole separation $W$) through the interference term as derived in (12) or approximately via (14). After the drive-by, the tire interference increases between the bearings $\gamma_2'$ and $\gamma_1'$, decreasing the first ES component to $\delta_{\gamma,2}$. The parameter $\delta_{\gamma,2}$ is usually close to zero. We note that the component $\rho_\gamma(\varphi)$ varies in a nonsymmetric fashion with respect to $\varphi$. The asymmetry

is due to the movement of the car: because of the reference frame change, any angle defined in the vehicle reference frame, denoted as $\phi$, is related to the angles in the microphone frame, denoted as $\varphi$, through an aberration relation [33]:

$$\tan\frac{\varphi}{2} = \sqrt{\frac{1 + v/c}{1 - v/c}} \tan\frac{\phi}{2}, \quad (15)$$

where the sign of the speed terms flip after the CPA. Hence, by assuming a symmetric interference pattern for the front and rear tires of the car based on constant car width, one can relate the following angle parameters:

$$\tan\left(\frac{\pi - \kappa'}{2}\right) = \left(\frac{1 - v/c}{1 + v/c}\right)\tan\frac{\kappa}{2}, \quad (16)$$

where $\kappa = \gamma_1$ and $\gamma_2$.

Sometimes, one observes a spike in the acoustic signature in the middle of the interference term very close to the CPA as illustrated in Fig. 3(d)-(f). This spike is caused by the tire interference, which is mostly vibrational in nature (Fig. 3(d)). Hence, it can be approximated by a delta function. In Fig. 4, we plot this effect into the signal interference component $\rho_\gamma(\varphi)$ at the CPA; however, this delta function was not used in our estimation stage even if it is sometimes present in drive-by recordings (Sect. VII).

*The second ES component* $\rho_\theta(\varphi)$ is due to the horn effect, which was explained in Sect. II and illustrated in Fig. 5. In the observed signal envelope, at the microphone bearing $\theta_1$ the horn effect amplification of the farthest front tire from the microphone starts to go down until the bearing $\theta_2$ to $\delta_{\theta,1}$, when the horn effect of the closest rear tire to the microphone also drops. The differential angle $\theta_2 - \theta_1$ is a very good indicator of the vehicle wheelbase length $L$, which can be used to compare the relative sizes of vehicles. The $L$ estimates are not precise because the tire interference also starts to drop between $\varphi = 27°-31°$, which gets combined with the drop of the horn effect. To convert the angle difference into actual size, we use the following approximate relationship using the geometry in Fig. 5

$$\frac{y_m}{\tan\theta_1} - L = \frac{y_m}{\tan\theta_2}. \quad (17)$$

Geometrically, $\theta_2$ is the microphone bearing of the front of the car after the car fully passes the line defined by the bearing $\theta_1$.

After the CPA, the horn effect amplifies the tire noise between bearings $\theta_2'$ and $\theta_1'$, which are related to $\theta_1$ and $\theta_2$ also by the abberation relation (16) (i.e., $\kappa = \theta_i$). The final level of the tire noise component $\delta_{\theta,2}$ is usually different from 0dB, because the rear tire curvature is different from the front tire curvature due to the torque on the tire. Any imbalance of the weight ratio on the front and rear tires also causes $\delta_{\theta,2}$ to deviate from 0dB.

Finally, *the third ES component* $\rho_\psi(\varphi)$ is a composite component that incorporates (i) engine noise, (ii) exhaust system noise, (iii) interference pattern of the tires on the side of the car, and (iv) the noise caused by the air turbulence. This component is hard to associate with any particular effect because many effects are happening simultaneously. A possible scenario is illustrated on $\rho_\psi(\varphi)$ in Fig. 4. To keep the number of ES components manageable and to avoid over-parameterization, we approximate the composite interference pattern as a step function that rises from 0dB to $\delta_\psi$ between bearings $\psi_1$ and $\psi_2$. This approximation works well in our field test cases.

The third ES component explains the variations when the vehicle is closest to the microphone. It should be modified for tracked vehicles since the vibrational tire noise effects or the engine noise effects coming from the side of the tracked vehicles might vary significantly, when compared to wheeled vehicles. When this approximation of the

third ES component becomes poor, $\delta_{\theta,2}$ of the second ES component $\rho_\theta(\varphi)$ compensates. We found that the angle difference $\psi_2 - \psi_1$ is also an indicator of the vehicle length. Hence, (17) is also used to relate the angles in the third interference component to the vehicle wheelbase length $L$.

### B. Vehicle Profile Vector

To jointly determine the speed and the vehicle's spatial acoustic pattern, we use the vehicle profile vector $\boldsymbol{\lambda}$, which is defined as follows:

$$\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_v & \boldsymbol{\lambda}_\varphi & \boldsymbol{\lambda}_\delta & \boldsymbol{\lambda}_f \end{bmatrix}, \text{ where} \tag{18}$$

$$\boldsymbol{\lambda}_v = \begin{bmatrix} \sigma_s^2 & v & W & L \end{bmatrix}, \boldsymbol{\lambda}_\varphi = \begin{bmatrix} \varphi_0 & \gamma_1 & \theta_1 & \psi_1 \end{bmatrix},$$
$$\boldsymbol{\lambda}_\delta = \begin{bmatrix} \delta_{\gamma,1} & \delta_{\gamma,2} & \delta_{\theta,1} & \delta_{\theta,2} & \delta_\psi \end{bmatrix}, \text{ and } \boldsymbol{\lambda}_f = \begin{bmatrix} \chi & p \end{bmatrix}, \tag{19}$$

where $\sigma_s^2 = \frac{S^2 F_s \mathcal{W}}{2\pi\tau}$ with $\mathcal{W}$ assumed known. The vector $\boldsymbol{\lambda}_v$ consists of the physical parameters of the vehicle including loudness $\sigma_s^2$, speed $v$, vehicle tire track length $W$, and the wheelbase length $L$. The vector $\boldsymbol{\lambda}_\varphi$ has the initial vehicle bearing $\varphi_0$ and the angles that define the ES components along with $\boldsymbol{\lambda}_\delta$, which contains the amplitude attenuations and amplifications for the ES components. Lastly, the vector $\boldsymbol{\lambda}_f$ has the RPM $\chi$ and the number of cylinders $p$ of the vehicle. The profile vector $\boldsymbol{\lambda}$ can be viewed as a *fingerprint* of the vehicle and can be used for appearance-based tracking and classification.

### C. Amplitude Observations

In this section, we derive a relationship between the vehicle profile vector $\boldsymbol{\lambda}$ and the square-root of the average signal power, which we denote as the power envelope. This relationship is used to determine the vehicle profile vector using standard maximum likelihood estimation techniques. In the derivation, we assume that there are no multipath effects. When there are multipath effects, a parametric form of the source signal must be known. Without these assumptions, range estimation using a single microphone is not possible [34].

The power envelope function $\mathcal{E}$ is calculated using $\tau$-discrete samples of the microphone output $z[n]$ as follows:

$$\mathcal{E}[n_\tau] = \mathcal{E}(t)\Big|_{t=\frac{n_\tau\tau}{F_s}} = \sqrt{P_z[n_\tau]} = \sqrt{\frac{1}{\tau}\sum_{k=0}^{\tau-1} z^2[n_\tau\tau + k]}, \tag{20}$$

where subscript $\tau$ under the sample index $n$ implies that the samples of the power envelope function are calculated at every $\tau/F_s$ second. Given the drive-by acoustic recording of the target, the microphone output signal is divided into $N_\tau$-segments, each of which has $\tau$ samples. Then, each of these segments are used to calculate one $\mathcal{E}[n_\tau]$ using (20). During the calculation of each $\mathcal{E}[n_\tau]$, we assume local stationarity so that the complete motion of the target is parameterized as follows:

$$\beta[n_\tau] = 1 + \frac{v}{c}\cos\varphi[n_\tau], \; r[0] = y_m \csc\varphi_0$$
$$r[n_\tau] = \sqrt{(v\tau/F_s)^2 + r^2[n_\tau - 1] - 2(v\tau/F_s)r[n_\tau - 1]\cos\varphi[n_\tau - 1]},$$
$$\varphi[n_\tau] = \varphi[n_\tau - 1] + \sin^{-1}\left(\frac{v\tau}{F_s r[n_\tau]}\sin\varphi[n_\tau - 1]\right). \tag{21}$$

The choice of the parameter $\tau$ was discussed in Sect. III. Because of our particular choice of $\tau$, the DFT coefficients used to calculate the power function at $\tau$-samples apart are statistically uncorrelated, and hence, each sample of $\mathcal{E}[n_\tau]$ ($n_\tau = 0, 1, \ldots, N_\tau - 1$) is also statistically uncorrelated of the others. In the remaining parts of this section, we drop the bracket notation here for brevity as we emphasize
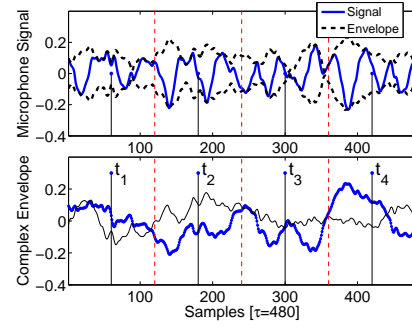


Fig. 6. (*Top*) A short segment of a Ford F150 truck drive-by noise is shown where $F_s = 48$kHz. A total of $\tau = 480$ samples are divided into $K = 4$ segments, where each segment indexed by $t_i$ ($i = 1, \ldots, 4$) has 120 samples. As illustrated by the top plot, the observed signal at each segment looks like a noisy sinusoid, whose dominant frequency is approximately 1000Hz. (*Bottom*) Real (solid) and imaginary (circles) parts of the complex envelope are illustrated.

here that we are determining a pdf for $\mathcal{E}[n_\tau]$, where the time indexes belong to a particular segment of the acoustic recording.

Note that the vehicle signals contain strong harmonics due to the engine noise and a strong multi coincidence peak at 1000Hz due to the tire noise [14]. Hence, the observed signal at the microphone appears as a noisy periodic signal over short time intervals. Define the complex envelope of the observed target signal at the microphone plus noise as $y(t)$, sampled at times $t = (t_1, \ldots, t_K)$. The absolute value of each complex envelope sample represents approximately the RMS power of a signal segment that is $\tau/K$ in length. Then, we have the following narrow-band observation model, which has been extensively used in the literature (e.g., for bearing estimation [6], [35]–[38]):

$$y(t) = \varepsilon(t)e^{-j\frac{2\pi f_0 R}{\beta c}} + u(t), \tag{22}$$

where $f_0$ is the dominant instantaneous source frequency at time $t$, $u(t)$ is an *i.i.d.*, zero mean, complex circularly symmetric Gaussian random variable $\mathcal{CN}\left(0, \sigma_u^2\right)$ with variance $\sigma_u^2$, and the source signal complex envelope samples $\varepsilon(t)$ are also assumed *i.i.d.* with the following form:

$$\varepsilon(t) \sim \mathcal{CN}\left(0, \mathcal{A}_{\boldsymbol{\lambda}}(t)\right), \quad \mathcal{A}_{\boldsymbol{\lambda}}(t) = \frac{\sigma_s^2 e^{2m_\tau}}{\beta(t)r^2(t)}\prod_{i=\gamma,\theta,\psi} 10^{\rho_i(\varphi(t))/10}. \tag{23}$$

In order for $\varepsilon(t_i)$'s to be uncorrelated for moderate or small $K$ values (e.g., $K \geq 3$), $\tau$ (and $F_s$) must be high enough to include more than a few cycles of the noisy signal. In (23), $\mathcal{A}_{\boldsymbol{\lambda}}(t)$ is the source signal power, which has directional variation, and $e^{m_\tau}$ is an *i.i.d.* multiplicative noise on the signal amplitude ($m \sim \mathcal{N}\left(0, \sigma_m^2\right)$) to account for (i) any additional complex tire interference effects that appear multiplicative in the received signal power in (11) (e.g., Fig. 3), (ii) air turbulence noise, and (iii) the directional filtering of the engine and exhaust noises by the vehicle body. The power envelope $\mathcal{A}_{\boldsymbol{\lambda}}$ is written as a function of the full vehicle profile vector $\boldsymbol{\lambda}$ for brevity even though it does not depend on $\boldsymbol{\lambda}_f$ of the vehicle profile vector. This choice eliminates the cumbersome notation of writing the envelope specifically as a function of $\boldsymbol{\lambda}_v$, $\boldsymbol{\lambda}_\varphi$, and $\boldsymbol{\lambda}_\delta$. Figure 6 illustrates the elements of the observation model (22) for a truck signal, where $\tau = 480$ samples.

Based on our signal and noise assumptions, it is straightforward to prove that the sensor output complex envelope $y(t)$ also has an *i.i.d.* zero mean circularly complex Gaussian distribution with variance $\sigma_y^2 = \mathcal{A}_{\boldsymbol{\lambda}}(t) + \sigma_u^2$. Due to the local stationarity assumption of the target motion during each $\tau$-length data segment, $\mathcal{A}_{\boldsymbol{\lambda}}(t_i) = \mathcal{A}_{\boldsymbol{\lambda}}$ for
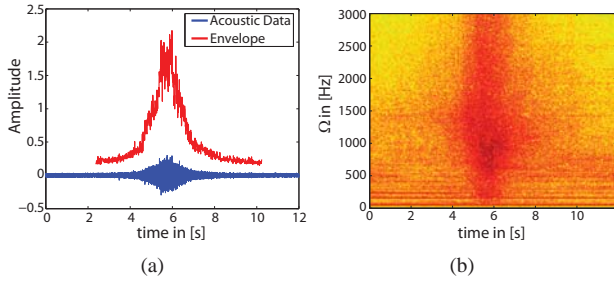
Fig. 7. Drive-by test by a 6-cylinder Chevy Impala moving with 18.7m/s at an approximate distance of 5.8m. (a). The acoustic signal was sampled at $F_s = 48$kHz. The power envelope $\mathcal{E}$ is calculated with $\tau = 480$. In the figure, $\sqrt{\tau}\mathcal{E}$ is plotted to emphasize the variation. There is an asymmetry in the envelope estimates that can be explained by the ES components. (b). The spectral content of the acoustic signal. Note the strong interference at 60Hz. The tire air-noise spectrum, which is concentrated around $700 - 1300$Hz, does not exhibit the frequency modulation pattern illustrated in Fig.1.

$i = 1, \ldots, K$. Then, we write the power envelope as follows:

$$\mathcal{E} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}|y(t_i)|^2} = \frac{\sigma_y}{\sqrt{2K}}\sqrt{\sum_{i=1}^{K}\left(\frac{y_{\text{real}}^2(t_i)}{\sigma_y^2/2} + \frac{y_{\text{imag}}^2(t_i)}{\sigma_y^2/2}\right)} = \frac{\sigma_y}{\sqrt{2K}}\nu, \tag{24}$$

where $\nu$ has a Chi distribution $p_\nu(\nu)$ with $2K$ degrees-of-freedom [39]. It is well-known that Chi distribution reaches normality rather quickly [40]–[42]. In [42], we provide normal approximations of the Chi distribution based on the Laplacian method [43] and moment matching. These approximations are highly accurate even for small sample sizes.

A normal approximation based on moment matching of $\nu$ results in the following expression for the pdf of $\nu$ [42]:

$$p_\nu(\nu) = \mathcal{N}\left(\sqrt{2K - 1/2}, \frac{1}{2}\right). \tag{25}$$

Then, using (24), an approximate pdf for $\mathcal{E}$ is given by

$$\mathcal{E} \sim \mathcal{N}\left(\sqrt{\frac{2K - 1/2}{2K}}\sigma_y, \frac{\sigma_y^2}{4K}\right). \tag{26}$$

Note that when $2K \gg 1$, by using (26) and $e^x \approx 1 + x$ for $x \ll 1$, we can write

$$\mathcal{E} \sim \sigma_y + \frac{\sigma_y}{\sqrt{4K}}\mathcal{N}(0,1) \approx \sigma_y e^{\frac{\mathcal{N}(0,1)}{\sqrt{4K}}}. \tag{27}$$

At this point, further simplification is possible. When the target *SNR* is also high, we typically have $\mathcal{A}_{\boldsymbol{\lambda}} \gg \sigma_u^2$ (see Sect. VI). Then, (27) becomes (denoted as $\mathcal{E} \sim p_1(\mathcal{E})$):

$$\mathcal{E} \sim \mathcal{A}_{\boldsymbol{\lambda}}^{1/2}e^{\frac{\mathcal{N}(0,1)}{\sqrt{4K}}} = \frac{\sigma_s e^{\left\{m_\tau + \frac{\mathcal{N}(0,1)}{\sqrt{4K}}\right\}}}{\sqrt{\beta}r}\prod_{i=\gamma,\theta,\psi}10^{\rho_i(\varphi)/20}. \tag{28}$$

Aside from the ES components, the approximation (28) is somewhat different than the commonly used additive noise model in the literature [44] (denoted as $\mathcal{E} \sim p_2(\mathcal{E})$):

$$\mathcal{E} \sim \frac{\sigma_s}{\sqrt{\beta}r} + \sigma_w\mathcal{N}(0,1), \tag{29}$$

where $\sigma_w$ is assumed independent of $\sigma_s$. In this paper, we use $p_1(\mathcal{E})$ in (28) for estimation. We compare the range estimation performance of both models in [42] when the data is generated by (22).

### D. Frequency Observations

The spectral content of a vehicle exhibits directional variation, making it difficult to use the frequency modulation effects of the

vehicle motion to determine speed. We emphasize that this directional variation is not due to the motion of the vehicle but it is due to tire noise effects, which are stochastic in nature as discussed in Sect. II. The useable frequency tracks for speed estimation are generated by the engine because the frequency modulation effects can be observed in the deterministic component of the engine noise. These deterministic engine frequencies span the $0 - 250$Hz range at nominal RPM's (e.g., Fig. 7(b)). At moderate vehicle speeds ($30 - 50$mph), the full Doppler shift swings $F_0$ approximately by $6\%$. This change in $F_0$ can also be achieved by an RPM change, e.g., $\Delta\chi \approx 150@(\chi = 2500)$ results in a 6% swing of $F_0$. Hence, if a driver changes the car's RPM by 50 during the vehicle drive-by, there will be a 33% error in the expected Doppler frequency change in $F_0$ based on the constant frequency source assumption $\mathcal{C}$. We emphasize that this RPM change is unnoticeable on the dashboard of the vehicle and is likely to happen. On the other hand, the effect of the same RPM change on the total car loudness is negligible.

Therefore, determining the vehicle speed by fitting a Doppler shift function to the engine and tire frequency tracks is an unreliable approach. For example, in [2], the speed estimation was performed using an autoregressive modeling of the acoustic signals under a point source assumption. It was concluded that the Doppler-based speed estimation on the source frequencies does not perform well with field-data [2]. It was also concluded that, with the same source assumptions, the envelope measurements yield improved speed estimates than the frequency measurements; however, the speed estimates are nonetheless biased. Later on, we comment on the possible reasons of this bias mentioned in [2].

On the other hand, the spectral harmonic content can be used to determine $\boldsymbol{\lambda}_f$ of the vehicle profile vector since the ratio of the engine harmonic frequencies is independent of the vehicle RPM. Moreover, conditioned on $\boldsymbol{\lambda}_v$ estimate, it possible to further refine $\boldsymbol{\lambda}_f$ by compensating for the Doppler shifts if we are simultaneously tracking the frequencies. The number of cylinders $p$ is usually the most elusive to estimate because the body of a vehicle may also act as a filter to directionally suppress the frequency at the engine fire rate $F_0$ (Fig. 8). Hence, it is rather easy to incorrectly estimate the number of cylinders of a vehicle because the strongest frequency is not necessarily $F_0$. If a characterization can be done[4], which is applicable to the vehicles of interest, the number of cylinders can also be estimated robustly. Estimation of $\chi$ can be done accurately using harmonic analysis methods [45]. In our estimation, we use the power spectral density of the acoustic signal to determine $\boldsymbol{\lambda}_f$. Further details can be found in [45].

## VI. PRACTICAL ASPECTS OF ESTIMATION

### A. Detector

With a gross simplification of the problem, we assume that we have an estimate of the variance $\sigma_u^2$, which is the noise variance of an additive Gaussian noise on the microphone acoustic signal. A much simpler detector can be formulated with this assumption, and consequently, an approximate data-likelihood formula can be used for estimating the vehicle profile vector as shown in the next subsection. We write our hypotheses as follows ($n = 0, \ldots, \tau - 1$):

$$H_0 : y[n] = u[n], \quad H_1 : y[n] = u[n] + z[n], \tag{30}$$

where $u \sim \mathcal{N}(0, \sigma_u^2)$ and $z[n]$ is an unknown waveform that characterizes any non-noise like quantity. Given these simplifications,

---

[4]This also includes a characterization of the microphone itself. If the microphone has an uneven frequency response, it must be accounted for while estimating $p$.
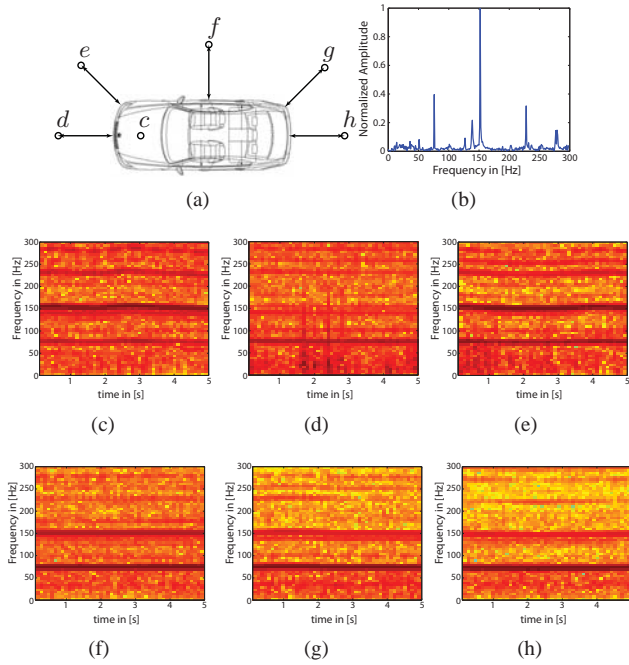
Fig. 8. (a) Our stationary BMW 325i sound measurement set up is illustrated. In the measurements, all microphones, except the engine microphone, were situated 1m away from the vehicle. The engine measurements were taken with a microphone approximately 25cm away from the center of the engine without the car hood. During the test, the engine RPM kept at approximately 3000, then measurements were taken with the same microphone, at the locations shown in the picture. (b) The strongest peak in the Fourier transform of the engine microphone data between $t = 1$s and $t = 2$s ($F_s = 22050$Hz) is 152Hz corresponding to $F_0$. Harmonic components are approximately 25.3Hz apart because the vehicle has 6 cylinders, however their amplitudes vary greatly with no apparent regularity. There is an additional peak around 139Hz, which is likely to be caused by the engine management. (c) Time-frequency plot of the sound collected at the engine microphone. (d) Time-frequency plot of the sound collected at the microphone situated at position $d$ in (a). Note that the strongest frequency is now 75Hz, corresponding to a 3-cylinder vehicle. (e-h) The estimated engine fire rate may be confused as the direction varies, leading to incorrect cylinder estimates.

the generalized likelihood ratio test is a square-law detector [6]:

$$\sum_{n=0}^{\tau-1} y^2[n] \underset{H_0}{\overset{H_1}{\gtrless}} \Upsilon, \qquad (31)$$

for some threshold $\Upsilon$. This threshold can be determined by using a false alarm probability $P_F$:

$$P_F = \Pr\left(\chi_\tau^2 > \frac{\Upsilon}{\sigma_u^2}\right) = \Pr\left(\mathcal{E} > \left(\frac{\Upsilon}{\tau}\right)^{1/2}\right), \qquad (32)$$

where $\chi_\tau^2$ is the Chi-squared distribution with $\tau$ degrees-of-freedom. The vehicle CPA can be approximately determined by windowing the short-time power estimates and finding the time of the maximum received power. This threshold can then be used to segment the desired vehicle signal [3], [9], [46].

### B. Approximate Cost Function and Its Solution

In our detector, if we threshold the amplitude measurements $\mathcal{E}$ by a positive constant $\varpi\sigma_u$, then the false alarm probability is given by $P_F = \Pr\left(\chi_\tau^2 > \varpi^2\tau\right)$. When $\varpi^2$ is chosen much greater than 1,

e.g., $\varpi^2 = 3$, the false alarm probability becomes quite small.[5] This assumption is also satisfied for acoustic vehicle detectors discussed in [9], [46]. In this case, the maximum likelihood (ML) estimator can be found by minimizing the following cost function, which does not depend on the multiplicative noise variance:

$$\boldsymbol{\lambda}_{\text{ML}} = \arg\min_{\boldsymbol{\lambda}} J(\boldsymbol{\lambda}), \quad J(\boldsymbol{\lambda}) = \sum_{n_\tau=0}^{N_\tau-1} \left(\log \mathcal{E}[n_\tau] - \frac{1}{2}\log \mathcal{A}_{\boldsymbol{\lambda}}[n_\tau]\right)^2. \tag{33}$$

To find the minimum of $J$ in (33), we can use iterative optimization algorithms to avoid calculating the entire ML surface for numerical efficiency.

**Constraint Set and Solution**: Most optimization algorithms use a cost function, its first and second order derivatives, and a *constraint set*. In our problem, the constraint set is crucial because it prevents the confusion among the individual interference components and provides robustness to the numerical solution. The constraints that we use in our solution are shown in Table I. We use the MATLAB's constrained optimization toolbox (*fmincon*) for the solution[6], which is summarized in Table II.

The constraints of the speed estimation problem are tied to the nature of the vehicles. The specific angle of the horn effect depends on the vehicle tire profile. According to [14], [16], [18], $25° < \theta_1 < 45°$ is typical. The amplification factor is also empirically between .5-4dB [14], [16], [18]. The dynamic range of the tire interference component is difficult to predict due to the horn effect; however it is typically less than 4dB and we constrain $\delta_{\gamma,1}$ to be greater than some small number, i.e., $0.1$dB. We assume that the tire interference effects are more or less symmetric, hence $|\delta_{\gamma,2}| < 0.1$dB is used to enforce this assumption. The third ES component is used to model the vehicle effects when it is closest to the microphone after the horn effects, hence we constrain $\theta_1 < \psi_1 < 120°$.

Finally, even with the recursive solution, the solution of the ES components is computationally demanding since it searches in a high dimensional state space. In this paper, we focus on demonstrating the importance and the use of the ES components and are not concerned with the computational issues. However, we note that the final solution is usually not sensitive to the initialization. Moreover, we note that in [47], Searle explains an efficient solution of a similar computational exercise, where bank of template signals are used to determine the vehicle speed. A similar approach may be taken in our problem, since the ES components have well-defined characteristics.

### TABLE I
### THE CONSTRAINT SET

| Name | Constraints |
|---|---|
| Bearings $\gamma, \theta, \psi$ | $\gamma_1 < \theta_1 < \psi_1, 10° < \gamma_1 < 20°, \psi_1 < 120°$ |
| $\delta_\gamma$ | $0.1$dB $< \delta_{\gamma,1} < 4$dB, $|\delta_{\gamma,2}| < 0.1$dB |
| $\delta_\theta$ | $-4$dB $< \delta_{\theta,1} < -0.5$dB, $|\delta_{\theta,2}| < 1$dB |
| $\delta_\psi$ | $0.1$dB $< \delta_\psi < 2$dB |

### C. Microphone Directionality and Frequency Response

Our derivations assumed an omnidirectional microphone with a flat spectral response. If this is not the case, the equations for the ES components can be modified to incorporate the spatio-spectral

---

[5]Also, note that to be able to calculate all the ES components using (28), the target *SNR* must be high enough.

[6]For the computational complexity of *fmincon* algorithm, please refer to MATLAB documentation.

**TABLE II**

VEHICLE PROFILING USING THE ES COMPONENTS

Given $\boldsymbol{\lambda}_v, \boldsymbol{\lambda}_\varphi$, and $\boldsymbol{\lambda}_\delta$, calculate the cost function $J(\boldsymbol{\lambda})$ in the following order:

1. Determine $\gamma_2$ by using $\gamma_1$, $W$, $\mathcal{W}$, and (14). A typical value $\mathcal{W}$ is $2\pi \times 2400$ for compact vehicles.
2. Determine $\theta_2$ by using $\theta_1$, $L$ and (17).
3. Determine $\psi_2$ by using $\psi_1$, $L$ and (17) (replace $\theta$ with $\psi$ in the equation).
4. Determine the primed bearings $\gamma_i'$ and $\theta_i'$ by using $v$ and the abberation relationship (16).
5. Use (23) and (21) and the ES components to calculate $\mathcal{A}_{\boldsymbol{\lambda}}$. Then, use (33), where $\mathcal{E}$ is calculated from the data.

---

microphone response:

$$\log \mathcal{E}[n_\tau] \sim \mathcal{N}\left(\frac{1}{2}\log\sum_k \mathcal{A}_{\boldsymbol{\lambda}}[n_\tau, k], \sigma_{\mathcal{E}}^2\right),$$

$$\mathcal{A}_{\boldsymbol{\lambda}}[n_\tau, k] = |H_{\text{mic}}[k, \varphi[n_\tau]]|^2 \times \frac{\sigma_s^2[k]}{\beta[n_\tau]r^2[n_\tau]} \prod_{i=\gamma,\theta,\psi} 10^{\rho_i(\varphi[n_\tau])/10},$$
(34)

where $k$ is the frequency index used in the $\tau$-sample DFT coefficients, and $|H_{\text{mic}}[k, \varphi[n_\tau]]|^2$ is the microphone response function at the frequency bin $k$ and the direction $\varphi$. In this formulation, the microphone orientation must be supplied or calibrated. It cannot be treated as an unknown. Moreover, due to the directionality, we can now determine the absolute travel direction of the vehicles.

## VII. EXPERIMENTS

To demonstrate the ideas, data was collected with $F_s = 48\text{KHz}$ at a two-way street with an omnidirectional microphone, emplaced 1.5m off the ground on a pole at the sidewalk. The distance of the bottom of the microphone pole to the center of the street is 7.4m. A video camera is used to establish the ground truth and identify the vehicles in the test [48].

### A. Vehicle Profiling

Table III and Figs. 9-11 show the results of the vehicle speed estimates obtained by three different methods using $\tau = 480$ samples:
**M1**) This method uses full vehicle profile vector with (28). The cost function is given by (33), where the target motion is defined by (21) and $\mathcal{A}_{\boldsymbol{\lambda}}[n_\tau]$ is defined by (23).
**M2**) This method uses only $\sigma_s^2$ and $v$ with (28) where the product of the ES components is set to 1 in (28). Hence, this method does not take into account the directional nature of the vehicle noise pattern and hence does not estimate the other parameters of the vehicle profile vector. The cost function is

$$J_{\text{M2}}(\sigma_s^2, v) = \sum_{n_\tau=0}^{N_\tau-1}\left(\log\mathcal{E}[n_\tau] - \log\left\{\frac{\sigma_s}{\sqrt{\beta[n_\tau]}r[n_\tau]}\right\}\right)^2,$$
(35)

where the target motion is defined by (21).
**M3**) This method uses only $\sigma_s^2$ and $v$ with the constant velocity motion model (21). This method is the same as the one used in [2] to estimate the speed of point sources. Hence, this method does not take into account the directional nature of the vehicle noise pattern as M2 and also does not estimate the other parameters of the vehicle profile vector. In addition, it does not use a multiplicative noise model

as discussed in Sect. V-C. The cost function is

$$J_{\text{M3}}(\sigma_s^2, v) = \sum_{n_\tau=0}^{N_\tau-1}\left(\mathcal{E}[n_\tau] - \frac{\sigma_s}{\sqrt{\beta[n_\tau]}r[n_\tau]}\right)^2.$$
(36)

To initialize M1, we use the results of M2 to initialize $\sigma_s^2$ and $v$. We then use $L = 3$m and $W = 1.5$m as our initial dimension estimates. Also, we use $\gamma_1 = 17°$, $\theta_1 = 27°$, and $\psi_1 = 109°$, and $\boldsymbol{\lambda}_\delta = \mathbf{0}$.

It is seen that the estimates of [2] are improved by incorporating the multiplicative noise model on the signal envelope. However, estimation using the ES components yields the best estimates. In Figs. 9-10(c), we also plot the log-likelihood surface variation with respect to the parameters $W$ and $L$, where we treat the rest of the vehicle profile vector parameters as known by substituting their ML estimates. The Hessian inverse of the second order Taylor series expansion of the cost function (33) with respect $(W, L)$ around the mode of the log-likelihood surface is used to estimate the variance of $(W, L)$, which are displayed in Table III and illustrated in Fig. 9(c) and Fig. 10(c). In the figures, the maximum of the log-likelihood surfaces are set to 0 by addition of appropriate constants.

We also estimated the number of cylinders and the RPM of the vehicles by using the methods in [45]. The number of cylinders $p$ are estimated by compensating for the microphone spectral characteristics. However, there was no compensation for any vehicle directional variation.
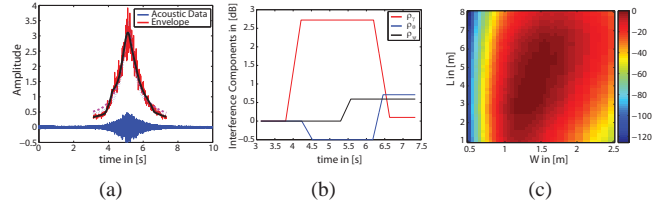


(a)　　　　(b)　　　　(c)

Fig. 9. **Ford F150**. (a) Estimated envelope by the interference components is shown with the solid line (M1). The dotted and the dashed lines belong to M2 and M3 noise models, respectively. (b) Estimated ES components are shown. According to the ES components, the vehicle is louder in the rear than front. (c) The log-likelihood surface of the vehicle dimensions.
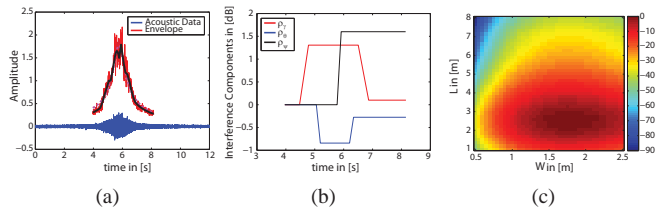


(a)　　　　(b)　　　　(c)

Fig. 10. **Chevy Impala**. (a) Observed envelope exhibits significant variations. The interference components (solid line) adequately explains the variations (also see Fig. 12). (b) Estimated ES components are shown. Parameter $\delta_\psi$ is relatively larger than the other components indicating significant air noise. (c) The log-likelihood surface of the vehicle dimensions.

### B. Noise Models

In Appendix I.A, we show that when an additive noise model is used to solve for a multiplicative noise model using the assumptions of our speed estimation problem, the speed estimates always have an expected negative bias. This is intuitive: because the power envelope samples around the CPA of the target has larger variance than the tails of the power envelope, a minimum cost solution would result in a wider envelope fit by minimizing the error around the higher variance region. A wider envelope fit implies slower speed estimates. This type of negative biases estimates was also noted by [2] as they

| | Ground Truth | | | | M1 | | | | | | M2 | | M3 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Vehicle* | $y_m$ | $v_{camera}$ | $L$ | $W$ | $v$ | $\sigma_s^2$ | $L(\mu \pm \sigma)$ | $W^{\ddagger}(\mu \pm \sigma)$ | $\chi$ | $p^{\natural}$ | $v$ | $\sigma_s^2$ | $v$ | $\sigma_s^2$ |
| Ford F150 | 6.3 | 17.54m/s | 3.20m | 1.70m | 17.86m/s | 12.60 | 5.38±.55m | 1.30±.09m | 3038 | 8 | 28.00m/s | 24.08 | 21.39m/s | 21.27 |
| Chevy Impala | 5.8 | 18.68m/s | 2.80m | 1.58m | 18.60m/s | 9.23 | 2.58±.31m | 1.75±.17m | 3300 | 6 | 18.29m/s | 11.55 | 15.05m/s | 10.90 |
| Honda Accord | 4.3 | 16.74m/s | 2.71m | 1.55m | 14.44m/s | 6.86 | 3.28±.29m | 1.40±.17m | 3074 | 6 | 17.34m/s | 10.17 | 14.49m/s | 9.67 |
| Nissan Maxima* | 4.6 | 13.32m/s | 2.70m | 1.53m | 13.20m/s | 12.45 | 3.28±.38m | 1.50±.18m | 3825 | 6' | 14.23m/s | 14.86 | 14.27m/s | 14.49 |
| Nissan Maxima* | 4.1 | 4.14m/s | 2.70m | 1.53m | 4.49m/s | 6.34 | 2.58±.19m | 1.50±.09m | 3150 | 4 | 4.75m/s | 9.90 | 3.46m/s | 9.20 |
| Isuzu Rodeo | 8.1 | 13.44m/s | 2.70m | 1.51m | 13.89m/s | 7.87 | 5.20±.53m | 1.35±.25m | 3450 | 6 | 11.32m/s | 7.97 | 11.79m/s | 7.95 |
| Mercedes E | 8.1 | 13.94m/s | 2.83m | 1.54m | 13.80m/s | 7.68 | 2.93±.96m | 1.50±.41m | 3075 | 6 | 15.51m/s | 10.47 | 11.78m/s | 9.93 |
| Volvo 850 SW | 8.1 | 14.11m/s | 2.66m | 1.51m | 14.69m/s | 9.60 | 3.10±.27m | 1.40±.27m | 2250 | 10' | 12.93m/s | 9.01 | 11.22m/s | 8.63 |
| Nissan Frontier | 4.3 | 17.56m/s | 3.20m | 1.56m | 17.84m/s | 9.31 | 4.85±.63m | 1.40±.25m | 2625 | 6 | 17.02m/s | 9.92 | 17.56m/s | 9.74 |
| VW Passat | 5.1 | 11.66m/s | 2.70m | 1.50m | 11.58m/s | 6.06 | 2.75±.66m | 1.80±.26m | 1950 | 6 | 8.58m/s | 6.06 | 8.66m/s | 6.11 |
| Error STD | | | | | 0.8246m/s | | | | | | 3.7203m/s | | 2.2627m/s | |
| Error STD¶ | | | | | 0.2777m/s | | | | | | 1.5154m/s | | 1.5126m/s | |
| Bias | | | | | -0.0735m/s | | | | | | 0.6845m/s | | -1.1458m/s | |
| Bias¶ | | | | | 0.1737m/s | | | | | | -0.4017m/s | | -1.7013m/s | |

‡ A fixed bandwidth of $\mathcal{W} = 2\pi \times 2400$ ($\Omega_2 = 2\pi \times 1200$) is used to determine the car widths using (14). Hence, the width estimates of the F150 and Nissan Frontier are biased because they have a significantly different tire profile than the sedan vehicles. When $\mathcal{W} = 2\pi \times 3200$ is used, the width estimates of F150 and Nissan Frontier become 1.60m and 1.80m, respectively. In turn, their wheelbase length estimates also change to 4.20m and 3.98m.

♮ Estimated by finding the frequency $F_0$ with the maximum power spectral density between frequencies $85$-$210$Hz and then dividing $F_0$ by the CFR $f_0$ estimate [45].

' Incorrectly estimated. The actual values are 4 (Maxima) and 5 (Volvo).

* Same vehicle.

¶ Leave-one-out estimates with the minimum variance.



(a) Honda Accord    (b) Nissan Maxima

(c) Nissan Maxima    (d) Isuzu Rodeo

(e) Mercedes Benz E-Class    (f) Volvo

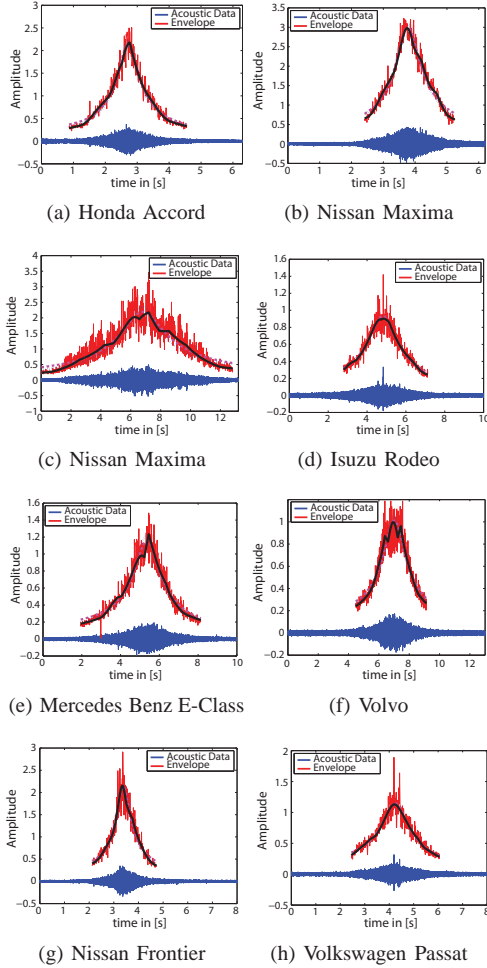(g) Nissan Frontier    (h) Volkswagen Passat

Fig. 11. Estimated envelopes by the different models. The solid, the dotted, and the dashed lines belong to M1, M2, and M3 noise models, respectively.

use the additive model. From Table III, we see that the estimation results of M3 using only $\sigma_s^2$ and $v$ with the additive Gaussian model as in [2] on the power-envelope observations exhibit the predicted expected bias on the average. This bias decreases when we change the additive noise model (M3) to the multiplicative noise model (M2). Hence, our results suggests that a multiplicative model on the envelope measurements might be a better model for estimation.

Figure 12 suggests another reason for the bias. Note that the speed estimates for vehicles 1, 4, and 9 in Table III are actually higher than the true speed values for M3. This is caused by the directional nature of the vehicle noise. We can think of the ES components as a time-varying mean shift acting on the logarithm of the envelope estimates using (20) (Fig. 12(b)). In this case, when the speed estimates are calculated without the ES components using the logarithm of the envelope (M2), then the speed estimates are also expected to have a bias (see Appendix I.B). When we estimate the speed using the ES components along with the multiplicative noise model, we further decrease the bias and also decrease the variance of the estimates.

Figures 12(c) and (f) show that the noise residual[7] of the logarithm of the envelope estimates has indeed a flat spectrum as assumed by our data model (20). Finally, other reasons for the possible bias are the ignorance of inherent estimator biases, multipath propagation, vertical dimension of the vehicles, and atmospheric turbulence [5], [7], [49].

### C. Vehicle Classification Results

The vehicle profile vector provides a natural feature vector for classifying vehicles. Figures 14(a) and (b) show that the vehicles can be separated into two classes based on their length and size. Note that the estimated vehicle lengths are not exact vehicle lengths; however, they can separate compact cars from SUV's or trucks. However, Fig. 13 illustrates a case where the vehicle profile ML estimate for $(L, W)$ is $(5.20, 1.35)$ m, plus a second mode around $(L, W) = (2.93 \pm .38, 1.5 \pm .23)$m, which is only fractionally lower in the dynamic scale of the log-likelihood function. The true values for Isuzu Rodeo are $(L, W) = (2.70, 1.51)$m. Hence, this vehicle is misclassified. Figure 14(c) also illustrates that it is possible to identify loud vehicles such as vehicles with mechanical problems or heavily loaded SUV's or pick-up trucks, which are expected to be louder than usual. This classification is based on the fact that the loudness of the vehicle has a certain functional distribution as indicated in [15], [16]. Hence, given two similar vehicles, it may be possible to identify if one of them is heavily loaded or has mechanical problems even if they move at different speeds.

[7] The noise residuals for each methods are $m_\tau = \log \mathcal{E}[n_\tau] - \frac{1}{2} \log \mathcal{A}_{\boldsymbol{\lambda}}[n_\tau]$ (M1), $m_\tau = \log \mathcal{E}[n_\tau] - \log \left\{ \frac{\sigma_s}{\sqrt{\beta[n_\tau] r[n_\tau]}} \right\}$ (M2), and $u_\tau = \mathcal{E}[n_\tau] - \frac{\sigma_s}{\sqrt{\beta[n_\tau] r[n_\tau]}}$ (M3).
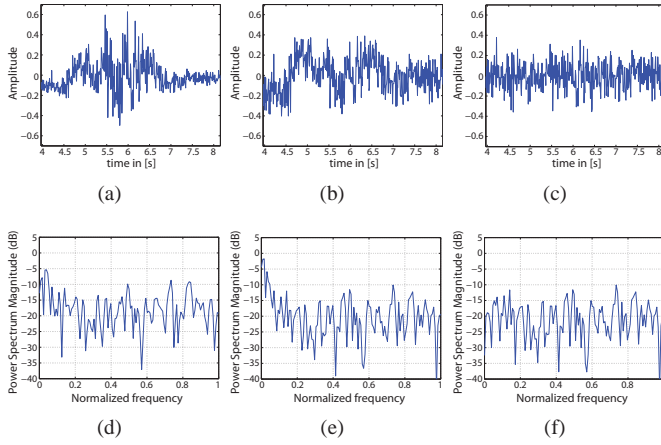
Fig. 12. Chevy Impala noise residuals from (a) the envelope estimates when the additive Gaussian model is used [2] (dashed line in Fig. 10), (b) the logarithm of the envelope estimates when the multiplicative noise model introduced here is used without the ES components (dotted line in Fig. 10), and (c) the logarithm of the envelope estimates when the multiplicative noise model is used along with the ES components (solid line in Fig. 10). (d-f) The noise spectra are plotted with respect to the normalized frequency. The spectra in (d-f) correspond to the noise residuals in (a-c), respectively. The noise residual in (c) is white as demonstrated by the flat spectrum in (f). Note that the interference components remove the time-varying mean from the noise residual of the multiplicative noise model in (b) while at the same time extracting useful information such as vehicle dimensions.
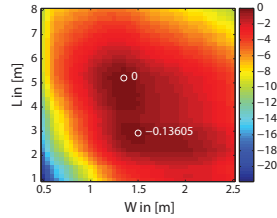


Fig. 13. Log-likelihood surface for the vehicle dimensions for Isuzu Rodeo. Note the bi-modality of the likelihood surface.
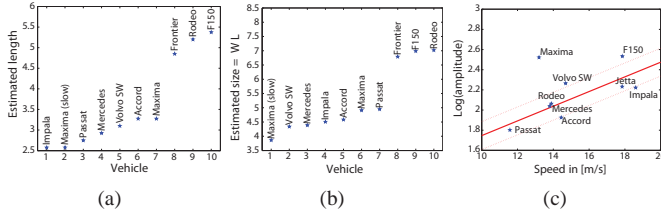


Fig. 14. (a) Estimated vehicle lengths are compared. There is a clear separation between compact cars and large vehicles. (b) Estimated vehicle sizes are compared. (c) Logarithm of the vehicle signal amplitudes are plotted with respect to their speed. There is a linear trend in the plot as also indicated by [15], [16]. The solid line represents a least squares fit to the data without the Nissan Maxima. The dotted lines are one standard deviation away from the mean. The Nissan Maxima is louder than the other cars because the vehicle has mechanical problems.

*D. Scaling Ambiguity*

We also investigated the sensitivity of the vehicle profile vector estimates to the scaling ambiguity issue. Figure 15 illustrates the estimation results for different $y_m$ values for the Chevy Impala and Honda Accord. The true values of the parameter $y_m$ are shown in Table III. The speed estimates are proportional to the assumed $y_m$ (Fig. 15(a) and (d)), because the vehicle speed is entangled with $y_m$ as discussed in Sect. III. The width estimates are somewhat insensitive to the assumed $y_m$ (Fig. 15(b) and (e)) because they only depend on

the asymptotic interference angles, which do not depend on $y_m$. The length estimates increase as $y_m$ is increased; however, they are not proportional to $y_m$. The cost function shows a flat response around the true $y_m$ value for the Chevy Impala whereas the cost function is at a local optimum for the Honda Accord around the true value of $y_m$.
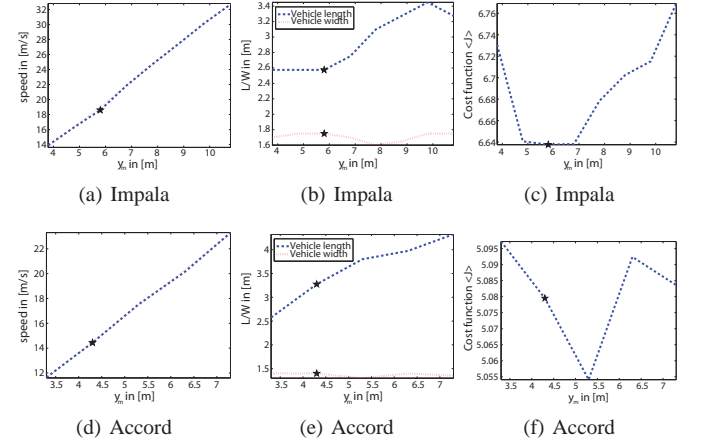


Fig. 15. (a) Chevy Impala speed estimates are proportional to the assumed $y_m$ values. Dashed line corresponds to the vehicle profile vector estimates whereas the dotted line is a linear fit that shows that the speed estimates are approximately proportional to $y_m$. The estimate at the true $y_m$ is marked with a star. (b) Vehicle width and length estimates are shown. The length estimates increase with $y_m$ whereas the width estimates are somewhat insensitive to the changes in $y_m$. (c) The cost function (33) exhibits a flat response around the true value. Hence, the cost function is likely to be at the global stationary point. (d) Honda Accord speed estimates are also proportional to the assumed $y_m$ values. (e) Vehicle width and length estimates follow a similar pattern as in Chevy Impala. (f) The cost function (33) for the Honda Accord at the true $y_m$ value is at a local optimum because the optimization algorithm yields a lower cost value when $y_m = 5.35$m is used as opposed to the true value $y_m = 4.3$m.

## VIII. CONCLUSIONS

We presented a method to determine a vehicle's speed via its acoustic drive-by sounds recorded at a microphone, by formulating the problem as a joint speed and acoustic pattern estimation problem. We achieve this estimation using a vector that profiles the directional variation of the vehicle acoustic pattern. The vehicle profile vector enables a signal processor to better address the vehicle correspondence problem since the vehicle profile vector reduces the bias in speed and loudness estimates as well as vehicle dimensions. It also generates better discriminative features that can be compactly represented by a 15-dimensional space. Parameters $\boldsymbol{\lambda}_v$ and $\boldsymbol{\lambda}_f$ of the vehicle profile vector can improve the confidence of the correspondence matches, also allowing minimal communication between a calibration microphone and a control microphone. However, given the difficulty of the correspondence problem, one should not expect perfect performance for all cases even with the vehicle profile vector.

While determining the vehicle speed, we relied on the signal power calculations and argued that the signal frequency information (Doppler) was not useful when only a single microphone is used. On the other hand, when an array of microphones is available, one can also obtain information from the phase of the received acoustic data across the array. In this case, we expect that the performance should improve more than what is gained from multiple independent amplitude observations. We envision that when multiple vehicles are present, the array can provide the acoustic beam steering necessary to remove the cocktail party effect on the ES components. Hence,

the approaches in the literature can be improved to reduce the biased speed estimates when an array is used. As future work, we plan to investigate how our proposed framework performs when the vehicle CPA's are relatively large.

## IX. ACKNOWLEDGEMENTS

## APPENDIX I
### BIAS ANALYSIS

#### A. Multiplicative Noise vs. Additive Noise

We use the following example to argue that a multiplicative noise model on the signal envelope should be used to estimate the speed in our problem. Let $f_k(v)$ denote monotonically decreasing functions of $v$ for all $k$, e.g., the envelope observations indexed by $k$. Define two generative models as follows

$$\text{Model I: } \mathcal{E}_k = f_k(v) + n_k, \quad \text{Model II: } \mathcal{E}_k = f_k(v)e^{m_k}, \quad (37)$$

where $n_k$ and $m_k$ are *i.i.d.* zero mean Gaussian random variables with variances $\sigma_n^2$ and $\sigma_m^2$. Let $v^*$ be the true speed and $v$ be the maximum likelihood estimate assuming observation noise Model I:

$$v = \arg\min_v J(v), \ J(v) = \sum_k \left(\mathcal{E}_k - f_k(v)\right)^2, \quad (38)$$

Given sufficient number of $\mathcal{E}_k$ measurements, we expect the ML solution to be around the true value of $v^*$. Hence, we can assume $v = v^* + \epsilon$, where $\epsilon$ is small enough so that $f_k(v) \approx f_k(v^*) + \dot{f}_k(v^*)\epsilon$ for all $k$.

The ML estimate of $v$ is found by taking the derivative of $J(v)$ with respect to $v$ and then equating to zero. Now, let us assume that the measurements are actually generated by Model II:

$$J(v) = \sum_k \left(f_k(v^*)e^{m_k} - f_k(v)\right)^2$$
$$\approx \sum_k \left(f_k(v^*)e^{m_k} - f_k(v^*) - \dot{f}_k(v^*)\epsilon\right)^2, \quad (39)$$

which only depends on $\epsilon$. By taking the derivative and equating to zero, we solve the bias and take its expected value:

$$\epsilon = \frac{\sum_k \dot{f}_k(v^*)f_k(v^*)\left(e^{m_k} - 1\right)}{\sum_k \left[\dot{f}_k(v^*)\right]^2},$$
$$\Rightarrow \bar{\epsilon} = \frac{\sum_k \dot{f}_k(v^*)f_k(v^*)\left(e^{\frac{\sigma_m^2}{2}} - 1\right)}{\sum_k \left[\dot{f}_k(v^*)\right]^2}, \quad (40)$$

which is always negative since $f_k(v) > 0$ is a monotonically decreasing function of $v$, i.e., $\dot{f}_k(v^*) < 0$. Hence, given $v^*$, the expected value of the ML estimate of Model I will always have a negative bias when the data is actually generated by Model II. That is, $E\{v|v^*\} - v^* = \bar{\epsilon} < 0$.

#### B. Bias due to Directional Variation

From the ML cost function (33) and the expression for the directional power variation function $\mathcal{A}_\lambda[n_\tau]$ in (28), we can see that the ES components account for a time-varying mean of the logarithm of the power envelope. Hence, method M2 solves for the vehicle speed without taking into account of this mean shift. To understand the effect of this ignorance, we follow the same strategy as above and consider the following generative models

$$\text{Model I: } \log \mathcal{E}_k = \log f_k(v) + n_k,$$
$$\text{Model II: } \log \mathcal{E}_k = \log f_k(v) + \rho_k + m_k, \quad (41)$$

where $m_k$ is *i.i.d.* noise independent of $\rho_k$, whereas $\rho_k$ is the time-varying mean shift that is proportional to the sum of the ES components, where the proportionality constant is positive.

With the same assumptions as in Appendix I.A, we expect the ML solution to be around the true value of $v^*$ for small mean shifts $\rho_k$. When the mean shifts are large, the analysis here is not valid. Hence, we can again assume $v = v^* + \epsilon$, where $\epsilon$ is small enough so that $f_k(v) \approx f_k(v^*) + \dot{f}_k(v^*)\epsilon$ for all $k$. Then, it is straightforward to show that if Model I is used to solve for the speed parameter when Model II is true, the expected bias is given by

$$\bar{\epsilon} = \frac{\sum_k \rho_k}{\sum_k \left(\frac{\dot{f}_k(v^*)}{f_k(v^*)}\right)}. \quad (42)$$

Hence, the expected bias is negative if $\sum_k \rho_k$ is positive (and vice versa) since $f_k(v) > 0$ and $\dot{f}_k(v^*) < 0$. If we look at the results in Figs. 9-11, we see that $\sum_k \rho_k > 0$ for our test cases. Therefore, this result gives an empirical reason why the method M2 results in negatively biased speed estimates.

## REFERENCES

[1] B. G. Quinn, "Doppler speed and range estimation using frequency and amplitude estimates," *J. Acoust. Soc. Am.*, vol. 98, pp. 2560–2566, November 1995.

[2] C. Couvreur and Y. Bresler, "Doppler-based motion estimation for wide-band sources from single passive sensor measurements," in *ICASSP 1997*, 21–24 April 1997.

[3] R. Lopez-Valcarce, C. Mosquera, and R. Perez-Gonzalez, "Estimation of road vehicle speed using two omnidirectional microphones: a maximum likleihood approach," *EURASIP Journal on Applied Signal Processing*, pp. 1059–1077, 2004.

[4] J. F. Forren and D. Jaarsma, "Traffic monitoring by tire noise," in *Proc. IEEE Conf. on Intelligent Transportation System*, Boston, MA, Nov 1997, pp. 177–182.

[5] K. W. Lo and B. G. Ferguson, "Broadband passive acoustic technique for target motion parameter estimation," *IEEE Trans. on Aero. and Elec. Syst.*, vol. 36, pp. 163–175, 2000.

[6] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, 1993.

[7] R. Lopez-Valcarce, D. Hurtado, C. Mosquera, and F. Perez-Gonzalez, "Bias analysis and removal of a microphone array based road traffic speed estimator," in *EUSIPCO*, 2004.

[8] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensor networks," *IEEE Signal Processing Magazine*, vol. 19, pp. 13–14, 2002.

[9] M. Duarte and Y-H. Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, pp. 826–838, 2004.

[10] S. A. Amman and M. Das, "An efficient technique for modeling and synthesis of automotive engine sounds," *IEEE Transactions on Industrial Electronics*, vol. 48, pp. 225–234, February 2001.

[11] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[12] Senate Public Works Committee, *Noise Pollution and Abatement Act of 1972*, S. Rep. No. 1160, 92nd Cong. Second session, 1972.

[13] 86 Stat. 1234 Public Law No. 92-574, *Noise Pollution and Abatement Act of 1972*, codification amended at 42 U.S.C. 4901-4918, 1988.

[14] U. Sandberg, "Tyre/road noise - myths and realities," in *Proceedings of the 2001 International Congress and Exhibition on Noise Control Engineering*, Hague, Netherlands, 27–30 August 2001.

[15] Road Directorate-Ministry of Transport, "Noise reducing pavements," Tech. Rep. 141, Road Directorate, Danish Road Institute, April 2005.

[16] U. Sandberg and A. J. Ejsmont, *Tyre/road noise reference book*, Infomex, SE-59040 Kisa, Sweden, 2002.

[17] P. Andersson, *High frequency tyre vibrations*, Ph.D. thesis, Dept. of Applied Acoustics, Chalmers Univ. of Tech., Gothenburg, Sweden, 2002.

[18] R. A. G. Graf, C. Y. Kuo, A. P. Dowling, and W. R. Graham, "On the horn effect of a tyre/road interfacepart I: experiment and computation," *Journal of Sound and Vibration*, vol. 256, pp. 417–431, 2002.

[19] C. Y. Kuo, R. A. G. Graf, A. P. Dowling, and W. R. Graham, "On the horn effect of a tyre/road interfacepart II: asymptotic theories," *Journal of Sound and Vibration*, vol. 256, pp. 433–445, 2002.

[20] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, pp. 973–973, 1999.

[21] J. G. Lilly, "Engine exhaust noise control," available online at *http://www.ashraeregion7.org*.

[22] R. E. Eskridge and J. C. R. Hunt, "Highway modeling. Part I: prediction of velocity and turbulence fields in the wake of vehicles," *American Meteorological Society*, vol. 79, pp. 387–400, 1979.

[23] N. Sarigul-Klijn, D. Dietz, D. Karnopp, and J. Dummer, "A computational aeroacoustic method for near and far field vehicle noise predictions," *The American Institute of Aeronautics and Astronautics Inc.*, 2001.

[24] P. Kastner-Klein, R. Berkowicz, and E. J. Plate, "Modelling of vehicle induced turbulence in air pollution studies for streets," in *5th Workshop on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes*, Rhodes, Greece, 18–21 May 1998.

[25] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Inc., 1989.

[26] P. Morse and K. Ingard, *Theoretical Acoustics*, McGraw-Hill, 1968.

[27] W. B. Davenport II and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, New York: McGraw-Hill, 1958, ch. 6, sec. 6-4, pp. 93–96.

[28] R. L. Burden and J. D. Faires, *Numerical Analysis*, Brooks/Cole, 2000, 7th Ed.

[29] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[30] L. J. Ziomek, "Three necessary conditions for the validity of the Fresnel phase approximation for the near-field beam pattern of an aperture," *IEEE Journal of Oceanic Engineering*, vol. 18, pp. 73–76, 1993.

[31] Lord Rayleigh, *The Theory of Sound*, Macmillan, London, 1877.

[32] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley New York, 2001.

[33] E. Einstein, "On the electrodynamics of moving bodies," *Annalen der Physik*, vol. 17:891, 1905.

[34] M. J. D. Rendas and J. M. F. Moura, "Cramer-Rao bound for location systems in multipath environments," *IEEE Transactions on Signal Processing*, vol. 39, no. 12, pp. 2593–2610, 1991.

[35] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "Explicit Ziv-Zakai lower bound for bearing estimation," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2810–2824, 1996.

[36] P. Stoica and A. Nehorai, "Music, maximum likelihood, and Cramér-Rao bound," *IEEE Trans. on ASSP*, vol. 37, no. 5, pp. 720–741, May 1989.

[37] S. K. Chow and P. Schultheiss, "Delay estimation using narrow-band processes," *IEEE Transactions on Signal Processing*, vol. 29, no. 3, pp. 478–484, 1981.

[38] A. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems," *IEEE Transactions on Signal Processing*, vol. 31, no. 2, pp. 472–486, 1983.

[39] M. Evans, N. Hastings, and B. Peacock, *Statistical distributions. Ed*, Wiley & Sons. New York, 2000.

[40] R. A. Fisher, *Statistical methods for research workers*, Oliver and Boyd, 1970.

[41] E. B. Wilson and M. M. Hilferty, "The distribution of Chi-square," *Proceedings of the National Academy of Sciences*, vol. 17, no. 12, pp. 684–688, 1931.

[42] V. Cevher, R. Chellappa, and J. H. McClellan, "Gaussian approximations for energy-based detection and localization in sensor networks," in *IEEE Statistical Signal Processing Workshop*, Madison, WI, 26–29 August 2007.

[43] L. Tierney and J. B. Kadane, "Accurate approximations for posterior moments and marginal densities," *Journal of the American Statistical Association*, , no. 81, pp. 82–86, 1986.

[44] J. Liu, X. Koutsoukos, J. Reich, and F. Zhao, "Sensing field: coverage characterization in distributed sensor networks," in *ICASSP 2003*, vol. 5, pp. 173–176.

[45] B. Boashash and P. O'Shea, "A methodology for detection and classification of some underwater acoustic signals using time-frequency analysis," *IEEE Trans. on ASSP*, vol. 38, pp. 1829–1841, Nov 1990.

[46] J. Ding, *Vehicle detection by sensor network nodes*, Ph.D. thesis, University of California, Berkeley, CA, 2003.

[47] S.J. Searle, "Efficient matched processing for localisation of a moving acoustic source," *Signal Processing*, vol. 85, no. 9, pp. 1787–1804, 2005.

[48] V. Cevher, F. Guo, A. C. Sankaranarayanan, and R. Chellappa, "Joint acoustic-video fingerprinting of vehicles, part II," in *ICASSP 2007*, Honolulu, Hawaii, 15–20 April 2007.

[49] M. Gavish and A.J. Weiss, "Performance analysis of bearing-only target location algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 817–828, 1992.