

JOINT ACOUSTIC-VIDEO FINGERPRINTING OF VEHICLES, PART I

V. Cevher*, R. Chellappa

Center for Automation Research,
University of Maryland,
College Park, MD 20742

J. H. McClellan

Center for Signal and Image Processing,
Georgia Institute of Technology,
Atlanta GA 30332

ABSTRACT

We address vehicle classification and mensuration problems using acoustic and video sensors. In this paper, we show how to estimate a vehicle's speed, width, and length by jointly estimating its acoustic wave-pattern using a single passive acoustic sensor that records the vehicle's drive-by noise. The acoustic wave-pattern is approximated using three envelope shape (ES) components, which approximate the shape of the received signal's power envelope. We incorporate the parameters of the ES components along with estimates of the vehicle engine RPM and number of cylinders to create a vehicle profile vector that forms an intuitive discriminatory feature space. In the companion paper, we discuss vehicle classification and mensuration based on silhouette extraction and wheel detection, using a video sensor. Vehicle speed estimation and classification results are provided using field data.

Index Terms— Object recognition, pattern recognition, acoustic applications, acoustic signal processing, intelligent sensors

1. INTRODUCTION

Vehicle classification and motion parameter estimation using signals received at passive sensors are classical signal processing problems [1–8]. Traditionally, these two problems are treated separately. In this paper, we address how to jointly estimate a vehicle's motion parameters and classification features, using a single acoustic microphone. This problem has applications in distributed sensor networks and traffic monitoring [5,8,9].

When a single acoustic microphone is used, wave propagation effects can determine the source movements based on the assumptions that the vehicle \mathcal{A} is a point source [6,7], \mathcal{B} has stationary signal characteristics that admit a model such as an autoregressive moving average (ARMA) model [7], or \mathcal{C} produces a pure tone [6]. These assumptions are only partially satisfied in practice and the speed estimation algorithms based on these assumptions do not perform as expected when applied to field data. In fact, they all exhibit a negative speed estimation bias [10].

On the other hand, the acoustic vehicle classification algorithms concentrate on the frequency spectra of the vehicles, without explicitly exploiting the directionality of vehicle noise [1–4]. The frequency spectra of training signals are somewhat judiciously segmented to provide features, where discriminant functions [1], neural networks [3,4], and Bayesian methods [2] are used for classification. In contrast, Forren and Jaarsma [8] aim to classify vehicles with an array of microphones based on their axle detections by exploiting the tire noise generated by vehicles. They use signal correlations among three known microphones under as-

sumption \mathcal{A} to independently estimate the vehicle speed to improve their classification results.

In this paper, we first describe the spectral and spatial content of vehicle signals and recast the estimation problem in [6,7] as a spatial acoustic pattern recognition problem parameterized by vehicle speed and classification features such as vehicle wheelbase length, width, and tire profile. 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. For estimation, we introduce a vehicle profile vector that characterizes the ES components and also includes the engine revolutions per minute (RPM) and the number of cylinders to characterize the vehicle engine. The vehicle profile vector constitutes a *fingerprint* of the vehicle that can be estimated using the vehicle's drive-by noise without any training data. In the companion paper [11], this acoustic vehicle profile vector is extended to also include a subset of complimentary video fingerprint information.

2. 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, break 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 depends 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 vehicle's signal as consisting of four noise components:

Engine Noise: The noise from an internal combustion engine contains a deterministic harmonic train and a stochastic component similar to the human speech [12,13]. The stochastic component is wide-band 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 $f_0 = \chi/120$, defined as the firing rate of any one cylinder in the engine, where χ is the RPM of a vehicle. The strongest tone in the engine noise is called the engine fire rate F_0 , and $F_0 = f_0 \times p$, where p is the number of cylinders.

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 after 50km/h. It consists of two components: vibrational noise and air noise [14]. The vibrational component is caused by the contact between the tire threads and the pavement texture. Its spectrum is most dominant between 100 – 1000Hz frequency range. The air noise is generated by the air being sucked-in or

Prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-02-0008.

forced out of the rubber blocks of a tire and is dominant in the frequency ranges between 1000 – 3000Hz. In the driving direction of the vehicle, the road and the tire forms a geometrical structure that amplifies the noise generated by the tire-road interaction [14, 15]. This effect is called the *horn effect* and has a directional pattern [15].

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. The exhaust noise has the same harmonics frequency structure as the engine and additional tail pipe resonances that occur at fundamental frequency of $f_e = c/(2l)$, where l is the tail pipe length and c is the speed of sound [16].

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. The turbulence noise depends on the vehicle aerodynamics as well as the ambient wind speed and orientation [17].

3. JOINT ESTIMATION OF SPEED AND SPATIAL ACOUSTIC PATTERNS

3.1. Envelope Shape Components

To determine a vehicle's speed v using acoustic observations from a single microphone, we jointly estimate the vehicle's spatial acoustic pattern. We denote any component that makes up a vehicle's spatial acoustic pattern as an *envelope shape* (ES) component. We use the reciprocity theorem derive the ES components in the microphone reference frame [18]. For simplicity, we model the ES components using three piecewise constant functions in dB scale with respect to the microphone bearing φ as illustrated in Fig. 1. We make the connection between the ES components and the received signal power in the following subsections.

The first ES component $\rho_\gamma(\varphi)$ in Fig. 1 models the signal interference due to the front and rear tires, which can be modeled as dipole sources. This component explains the perturbation in the envelope function of the vehicle acoustic drive-by signals around $\varphi = 18^\circ$. The tire interference decreases between the bearings γ_1 and γ_2 , increasing the first component to $\delta_{\gamma,1}$. The angles γ_i are related to the tire track length of the vehicle W (dipole separation) through an interference term [10].

After the drive-by, the tire interference increases between the bearings γ_2 and γ_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 φ . 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 by ϕ , is related to the angles in the microphone frame, denoted φ , through an aberration relation [19]. 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 where $\kappa = \gamma_1$ and γ_2 :

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

The second ES component $\rho_\theta(\varphi)$ is due to the horn effect [15]. In the observed signal envelope, at the microphone bearing θ_1 the horn effect amplification of the farthest front tire from the microphone starts to go down until the bearing θ_2 to $\delta_{\theta,2}$, when the horn effect of the closest rear tire to the microphone also drops. The differential angle $\theta_2 - \theta_1$ is a good indicator of the vehicle wheelbase, which can be used to compare the relative size of vehicles. To convert the angle difference into actual

size, we use the following approximate relationship

$$\frac{y_m}{\cos \theta_1} - L = \frac{y_m}{\cos \theta_2}, \quad (2)$$

where L is the wheelbase length. After the vehicle CPA, the horn effect amplifies the tire noise between bearings θ_2' and θ_1' , which are related to θ_1 and θ_2 also by the aberration relation (1) (i.e., $\kappa = \theta_i$).

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. To keep the number of ES components manageable, we approximate the composite interference pattern as a step function that rises from 0dB to δ_ψ between bearings ψ_1 and ψ_2 . When this approximation 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, (2) is also used to relate the angles in the third interference component to the vehicle wheelbase length L .

3.2. Vehicle Profile Vector

To jointly determine the speed and the vehicle's spatial acoustic pattern, we use the vehicle profile vector λ , which is defined as follows:

$$\lambda = [\lambda_v \quad \lambda_\varphi \quad \lambda_\delta \quad \lambda_f], \text{ where} \quad (3)$$

$$\lambda_v = [S \quad v \quad W \quad L], \lambda_\varphi = [\varphi_0 \quad \gamma_1 \quad \theta_1 \quad \psi_1], \quad (4)$$

$$\lambda_\delta = [\delta_{\gamma,1} \quad \delta_{\gamma,2} \quad \delta_{\theta,1} \quad \delta_{\theta,2} \quad \delta_\psi], \text{ and } \lambda_f = [\chi \quad p].$$

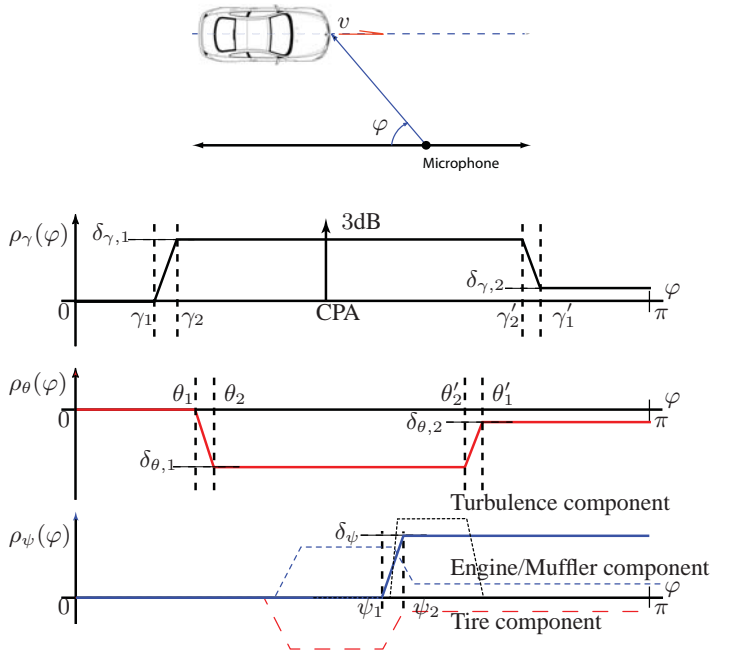


Fig. 1. The microphone bearing reference orientation is defined as the moving direction of the vehicle. Then, a vehicle's spatial acoustic pattern can be approximated by three ES components in φ . The first component ρ_γ is due to the signal interference from the front and rear tires. The second component ρ_θ explains the variation as the microphone comes out of the horn-effect area of the tires. The third component ρ_ψ is an approximate component that accounts for a composite engine/exhaust/tire/turbulence noise effect around the vehicle CPA.

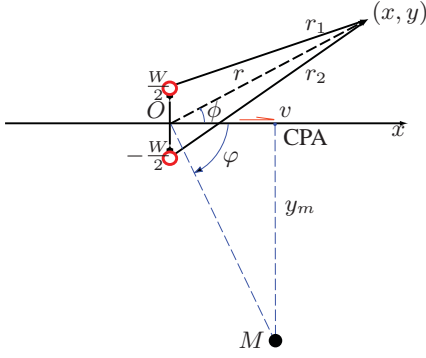


Fig. 2. Dipole geometry. When the dipole sources are correlated, the resulting wave propagation effect on the received signal power is not a superposition of individual monopole effects.

The vector λ_v consists of the physical parameters of the vehicle such as the loudness S , speed v , tire track length W , and the vehicle wheelbase length L . The vector λ_φ has the initial vehicle bearing φ_0 and the angles that define the ES components along with λ_δ , which contains the amplitude attenuations and amplifications for the ES components. Lastly, the vector λ_f has the RPM χ and the number of cylinders p of the vehicle.

3.3. Amplitude Observations

Let $s(t)$ be a zero-mean *i.i.d.* acoustic signal emitted by the monopole source. To simplify the results, we concentrate on the following special case, where the Fourier transform of the source signal is assumed to be bandlimited as follows:

$$|S(\Omega)| = \begin{cases} S, & \Omega_1 \leq \Omega \leq \Omega_2, \mathcal{W} = \Omega_2 - \Omega_1; \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

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

$$z[n] = \sum_{i=1,2} \frac{1}{r_i[n]} s \left(\beta_i[n] \frac{n}{F_s} - \frac{r_i[n]}{c} \right), \quad (6)$$

where $\beta_i(t)$ ($i = 1, 2$) is the Doppler shift factor of each monopole source in the dipole. Under the far field assumptions [20], one can approximate $r_i \approx r$ and $\beta_i \approx \beta$ as shown in Fig. 2.

We define the power envelope function by using τ -discrete samples of $z[n]$ as follows:

$$\mathcal{E}[n_\tau] = \mathcal{E}(t) \Big|_{t=\frac{n_\tau \tau}{F_s}} = \sqrt{\frac{1}{\tau} \sum_{m=n}^{n+\tau-1} |z[m]|^2}, \quad (7)$$

where subscript τ under the sample index n implies that the samples of the continuous function are calculated at every τ/F_s second. The parameter τ is chosen so that the DFT coefficients used to calculate the power function at τ -samples apart are statistically uncorrelated, and hence, each sample of $\mathcal{E}[n_\tau]$ ($n_\tau = 0, \dots, N_\tau - 1$) is also statistically uncorrelated.

Assuming that the noise acting on the microphone signal $z[n]$ is zero-mean additive white Gaussian noise with variance σ_u^2 , we relate the enve-

lope observations to the vehicle profile vector as follows:

$$\begin{aligned} \mathcal{E}^2[n_\tau] &\approx \mathcal{A}_\lambda[n_\tau] e^{2m_\tau} + \frac{\sigma_u^2}{\tau} w_\tau, \\ \mathcal{A}_\lambda[n_\tau] &= \frac{C}{\beta[n_\tau] r^2[n_\tau]} \prod_{i=\gamma, \theta, \psi} 10^{\rho_i(\varphi[n_\tau])/10}, \quad C = \frac{S^2 F_s W}{2\pi\tau} \end{aligned} \quad (8)$$

where $\mathcal{A}_\lambda[n_\tau]$ is the directional power variation, e^{m_τ} is an *i.i.d.* multiplicative noise on the signal amplitude ($m \sim \mathcal{N}(0, \sigma_m^2)$), w_τ is an *i.i.d.* additive Chi2 $_\tau$ noise (chi-squared distribution with τ degrees of freedom) that is also independent of m_τ , and

$$\beta[n_\tau] = 1 + \frac{v}{c} \cos \varphi[n_\tau],$$

$$r[n_\tau] = \sqrt{(v\tau/F_s)^2 + r^2[n_\tau - 1] - 2(v\tau/F_s) r[n_\tau - 1] \cos \varphi[n_\tau - 1]},$$

$$\varphi[n_\tau] = \varphi[n_\tau - 1] + \sin^{-1} \left(\frac{v\tau}{F_s r[n_\tau]} \sin \varphi[n_\tau - 1] \right). \quad (9)$$

When a constant false alarm rate detector such as in [5] is used, the Chi2 noise can be ignored and the following cost function can be minimized to determine the maximum likelihood estimate of λ :

$$\lambda_{\text{ML}} = \arg \min_{\lambda} J(\lambda), \quad J(\lambda) = \sum_{n_\tau=0}^{N_\tau-1} \left(\log \mathcal{E}[n_\tau] - \frac{1}{2} \log \mathcal{A}_\lambda[n_\tau] \right)^2. \quad (10)$$

3.4. Frequency Observations

The spectral content of a vehicle exhibits directional variation, making it difficult to use the Doppler 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. 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 – 250Hz range at nominal RPM's. At moderate vehicle speeds (30 – 50mph), the full Doppler shift swings F_0 approximately %6, also corresponding to an RPM change of the same percentage ($\Delta\chi \approx 200$). Hence, if a driver changes the car's RPM by 50 during the vehicle drive-by, there will be a %25 error in F_0 when one assumes a constant frequency source. 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 a probability density function for the vehicle profile vector by fitting a Doppler shift function to the engine and tire frequency tracks is an unreliable approach [7].

On the other hand, the spectral harmonic content can be used to determine λ_f of the vehicle profile vector. Moreover, conditioned on λ_v estimate, it possible to further refine λ_f by compensating for the Doppler shifts. Estimation of χ can be done accurately using harmonic analysis methods [21]. In our estimation, we use the power spectral density of the acoustic signal to determine λ_f . Details can be found in [21].

4. A VEHICLE PROFILING EXAMPLE

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 microphone pole to the center of the street is 7.4m. A video camera is also used for estimation, whose results are discussed in the companion paper.

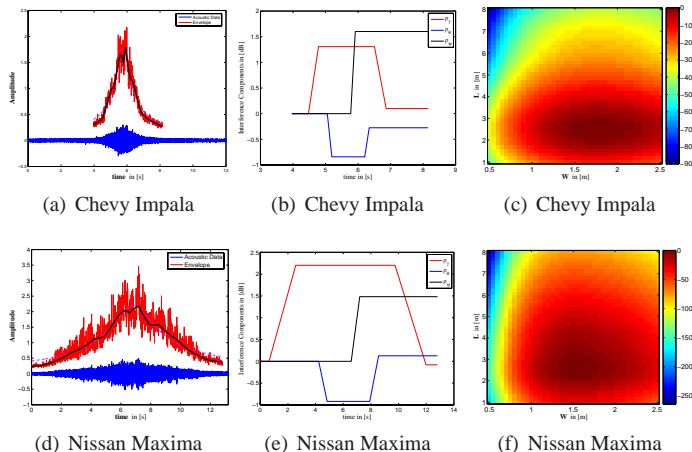


Fig. 3. (a,d) Observed envelope exhibits significant variations as explained by our model (solid lines). (b,e) Estimated ES components are shown. (c,f) The log-likelihood surface of the vehicle dimensions.

Figure 3 gives two sample estimation results using recorded acoustic signals of a Chevy Impala and Nissan Maxima drive-by's. More acoustic processing results can be found in the second part in [10, 11]. For this example, the envelope estimates are obtained using $\tau = 480$ samples. The vehicles' distances to the camera are (5.8, 4.1)m, respectively. The vehicle speeds are estimated as (18.68, 4.14)m/s by the video camera and (18.60, 4.49)m/s by the acoustic method described in this paper. In contrast, the acoustic method outlined in [7] estimates the vehicle speeds as (15.05, 3.46)m/s (dashed lines in Fig. 3(a,d)) consistent with the negative bias noted in their paper. Other vehicle profile vector parameters are estimated as $C = (12.6, 6.34)$, $L = (2.58, 2.58)$ m, $W = (1.75, 1.50)$ m, $\chi = (3300, 3150)$, and $p = (6, 4)$. We note that the manufacturer's specifications for the vehicle dimensions are $L = (2.8, 2.7)$ m and $W = (1.58, 1.53)$ m. The vehicle engines are also known to have (6, 4)-cylinders. We use the inverse of the log-likelihood surface Hessian at the mode (Fig. 3(c)) to estimate the variance of the acoustic estimation performance for the vehicle dimensions. For these examples, $\sigma_W \approx (17.3, 8.9)$ cm, whereas $\sigma_L \approx (31.1, 19.0)$ cm.

5. CONCLUSIONS

We presented a method that can estimate a vehicle's motion parameters and classification features by formulating the problem as a joint speed and acoustic pattern estimation problem. We achieve this estimation using the vehicle profile vector that encodes the directional variation of the vehicle acoustic pattern. The vehicle profile vector enables a signal processor to better address the vehicle classification and mensuration problem using acoustic signals since it results in vehicle dimensions as well as unbiased speed and loudness estimates [10].

6. REFERENCES

- [1] H. Wu, M. Siegel, and P. Khosla, "Vehicle sounds signature recognition by frequency vector principal component analysis," in *IEEE Instr. and Meas. Tech. Conf.*, St. Paul, MN, 18–20 May 1998.
- [2] M. E. Munich, "Bayesian subspace methods for acoustic signature recognition of vehicles," in *EUSPICO*, Vienna, Austria, 2004.
- [3] S. Sampan, *Neural fuzzy techniques in vehicle acoustic signal classification*, Ph.D. thesis, Virginia Polytechnic Ins., VA, 1997.
- [4] L. Liu, *Ground vehicle acoustic signal processing based on biological hearing models*, Ph.D. thesis, Institute of Systems Research, UMD, College Park, MD, 1999.
- [5] M. Duarte and Y-H. Hu, "Vehicle classification in distributed sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, pp. 826–838, 2004.
- [6] B. G. Quinn, "Doppler speed and range estimation using frequency and amplitude estimates," *J. Acoust. Soc. Am.*, vol. 98, pp. 2560–2566, November 1995.
- [7] 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.
- [8] 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.
- [9] S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensors networks," *IEEE Signal Processing Magazine*, vol. 19, pp. 13–14, 2002.
- [10] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," under revision at *IEEE Transactions on Signal Processing*, available at http://www.umiacs.umd.edu/users/volkan/Speed_Estimation.pdf.
- [11] V. Cevher and *et al.*, "Joint acoustic-video fingerprinting of vehicles, part II," submitted to *ICASSP 2007*, available at <http://www.umiacs.umd.edu/users/volkan/javf2.pdf>.
- [12] 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.
- [13] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- [14] U. Sandberg and A. J. Ejsmont, *Tyre/road noise reference book*, Infomex, SE-59040 Kisa, Sweden, 2002.
- [15] 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," *J. of Sound and Vib.*, vol. 256, pp. 417–431, 2002.
- [16] J. G. Lilly, "Engine exhaust noise control," available online at <http://www.ashraeregion7.org>.
- [17] 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.
- [18] Lord Rayleigh, *The Theory of Sound*, Macmillan, London, 1877.
- [19] E. Einstein, "On the electrodynamics of moving bodies," *Annalen der Physik*, vol. 17:891, 1905.
- [20] L. J. Ziomek, "Three necessary conditions for the validity of the Fresnel pahse approximation for the near-field beam pattern of an aperture," *IEEE Journal of Oceanic Engineering*, vol. 18, pp. 73–76, 1993.
- [21] 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.