

Estimation of road vehicle speed using two omnidirectional microphones: a Maximum Likelihood approach

Roberto López-Valcarce*, Carlos Mosquera and Fernando Pérez-González

Departamento de Teoría de la Señal y las Comunicaciones

Universidad de Vigo, 36200 Vigo, Spain

Ph. +34 986 812677, Fax +34 986 812116

valcarce@gts.tsc.uvigo.es, {mosquera,fperez}@tsc.uvigo.es

Abstract

We address the problem of estimating the speed of a road vehicle from its acoustic signature, recorded by a pair of omnidirectional microphones located next to the road. This choice of sensors is motivated by their nonintrusive nature as well as low installation and maintenance costs. A novel estimation technique is proposed, which is based on the maximum likelihood principle. It directly estimates car speed without any assumptions on the acoustic signal emitted by the vehicle. This has the advantages of bypassing troublesome intermediate delay estimation steps as well as eliminating the need for an accurate yet general enough acoustic traffic model. An analysis of the estimate for narrowband and broadband sources is provided and verified with computer simulations. The estimation algorithm uses a bank of modified crosscorrelators and therefore it is well suited to DSP implementation, performing well with preliminary field data.

Keywords

Speed estimation, traffic monitoring, microphone arrays.

*Corresponding author. The work of R. López-Valcarce is supported by a *Ramón y Cajal* grant from the Spanish Ministry of Science and Technology.

I. INTRODUCTION

Nowadays several alternatives exist for collecting numerical data about the transit of road vehicles at a given location. From these data, parameters such as traffic density and flow are estimated in order to develop effective traffic management strategies. Thus, traffic management schemes heavily depend on an infrastructure of sensors capable of automatic monitoring traffic conditions. The design of such systems must include the choice of the type of sensor and the development of adequate signal processing and estimation algorithms [9]. Cheap sensor-based networks enable dense spatial sampling on a road grid, so that meaningful global results can be extracted; this is the so-called Collaborative Information Processing paradigm [7], an emerging interdisciplinary research area tackling different issues such as data fusion, adaptive systems, low power communication and computation, etc.

Traffic sensors commercially available at present include magnetic induction loop detectors; radar, infrared or ultrasound-based detectors; video cameras and microphones. All of them present different characteristics in terms of robustness to changes in environmental conditions; manufacture, installation and repair costs; safety regulation compliance, etc. A desirable system would: i) be passive, to avoid radiation emissions and/or operate at low power; ii) operate in all-weather day-night conditions, and iii) be cheap and easy to install and maintain. Although these objectives can be achieved by microphone-based schemes, commercially available systems employ highly directive microphones which considerably increase the cost. Alternatively, the use of cheap (i.e. omnidirectional) sensors must be compensated for with more sophisticated algorithms. In addition, power-aware signal processing methods are mandatory to meet the energy constraints of battery-powered sensors.

In this paper we address the problem of how to directly estimate the speed of a vehicle moving along a known transversal path (e.g. a car on a road) from its acoustic signature. Previous related work using a single sensor usually relied on some sort of assumption on the source (e.g. narrowband signals of known frequency [18] or time-varying ARMA models [8]). It is known, however, that an important component of the acoustic signal emitted by a vehicle consists of several tones harmonically related [2], as expected from a rotating machine. Furthermore, the noise caused by the friction of the vehicle tires can also be relevant, especially for high speeds, incorporating a broadband component which is hard to model [10]. As a consequence, acoustic waveforms generated by wheeled and tracked vehicles may have significant spectral content ranging from a few tens of Hz up to several kHz, yielding a ratio of the maximum to the minimum frequency components of at least 100 [6]. These characteristics of road vehicle acoustic signals make robust modeling a difficult task, given the great variability within the vehicle population [12].

This problem could be avoided by including a second sensor, which is the approach we adopt: a pair of omnidirectional microphones are placed alongside the known

path of the moving source. For a review on the topic of parameter estimation from an array of sensors, see the excellent paper by Krim and Viberg [11]. However, most research on array processing is devoted to the problem of direction of arrival (DOA) or differential time delay (DTD) estimation of narrowband or broadband sources for radar and sonar applications. Target motion is usually considered a nuisance that must be compensated for [3], [4], or is studied through the analysis of the time variation of the DTD over consecutive processing windows [15]. An exception is the Stochastic Maximum Likelihood (SML) approach of Stuller [16], [17], who assumed a random Gaussian source with known power spectrum and an arbitrarily parameterized time-varying DTD, and then provided the generic form of the likelihood function for the estimation of the DTD parameters.

As noted above, the Gaussian model does not seem adequate for acoustic traffic signals. Therefore, we adopt a Deterministic Maximum Likelihood (DML) approach: waveforms are treated as deterministic (arbitrary) but unknown within this framework in order to estimate the only parameter we are interested in, i.e. vehicle speed, which is assumed constant. The resulting (approximate) likelihood function can be efficiently computed, and the geometric structure of the problem allows for an approximate analysis that reveals the influence of the different parameters such as frequency, range, and sensor separation.

Two works directly studying the same problem as here are [5], designed for ground vehicles, and [13] for airborne targets. Both use the same principle, namely, short time crosscorrelations assuming local stationarity to extract the temporal variation of the delay between the received signals. As opposed to these, ours is a direct approach which estimates the speed in a single step, without intermediate time-delay estimations which would increase the error in the final result.

Section II gives a detailed description of the problem, and a near maximum likelihood estimate is derived in Section III together with an efficient DSP oriented implementation. Analyses are developed in Sections IV and V, followed by simulation and experimental results in Sections VI and VII.

II. PROBLEM DESCRIPTION

Figure 1 illustrates the problem. The microphones M_1 , M_2 are separated $2b$ m and placed D m from the road center. The vehicle travels at constant speed v_0 on a straight path along the road. The time reference is set at the closest point of approach (CPA) so that $t = 0$ when the vehicle is equidistant to M_1 and M_2 . The (time-varying) distances from the vehicle to the microphones are

$$d_1(t; v_0) = \sqrt{D^2 + (v_0 t + b)^2}, \quad d_2(t; v_0) = \sqrt{D^2 + (v_0 t - b)^2} \quad (1)$$

so that the propagation time delays are $\tau_i(t; v_0) = d_i(t; v_0)/c$, where c is the sound propagation speed. The observation window is $(-T/2, T/2)$. We also define the

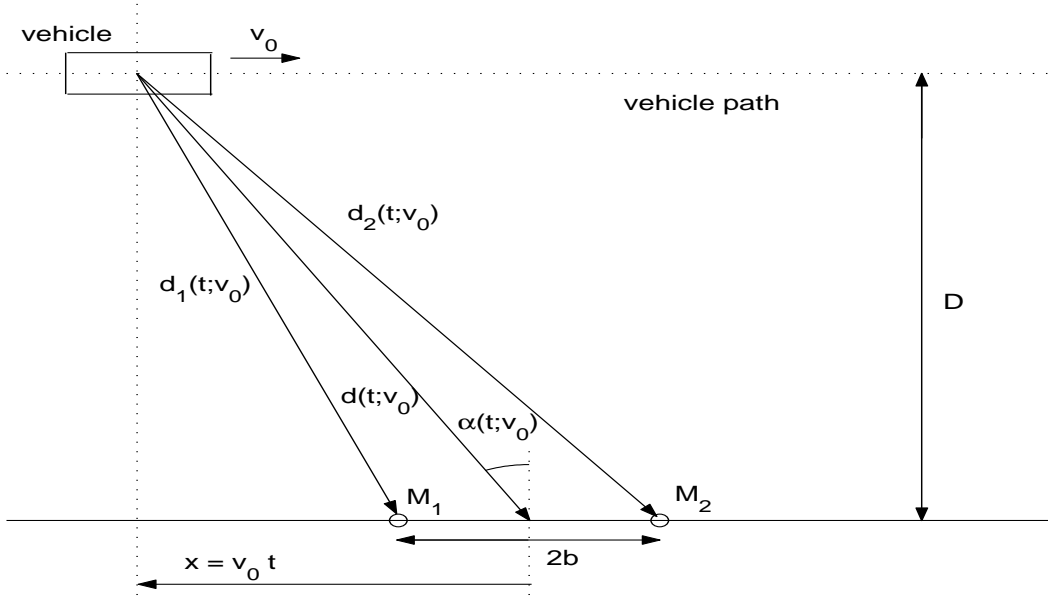


Fig. 1. Geometry of the problem

angle and distance between the source and the array center respectively as

$$\alpha(t; v_0) = \text{atan} \frac{v_0 t}{D}, \quad d(t; v_0) = \frac{D}{\cos \alpha(t; v_0)}, \quad (2)$$

and the ‘angular aperture’ α_0 denoting the observation limit in the angular domain:

$$\alpha_0 \triangleq \alpha(T/2; v_0) = \text{atan} \frac{v_0 T}{2D}. \quad (3)$$

Let the sound wave generated by the vehicle be $s(t)$, which is assumed to be deterministic but unknown. Taking into account the attenuation of sound with distance, we can express the received signal at sensor M_i as

$$r_i(t) = s_i(t) + w_i(t) \quad \text{with} \quad s_i(t) \triangleq \frac{s(t - \tau_i(t; v_0))}{d_i(t; v_0)} \approx \frac{s(t - \tau_i(t; v_0))}{d(t; v_0)}. \quad (4)$$

The approximation in (4) will be adopted throughout. The noise processes $w_1(\cdot)$, $w_2(\cdot)$ are assumed stationary, independent and Gaussian with zero mean. Assuming an ideal antialiasing filter preceding the A/D conversion in the signal processor, we model their power spectral density and autocorrelation respectively as

$$S_w(f) = \begin{cases} N_0/2 \text{ W/Hz}, & |f| < f_s/2 \\ 0, & \text{otherwise,} \end{cases} \quad R_w(\tau) = \frac{N_0 f_s}{2} \text{sinc}(f_s \tau), \quad (5)$$

where $f_s = 1/T_s$ denotes the sampling frequency. Hence, the samples $w(kT_s)$ are uncorrelated zero-mean Gaussian with variance $\sigma^2 = N_0 f_s/2$. The problem is to find an estimate of v_0 given the signals $r_i(t)$, and without knowledge of $s(t)$.

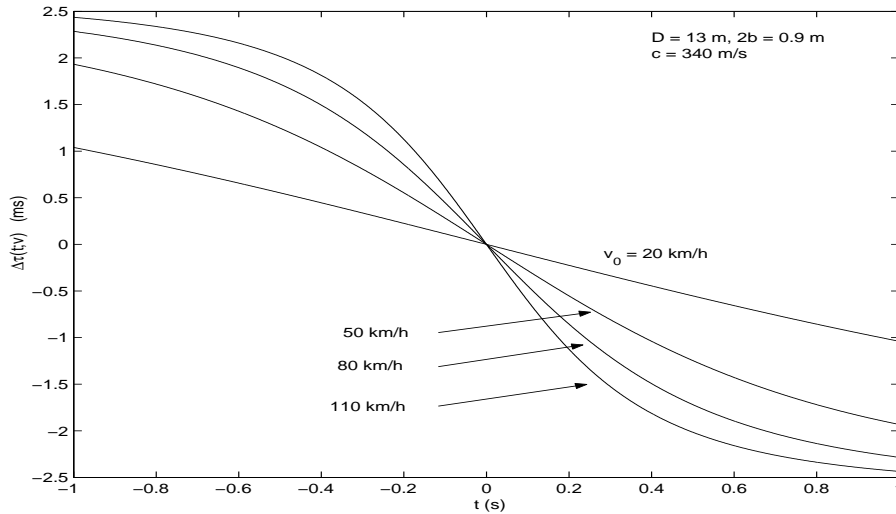


Fig. 2. The differential delay $\Delta\tau(t; v_0)$ for different values of the source speed.

[5] proposes to estimate the differential time delay (DTD) between $r_1(t)$ and $r_2(t)$

$$\Delta\tau(t; v_0) \triangleq \tau_2(t; v_0) - \tau_1(t; v_0) \quad (6)$$

$$\approx -\frac{2b}{c} \sin \alpha(t; v_0) \quad \text{if } b/D \ll 1, \quad (7)$$

using short-time cross correlations and peak picking. Then, noting that

$$\left. \frac{\partial \Delta\tau(t; v_0)}{\partial t} \right|_{t=0} = -\frac{2b}{Dc} v_0,$$

(see Fig. 2), it is seen that v_0 can be estimated from the slope of the (itself estimated) DTD at the CPA. [5] considers directional microphones and does not provide an explicit method to extract the estimate of v_0 from that of the DTD. Instead we derive a direct ML approach in the next section, which will be shown to compare favorably to the indirect method of [5].

III. APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATE

A. Derivation

Consider first the problem of estimating v_0 without knowledge about $s(t)$ and with a single sensor M_1 . Then the ML estimate is given by $\hat{v}_{\text{ml}} = \arg \max_v p(\mathbf{r}_1|v)$, where \mathbf{r}_1 is the vector of observations. However, since $s(t)$ is completely unknown, one cannot extract any information about v_0 from \mathbf{r}_1 : any effect that we may expect v_0 to produce on \mathbf{r}_1 can be canceled by proper choice of $s(t)$. Thus, *without any knowledge of $s(t)$* , $p(\mathbf{r}_1|v) = p(\mathbf{r}_1)$, i.e. all values of v are equally likely.

With two sensors, one has $\hat{v}_{\text{ml}} = \arg \max_v p(\mathbf{r}_1, \mathbf{r}_2|v)$. By the reasoning above,

$$p(\mathbf{r}_1, \mathbf{r}_2|v) = p(\mathbf{r}_2|\mathbf{r}_1, v)p(\mathbf{r}_1|v) = p(\mathbf{r}_2|\mathbf{r}_1, v)p(\mathbf{r}_1). \quad (8)$$

Hence the ML estimate reduces to $\arg \max_v p(\mathbf{r}_2 | \mathbf{r}_1, v)$. In order to obtain this pdf, we must find a relation between the two received signals $r_1(t)$, $r_2(t)$. Intuitively, if we time compand $r_1(t)$ by an appropriate amount which will depend on v_0 , then the resulting signal should be time aligned with $r_2(t)$. Letting $f(t) \triangleq t - \tau_1(t; v_0)$, and neglecting the effect of small time shifts in $1/d(t; v_0)$ (since it varies much more slowly than $s(t)$), the noiseless signals can be related via

$$s_2(t) = s_1(f^{-1}(t - \tau_2(t; v_0))) = s_1(u(t)), \quad (9)$$

where $u(t) \triangleq f^{-1}(t - \tau_2(t; v_0))$. To find u , note from the definitions of f and u that

$$f(u) = u - \tau_1(u; v_0) = t - \tau_2(t; v_0) \quad (10)$$

$$\Rightarrow u - \tau_1(u; v_0) + \tau_1(t; v_0) = t - \Delta\tau(t; v_0). \quad (11)$$

Since u is close to t , it is reasonable to make the following first-order approximation:

$$\tau_1(t; v_0) \approx \tau_1(u; v_0) + (t - u) \left. \frac{\partial \tau_1(t; v_0)}{\partial t} \right|_{t=u} \quad (12)$$

which is used to substitute $\tau_1(t, v_0)$ in (11):

$$u(t) \approx t - \frac{\Delta\tau(t; v_0)}{1 - \left. \frac{\partial \tau_1(t; v_0)}{\partial t} \right|_{t=u}}. \quad (13)$$

Observe that for practical values of the speed ($|v_0| \ll c$), one has

$$\left| \frac{\partial \tau_1(t; v_0)}{\partial t} \right|_{t=u} = \left| \frac{v_0}{c} \cdot \frac{(v_0 u + b)}{\sqrt{D^2 + (v_0 u + b)^2}} \right| \leq \frac{|v_0|}{c} \ll 1, \quad (14)$$

so that $u(t) \approx t - \Delta\tau(t; v_0)$, and we obtain the following fundamental approximation:

$$s_2(t) \approx s_1(t - \Delta\tau(t; v_0)). \quad (15)$$

Using this intuitively appealing relation, the ML estimate readily follows. Note that

$$\begin{aligned} r_2(t) &= s_2(t) + w_2(t) \\ &\approx s_1(t - \Delta\tau(t; v_0)) + w_2(t) \\ &= r_1(t - \Delta\tau(t; v_0)) - w_1(t - \Delta\tau(t; v_0)) + w_2(t). \end{aligned} \quad (16)$$

Let $w(t) = w_2(t) - w_1(t - \Delta\tau(t; v_0))$. Since for all practical values of v_0 , b , D , the DTD $\Delta\tau(t; v_0)$ varies much more slowly than t (see Fig. 2), in view of (5) the samples $w(kT_s)$ are approximately uncorrelated, with variance $2\sigma^2$. Therefore the conditional pdf $p(\mathbf{r}_2 | \mathbf{r}_1, v)$ is approximately normal, so that the ML estimate should minimize the squared Euclidean norm $\|\mathbf{r}_2 - \mathbf{r}_1(v)\|^2$, where $\mathbf{r}_1(v)$ is the vector of samples from the signal $r_1(t - \Delta\tau(t; v))$. Equivalently, it should maximize

$$\begin{aligned} &\langle \mathbf{r}_1(v), \mathbf{r}_2 \rangle - \frac{1}{2} \|\mathbf{r}_1(v)\|^2 \\ &= \int r_1(t - \Delta\tau(t; v)) r_2(t) dt - \frac{1}{2} \int r_1^2(t - \Delta\tau(t; v)) dt. \end{aligned} \quad (17)$$

The second term in the right-hand side of (17) is approximately constant with v . Therefore we propose the following estimator:

$$\hat{v}_0 = \arg \max_v \psi(v) = \arg \max_v \int_{-T/2}^{T/2} r_1(t - \Delta\tau(t; v)) r_2(t) dt. \quad (18)$$

B. Discussion

It is seen that the ML estimate (18) does not require short-time based estimates of the DTD. Instead it exploits knowledge of the parametric dependence of the DTD with v in order to accordingly time-comband the signals that enter the crosscorrelation, which is computed over the whole observation window for each candidate speed. It can be asked whether this approach may provide a substantial advantage over the indirect one of [5]. To give a quantitative comparison, consider a simplified model $r_1(t) = s(t) + w_1(t)$, $r_2(t) = s(t - \Delta\tau(t; v_0)) + w_2(t)$ in which attenuations have been neglected. Further, assume that the observation window is small so that the DTD appears to be linear for all practical values of v_0 , i.e. $\Delta\tau(t; v_0) \approx q_0 t$ for $|t| < T/2$, with $q_0 = -2bv_0/Dc$. Under such conditions, estimating v_0 is equivalent to estimating the relative time companding (RTC) parameter q_0 . This problem was considered by Betz [3], [4] under Gaussianity of signal and noise. In that case, following his development it can be shown that the estimation accuracy of the indirect approach with respect to the Cramer-Rao bound (CRB) is given by

$$\frac{\text{var}[\hat{q}_0]}{\text{CRB}(q_0)} = \frac{1}{9} \Omega^2(\pi B T' q_0), \quad (19)$$

where B is the signal bandwidth, $T' < T$ is the subwindow size used for short-time DTD estimation in the indirect method, and $\Omega(x) = x^3/(\sin x - x \cos x)$. The loss (19) is minimized when T' is, for given B and q_0 . Note that T' should be at least twice the value of the largest expected value of the DTD, which in our case is $2b/c$ (≈ 3 ms for a typical sensor separation of 1 m). Fixing $T' = 6$ ms, the loss (19) at $q_0 = 0.04$ (a typical RTC value for high speeds in arrays set close to the road) is of 2, 5 and 9 dB for bandwidths of 2, 3 and 4 kHz respectively.

These observations do favor the direct ML estimate over the indirect one. The simulation and experimental results in Sections VI and VII (obtained under the more general model (4)) will provide additional support for this claim.

C. Implementation issues

After sampling at a rate $f_s = 1/T_s$, the score function $\psi(v)$ is approximated by

$$\psi(v) \approx T_s \sum_{k=-K}^K r_1[k - k_0(k; v)] r_2[k], \quad (20)$$

where $r_i[k] \triangleq r_i(kT_s)$, $K = \lfloor T/2T_s \rfloor$ and

$$k_0(k; v) \triangleq \text{round} \left[\frac{\Delta\tau(kT_s; v)}{T_s} \right]. \quad (21)$$

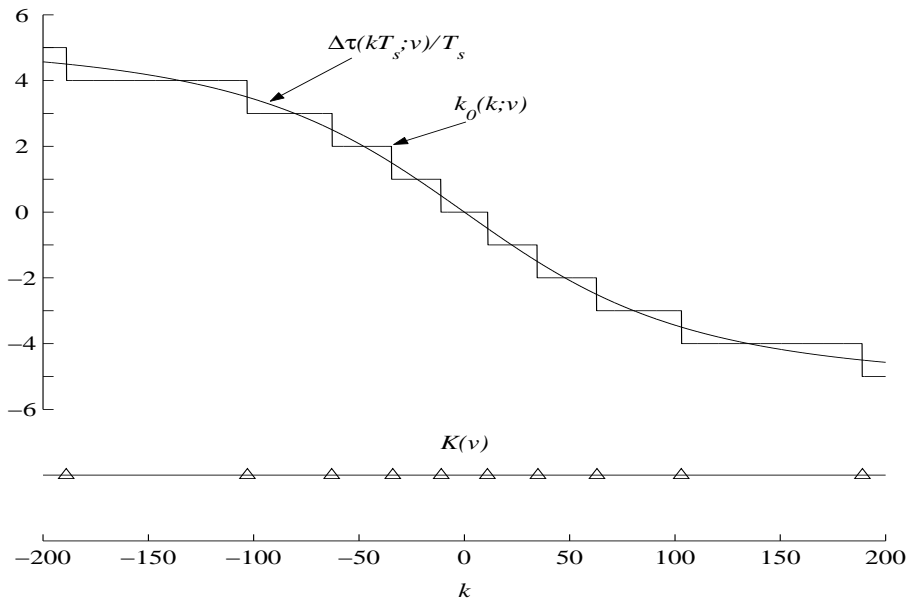


Fig. 3. $\Delta\tau(kT_s; v)$ and $k_0(k; v)$ for $v = 80$ km/h, $D = 13$ m, $2b = 0.9$ m and $T_s = 5$ ms. The constellation of triangles constitutes the set $\mathcal{K}(v)$.

In practice, \hat{v}_0 is obtained by maximizing (20) over a finite set of candidate speeds. Unfortunately, each of these requires full evaluation of the modified crosscorrelation (20) being not possible to reuse computations for any other speed. On the other hand, the implementation of (20) for each candidate v can be done very efficiently in a DSP chip by noting that the operation $k - k_0(k; v)$ in (20) is equivalent to a (slowly) time-varying delay. Since the slope of $\Delta\tau(kT_s; v)/T_s$ is very small, for each v it becomes advantageous to store the set $\mathcal{K}(v)$ of indices k where $k_0(k; v)$ changes (by one), see Fig. 3. Then (20) can be implemented within a DSP in the customary way, with two memory banks (each one associated to a different microphone) and two pointers, with the only difference that every time the pointer to the sequence $r_1[k]$ reaches a value in $\mathcal{K}(v)$, it is increased by one and thus, a sample is skipped.

It is important to remark that in arriving at the approximate ML estimate, the CPA, the sound speed c , and the vehicle range D have been assumed known. Although the actual c and D in a practical implementation will vary around their nominal values, these variations are not expected to be critical. With omnidirectional microphones CPA estimation becomes a nontrivial task, although it is possible to take advantage of the fact that signal power decreases as $1/d^2(t; v_0)$ to derive simple (although suboptimal) algorithms [12]. Joint estimation of CPA and speed following the ML paradigm, as well as analyses of the effect of uncertainty in the values of c and D , constitute an ongoing line of research and are not pursued here. In the remainder we will assume that the CPA, c and D are all known.

IV. ANALYSIS FOR NARROWBAND SOURCE

We now analyze the behavior of the proposed estimator for purely sinusoidal sources. As stated in the Introduction, car-generated waveforms are wideband and consequently do not fit in a tonal model; nevertheless, this simpler case will provide us with meaningful conclusions regarding the various physical parameters. Moreover, Section V will show how these results generalize to the wideband source case.

For the purpose of analysis, vehicle movement during the propagation of its acoustic signature to the sensors must be taken into account. For this, we introduce the following ‘delay error’ term:

$$\xi(t; v_0, v) \triangleq \tau_1(t - \Delta\tau(t; v); v_0) - \tau_1(t; v_0) \quad (22)$$

$$\approx -\frac{2bv_0}{c^2} \left[\sin \alpha(t; v_0) + \frac{b}{D} \cos \alpha(t; v_0) \right] \sin \alpha(t; v), \quad (23)$$

where the last approximation is valid near the true speed value ($|v - v_0|$ small). This term becomes necessary for the analysis because equality does not hold in (15), and the accuracy of the approximation worsens with higher values of the speed.

A. Mean score function

It is shown in the appendix that the mean value of $\psi(v)$ is given by

$$E[\psi(v)] = \int_{-T/2}^{T/2} s_1(t - \Delta\tau(t; v)) s_2(t) dt \quad (24)$$

$$\approx J_0 \left(\frac{\omega b(v - v_0)}{c\sqrt{2v_0v}} \right) \underbrace{\frac{A^2}{2} \int_{-T/2}^{T/2} \frac{\cos[\omega\xi(t; v_0, v)]}{d^2(t)} dt}_{\triangleq Q(v)}, \quad (25)$$

where J_0 is the zeroth-order Bessel function of the first kind. The effect of the ‘delay error’ $\xi(t; v_0, v)$ is perceived from its impact on $Q(v)$. In view of (23), for low frequencies and speeds such that $2\omega bv_0/c^2 \ll 2\pi$, the product $|\omega\xi|$ remains small. In that case, $\cos \omega\xi \approx 1$ and $Q(v)$ is approximately constant and equal to the signal energy per channel $\mathcal{E} \triangleq \int s_i^2(t) dt$, so that

$$E[\psi(v)] \approx \mathcal{E} \cdot J_0 \left(\frac{\omega b(v - v_0)}{c\sqrt{2v_0v}} \right). \quad (26)$$

Fig. 4 plots $E[\psi(v)]$ and (26) for $f = \omega/2\pi = 2$ kHz, $v_0 = 60$ km/h. Several properties of $E[\psi(v)]$ can be derived from those of J_0 : since (26) is maximized for $v = v_0$, for low frequencies and speeds one could expect the bias of the estimate to be small. Also, note that the width of the ‘main lobe’ is proportional to the source speed v_0 , and inversely proportional to the source frequency and microphone spacing. These observations, illustrated in Fig. 5, suggest that the variance of the estimate will increase with increasing source speed (since the main lobe of the score function becomes wider), and decrease as the source frequency and/or sensor spacing

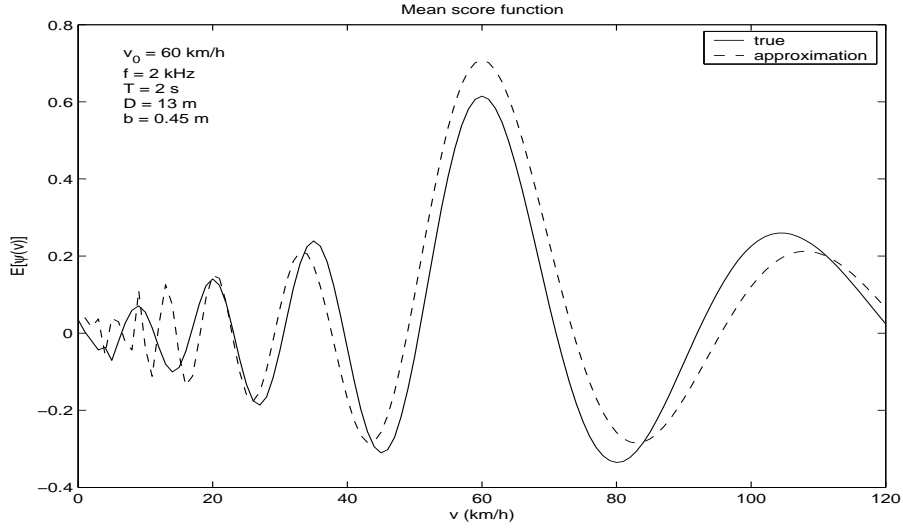


Fig. 4. Plots of $E[\psi(v)]$ and (26) for a 2 kHz narrowband source moving at 60 km/h.

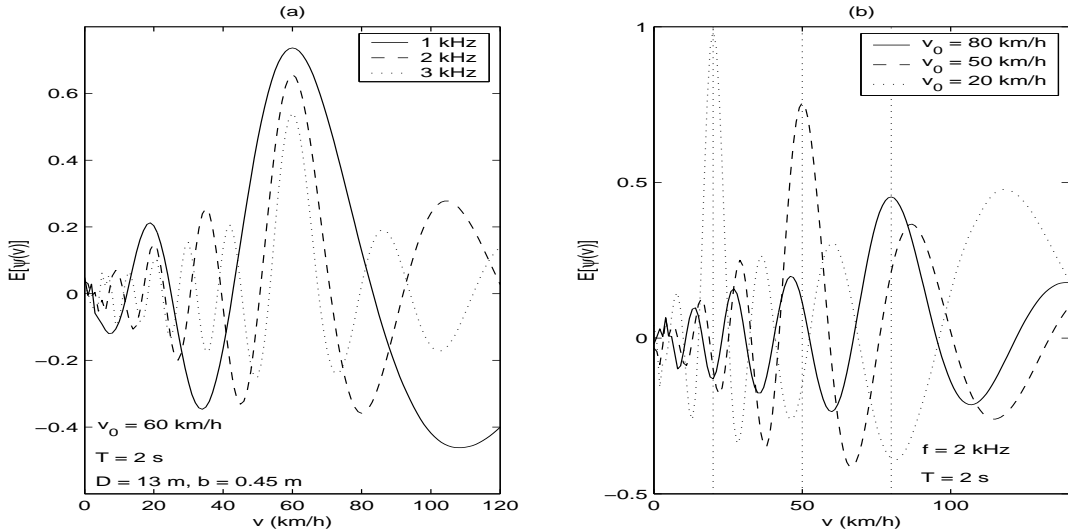


Fig. 5. Plots of $E[\psi(v)]$ for a narrowband source. (a) $v_0 = 60$ km/h and different frequencies; (b) $f = 2$ kHz and different speeds.

increase (since the main lobe becomes narrower). In Fig. 5(b), the peak value of $E[\psi(v)]$ falls with increasing v_0 , as expected since the signal energy \mathcal{E} is inversely proportional to v_0 (For long observation intervals, $\mathcal{E} \approx \pi A^2/2|v_0|D$). The fall with increasing frequency of the peak value of $E[\psi(v)]$ shown in Fig. 5(a), however, is not predicted by (26). Neither is the reduction of the main peak to side peak ratio of $E[\psi(v)]$ as v_0 is increased, as seen in Fig. 5(b).

If $|\omega\xi|$ is not small enough, one cannot regard $Q(v)$ as constant. Lacking an accurate closed-form approximation of $Q(v)$, suffice it to say that in general it does *not* peak at $v = v_0$, and hence the estimate will be biased. The bias will increase

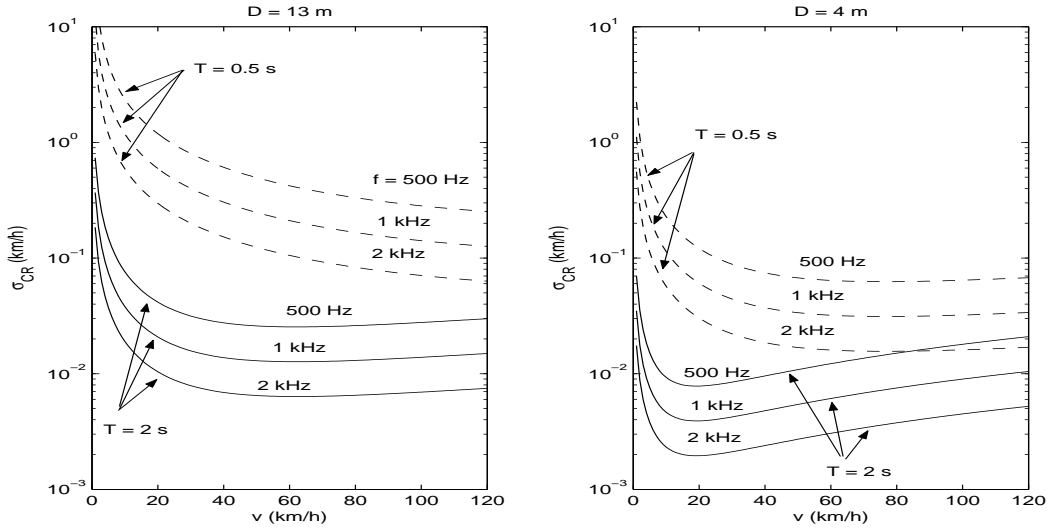


Fig. 6. Cramer-Rao bound for a narrowband source. $A^2/\sigma^2 = 3$ dB, $b = 0.45$ m.

with source frequency and speed. Fortunately, numerical evaluation shows that this bias remains small in the frequency and speed ranges of interest for our application.

B. Cramer-Rao lower bound

The CRB applies to the estimator (18) if the speed and frequency of the source are small enough, since in that case the estimate is unbiased. Also, the CRB is illustrative of the effect of the different parameters involved in the problem.

It must be noted that, if no assumptions on the acoustic waveform $s(t)$ are imposed, it is not possible to derive a generic form of the CRB. In such situation, the best that can be done is to obtain a CRB conditioned on every particular realization of the received signals. Such bound would not be very informative; thus, we derive the CRB assuming that $s(t)$ is known. Clearly, since the proposed estimator is blind, its variance will be much higher than this CRB. (For example, knowledge of the signal bandwidth would allow the designer to bandpass filter the received signals, considerably reducing the noise power and hence the estimate variance.)

Assuming a narrowband source $s(t) = A \sin \omega t$, it is shown in the Appendix that the CRB for arrays with a small ‘aspect ratio’ $b/D \ll 1$ is approximately given by

$$\sigma_{\text{CR}}^2 = \frac{c^2 v^3}{2D\omega^2 f_s G_0(\alpha_0) [A^2/\sigma^2]}, \quad (27)$$

where we have introduced the function

$$G_0(\alpha) \triangleq \tan \alpha + \frac{1}{4} \sin 2\alpha - \frac{3}{2} \alpha \quad (28)$$

$$\approx \frac{1}{5} \tan^5 \alpha, \quad |\alpha| < \frac{\pi}{4}. \quad (29)$$

Fig. 6 shows the variation of σ_{CR} with v for $T = 0.5$ and 2 s, $D = 13$ and 4 m, and different source frequencies.

C. Small-error analysis

Bias and variance analyses can be pursued under a small error approximation, for a narrowband source $s(t) = A \sin \omega t$. The second-order Taylor series expansions around $v = v_0$ corresponding to the terms depending on v in (18) read as

$$s_1(t - \Delta\tau(t; v)) \approx p_0(t) + (v - v_0)p_1(t) + \frac{1}{2}(v - v_0)^2 p_2(t), \quad (30)$$

$$w_1(t - \Delta\tau(t; v)) \approx q_0(t) + (v - v_0)q_1(t) + \frac{1}{2}(v - v_0)^2 q_2(t), \quad (31)$$

where

$$p_k(t) \triangleq \left. \frac{\partial^k s_1(t - \Delta\tau(t; v))}{\partial v^k} \right|_{v=v_0}, \quad q_k(t) \triangleq \left. \frac{\partial^k w_1(t - \Delta\tau(t; v))}{\partial v^k} \right|_{v=v_0}, \quad k = 0, 1, 2. \quad (32)$$

These second-order expansions give a unique solution for the maximization problem (18) in the local vicinity of v_0 at the point for which the derivative vanishes, i.e. $\partial\psi(v)/\partial v|_{\hat{v}_0} = 0$, leading to the following expression for the error

$$v_0 - \hat{v}_0 \approx \frac{\int_{-T/2}^{T/2} [p_1(t) + q_1(t)][s_2(t) + w_2(t)] dt}{\int_{-T/2}^{T/2} [p_2(t) + q_2(t)][s_2(t) + w_2(t)] dt} = \frac{\rho_1 + N_1}{\rho_2 + N_2} \quad (33)$$

where ρ_1, ρ_2 are deterministic values given by

$$\rho_i \triangleq \frac{1}{A^2} \int_{-T/2}^{T/2} p_i(t) s_2(t) dt, \quad i = 1, 2 \quad (34)$$

and N_i are zero-mean Gaussian random variables with variances σ_i^2 , $i = 1, 2$. These are computed in the Appendix, where it is also shown that $\sigma_2 \ll \rho_2$. Hence, one has the following approximations for the bias and variance of the estimation error:

$$E\{v_0 - \hat{v}_0\} \approx \frac{\rho_1}{\rho_2}, \quad \text{var}\{\hat{v}_0 - v_0\} \approx \frac{\sigma_1^2}{\rho_2^2}. \quad (35)$$

Note that the bias ρ_1/ρ_2 that arises is not due to noise (it is independent of the SNR) but to the approximation (15) implicit in the estimation algorithm. In the Appendix, it is shown that $\rho_1, \rho_2, \sigma_1^2$ can be approximated as follows:

$$\rho_1 \approx \frac{\omega b}{D v_0^2 c} \int_{-\alpha_0}^{\alpha_0} \sin \alpha \cos^2 \alpha \left(1 - \frac{v_0}{c} \sin \alpha\right) \sin[\omega \xi(\alpha)] d\alpha, \quad (36)$$

$$\rho_2 \approx -\frac{2\omega^2 b^2}{D v_0^3 c^2} \int_{-\alpha_0}^{\alpha_0} \sin^2 \alpha \cos^4 \alpha \cos[\omega \xi(\alpha)] d\alpha, \quad (37)$$

$$\sigma_1^2 \approx \frac{\pi^2}{3} \frac{f_s b^2 D}{v_0^3 c^2 [A^2/\sigma^2]^2} [\alpha_0 - \frac{1}{4} \sin 4\alpha_0] \quad (38)$$

where $\xi(\alpha)$ denotes the delay error term (22) for $v = v_0$ in terms of the angle α :

$$\xi(\alpha) = -\frac{2bv_0}{c^2} \sin \alpha \left[\sin \alpha + \frac{b}{D} \cos \alpha \right]. \quad (39)$$

It is not possible to find closed-form expressions for ρ_i due to the presence of this term in (36)-(37). However, if the product $\omega\xi$ remains small enough in the observation window, then $\sin \omega\xi \approx \omega\xi$, $\cos \omega\xi \approx 1 - \frac{1}{2}\omega^2\xi^2$. Hence, after integrating,

$$\frac{\rho_1}{\rho_2} \approx \frac{v_0^3}{c^2} \frac{1 - \frac{2bc}{Dv_0} \frac{\sin \alpha_0}{\alpha_0}}{1 - \frac{3}{8} \left(\frac{\omega b v_0}{c^2} \right)^2}, \quad (40)$$

$$\frac{\sigma_1^2}{\rho_2^2} \approx \frac{16\pi^2}{3} \frac{f_s D^3 v_0^3 c^2}{\omega^4 b^2 [A^2/\sigma^2]^2} \frac{\frac{1}{\alpha_0} \left(1 - \frac{\sin 4\alpha_0}{4\alpha_0} \right)}{\left[1 - \frac{3}{8} \left(\frac{\omega b v_0}{c^2} \right)^2 \right]^2}. \quad (41)$$

Observe that as ωv_0 approaches the value $\eta \triangleq \frac{c^2}{b} \sqrt{\frac{8}{3}}$, these expressions tend to infinity. Therefore for $\omega v_0 \rightarrow \eta$, the small error assumption on which the analysis is based ceases to be valid. In the small ωv_0 region, the bias is not very sensitive to the source frequency, while the variance falls as $1/\omega^4$. If α_0 is assumed constant (e.g. for large observation windows), then both bias and variance increase as v_0^3 .

V. BROADBAND SIGNALS

Assume now that $s(t)$ is a deterministic broadband signal with Fourier transform $S(\omega)$. It is shown in the Appendix that for low values of the speed v_0 , the mean score function takes the following form:

$$E[\psi(v)] \approx \frac{\alpha_0}{2\pi^2 D v_0 T} \int_{-\infty}^{\infty} |S(\omega)|^2 J_0 \left(\frac{\omega b (v - v_0)}{c \sqrt{2v_0 v}} \right) d\omega. \quad (42)$$

This expression is also valid if $s(t)$ is regarded as a wide sense stationary random process with power spectral density $|S(\omega)|^2$. Hence, for broadband signals the mean score function approximately reduces to the superposition of those corresponding to each frequency as computed in Section IV-A, weighted by the power spectrum of the signal. Given the dependence with frequency of the variance of the estimate found in the preceding sections, this suggests that in a practical implementation higher frequency components of the received signals should be enhanced with respect to lower ones. This will be verified by the experiments presented in Section VII.

The CRB in the broadband case, again for $b/D \ll 1$, is derived in the Appendix:

$$\sigma_{\text{CR}}^2 = \frac{\pi^2 T c^2 v^3 \sigma^2}{D f_s G_0(\alpha_0) \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega}. \quad (43)$$

It is seen that σ_{CR}^2 is inversely proportional to the power of the derivative of the source signal. That is, the CRB will be lower for acoustic signals with a highpass spectrum. The behavior of σ_{CR}^2 with respect to the remaining parameters (v , b , D , T) is the same as that in the narrowband case.

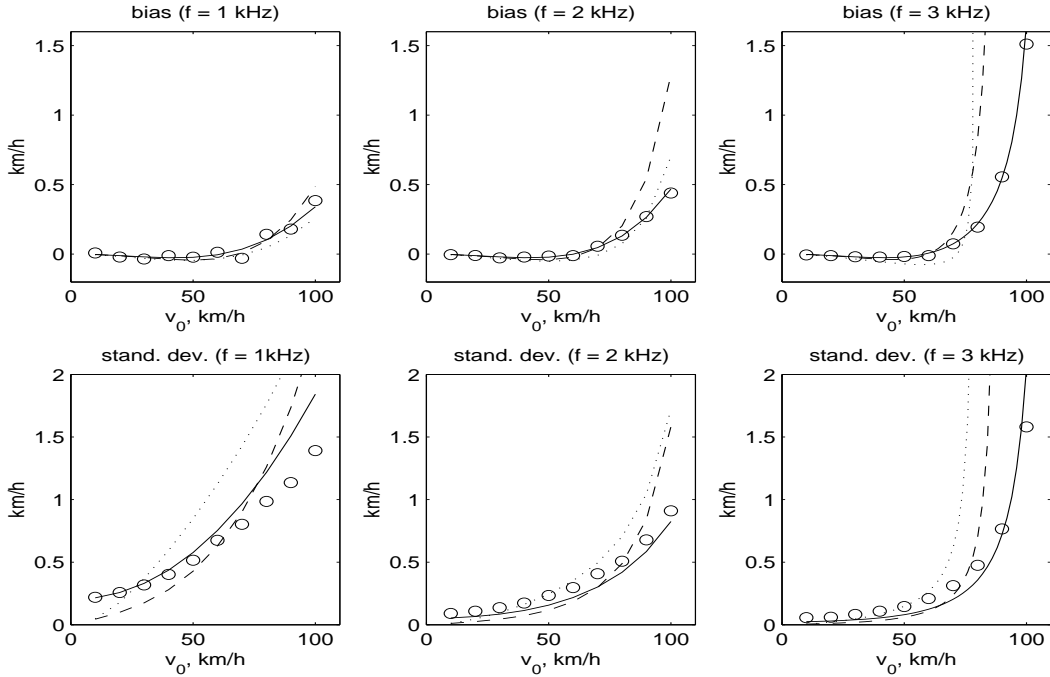


Fig. 7. Bias and standard deviation of \hat{v}_0 : theoretical (lines) and estimated (circles).
 $f_s = 40$ kHz, SNR = 3 dB, $T = 2$ s, $D = 13$ m, $b = 0.45$ m.

VI. SIMULATION RESULTS

In order to test the performance of the estimation algorithm, several computer experiments were carried out. For all of them we took $c = 340$ m/s, and for each data point, results were averaged over 1000 independent Monte Carlo runs.

First we considered narrowband sources $s(t) = A \sin \omega t$, and array dimensions $D = 13$ m, $b = 0.45$ m. With $A_0 \triangleq A/D$ the received signal amplitude at the CPA, we define the signal to noise ratio per channel as

$$\text{SNR} = \frac{A_0^2}{\sigma^2}. \quad (44)$$

In the first experiment we set $f_s = 40$ kHz, $T = 2$ s and SNR = 3 dB. Source speed and frequency varied from 10 to 100 km/h and from 1 to 3 kHz respectively. Fig. 7 shows the bias and standard deviation of the estimate \hat{v}_0 from the simulations (circles), as well as the values predicted by the analysis in Section IV-C using several degrees of accuracy in the approximations for ρ_i . The dotted line values were directly obtained from (40)-(41). For the dashed line values, we numerically integrated (36)-(37). Finally, the solid line was obtained without using the far-field approximation implicit in (36)-(37). This was done by numerical integration of (68)-(69) in the Appendix, using the exact time domain expressions of the integrands [i.e. without using the approximations in (63)-(65)]. The critical speed values η/ω are 240, 120 and 80 km/h for frequencies 1, 2 and 3 kHz respectively. The far field

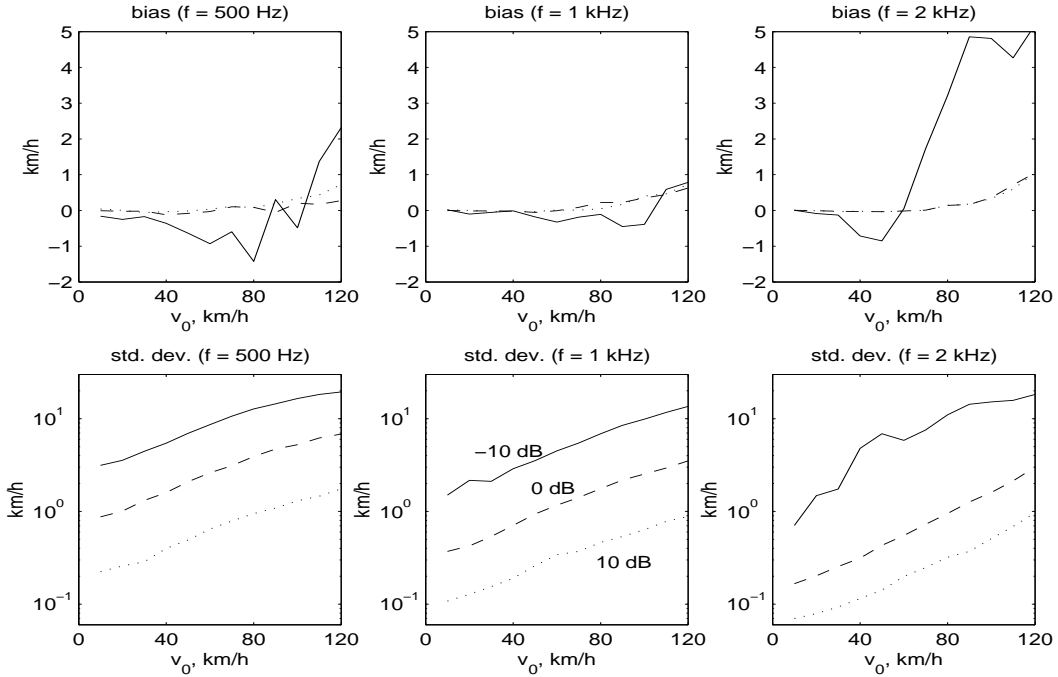


Fig. 8. Bias and standard deviation of \hat{v}_0 . SNR = -10 dB (solid), 0 dB (dashed) and 10 dB (dotted). $f_s = 10$ kHz, $T = 2$ s, $D = 13$ m, $b = 0.45$ m.

approximations show good agreement with the simulations for small v_0 , losing accuracy for higher speeds but still capturing the general trend of the estimate (bias and variance increase sharply near the critical values).

It is seen that for low speeds ($v_0 < 60$ km/h) the bias remains very small for all frequencies and the variance steadily decreases with frequency. For $v_0 > 60$ km/h the bias becomes noticeable, increasing with frequency, while there seems to be an optimal, speed dependent frequency value which minimizes the estimation variance.

In the second experiment, the sampling frequency was reduced to $f_s = 10$ kHz, while keeping $T = 2$ s. Fig. 8 shows the statistics of the estimate \hat{v}_0 , for different frequencies and SNRs. With this reduced sampling rate, the variance of the estimate presents an additional component due to the rounding operation (21) in the computation of the score function. This effect was not considered in the analysis of Section IV-C, so that the predicted variance values tend to be smaller than those obtained from the simulations for high SNR (in which case the rounding and noise components of the variance become comparable). The data reveals that the variance is inversely proportional to the SNR and to ω^2 . The behavior of the bias curves for -10 dB SNR is believed to be a result of insufficient averaging and/or the aforementioned rounding effects (recall that the bias is expected to be independent of the noise level). In any case, the bias remains within a few km/h.

The effect of the observation window T was also studied. Fig. 9 shows the standard deviation of \hat{v}_0 for $f_s = 10$ kHz, SNR = 0 dB and different values of T and ω . (The

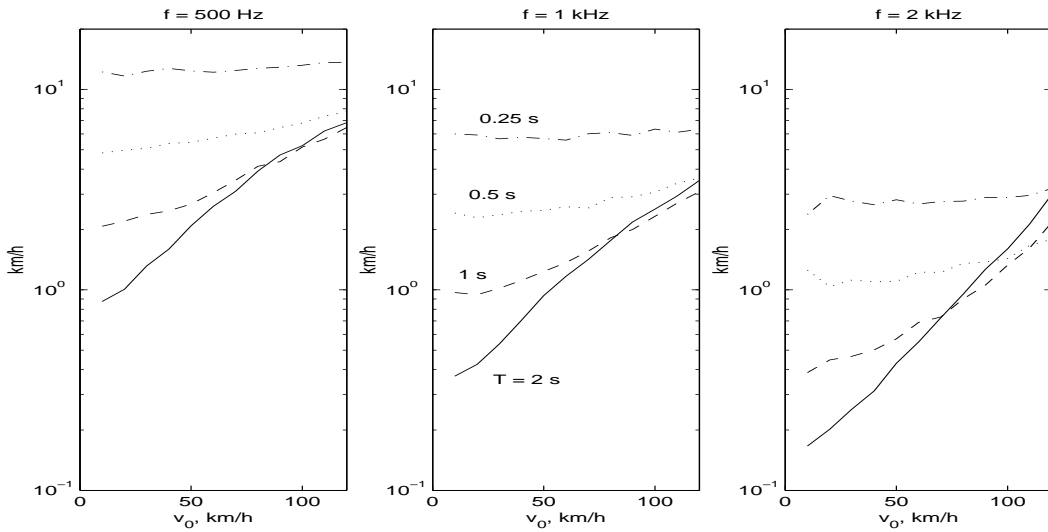


Fig. 9. Standard deviation of \hat{v}_0 . $T = 2$ s (solid), 1 s (dashed), 0.5 s (dotted) and 0.25 s (dashdot). $f_s = 10$ kHz, SNR = 0 dB, $D = 13$ m, $b = 0.45$ m.

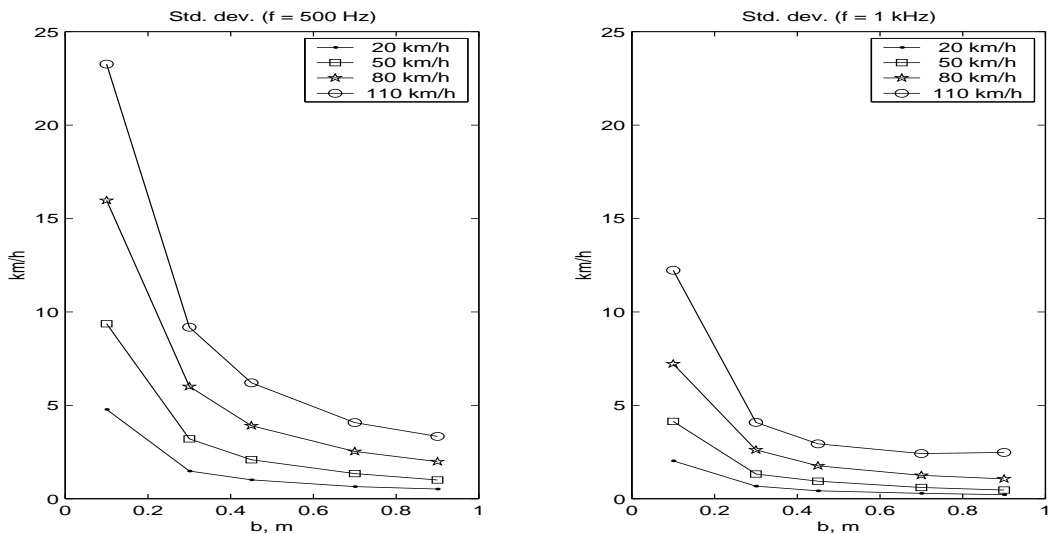


Fig. 10. Standard deviation of \hat{v}_0 as a function of the sensor separation. SNR = 0 dB, $T = 2$ s, $f_s = 10$ kHz, $D = 13$ m.

bias, not shown, remained within ± 2 km/h). Reducing T has a greater impact for low speeds, as expected since in that case a significant part of the signal energy is likely to lie outside $|t| < T/2$. However, it is also seen that, for higher speeds, increasing T beyond a certain speed-dependent value T_v has a negative impact on performance. If $T < T_v$, performance quickly degrades; for $T > T_v$ the variance also increases although not as sharply. Such ‘optimal window size’ effect is thought to be due to the underlying approximation (15).

The influence of sensor separation can be seen in Fig. 10. We fixed $D = 13$ while

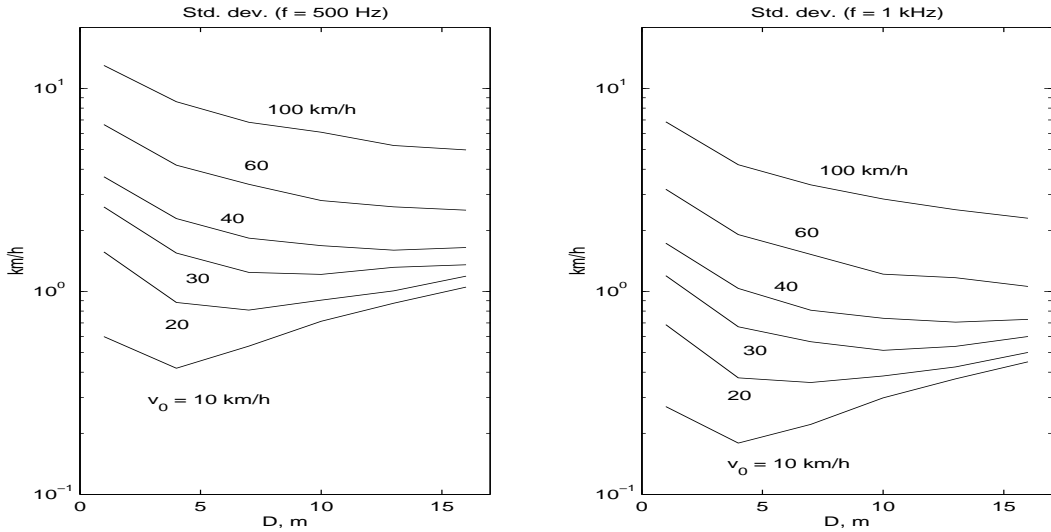


Fig. 11. Standard deviation of \hat{v}_0 as a function of array to road distance. SNR = 0 dB, $T = 2$ s, $f_s = 10$ kHz, $b = 0.45$ m.

varying b from 0.1 to 0.9 m, taking $T = 2$ s, $f_s = 10$ kHz and SNR = 0 dB. Clearly, placing the sensors too close to each other considerably worsens the performance, while the improvement is marginal if b is increased beyond 0.6 m. This is fortunate since achieving large separations may be problematic in practical settings.

Next, we fixed $b = 0.45$ m and varied the array to road distance D , keeping $T = 2$ s, $f_s = 10$ kHz, and SNR = 0 dB. It is observed in Fig. 11 that the variance initially falls as D is increased until a minimum is reached, after which a slow increase takes place. The location of this minimum depends on the source speed, but not on its frequency. Note that with the definition (44), varying D does not result in a change in the effective SNR, and therefore the results truly reflect the effect of the geometry. (On the other hand, if the source amplitude A is assumed constant, then the effective SNR should decrease as $1/D^2$ as the separation from the road is increased).

Simulations with wideband sources were also run. Samples of $s(t)$ were generated as independent Gaussian random variables with zero mean and variance D^2 , so that the instantaneous received power per channel at the CPA is normalized to unity. In this way, the SNR per channel is defined as $\text{SNR} = 1/\sigma^2$. The delayed values required to generate the synthetic received signals were computed via interpolation.

For comparison purposes we also tested an indirect approach based on DTD estimation, as in [5]. The observation window was divided in disjoint, consecutive segments of length M samples over which the received signals were crosscorrelated. By picking the delay at which the maximum of this crosscorrelation takes place, an estimate $\Delta\hat{\tau}(t)$ of $\Delta\tau(t; v_0)$ is obtained. Then the speed estimate is chosen in order

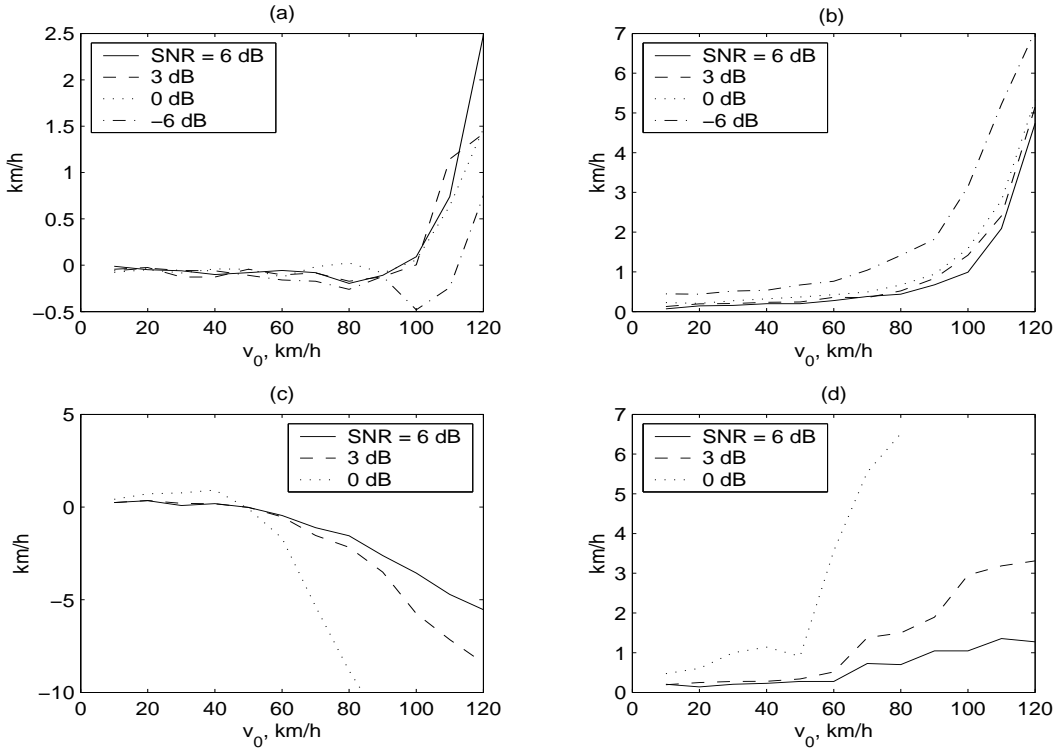


Fig. 12. Results for a wideband random source. $T = 2$ s, $f_s = 10$ kHz, $D = 13$ m, $b = 0.45$ m. (a) Proposed approach, bias; (b) Proposed approach, standard deviation; (c) Indirect approach, bias; (d) Indirect approach, standard deviation.

to minimize the following weighted least squares (WLS) cost:

$$C(v) \triangleq \sum_{n=-N}^N \frac{[\Delta\hat{\tau}(nMT_s) - \Delta\tau(nMT_s; v)]^2}{d^4(nMT_s; v)},$$

where $N \triangleq \lfloor T/2MT_s \rfloor$. (Since the shape of $\Delta\tau$ is more sensitive to speed variations near the CPA, a weighting factor of the form $1/d^p(t; v)$ seems reasonable. The choice $p = 4$ was found to result in best performance.)

Fig. 12 shows the performance of both approaches using an array with $D = 13$ m, $b = 0.45$ m, processing parameters $f_s = 10$ kHz, $T = 2$ s, and $M = 128$ samples. Analogous results after reducing T to 0.5 s are shown in Fig. 13. The estimate $\Delta\hat{\tau}(t)$ in the indirect approach was smoothed by a 7th-order median filter before WLS minimization. Both algorithms are given the exact CPA location. The bias of the proposed method remains very small for low speeds, as in the narrowband case. The variance increases with speed and decreases with the SNR, as expected. These trends are also observed in the indirect approach, although this estimate seems to be very sensitive to the additive noise with respect to both bias and variance. The proposed method is much more robust in this respect. This is because it uses the whole available signal at once in the estimation process, therefore providing a much

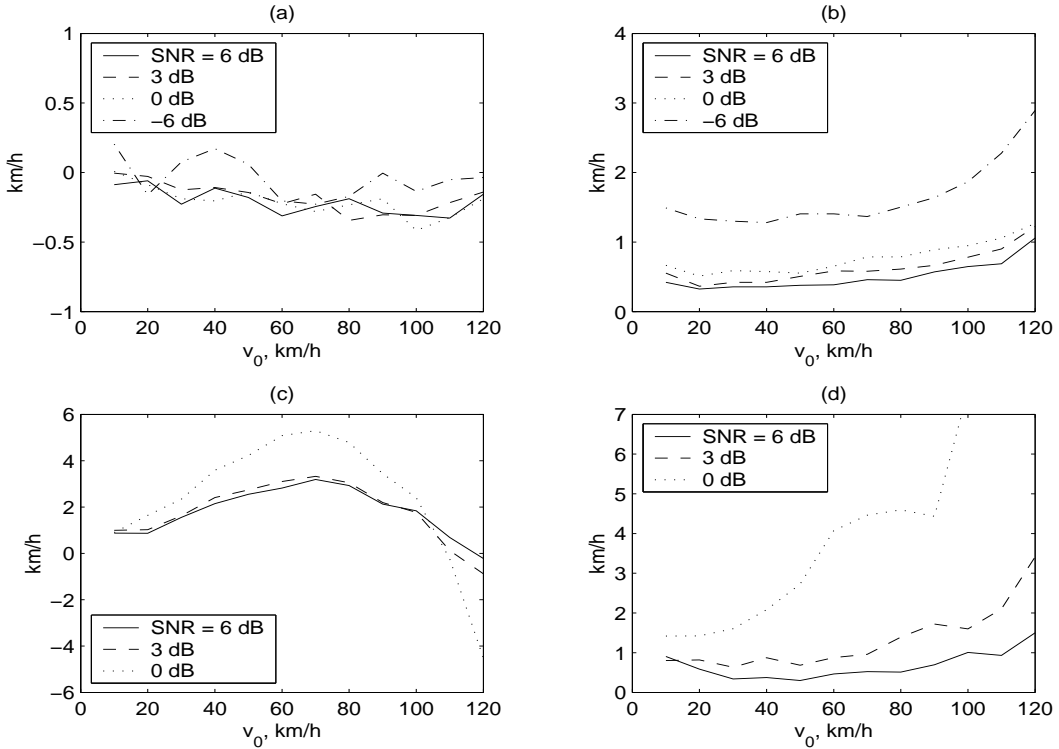


Fig. 13. Results for a wideband random source. $T = 0.5$ s, $f_s = 10$ kHz, $D = 13$ m, $b = 0.45$ m. (a) Proposed approach, bias; (b) Proposed approach, standard deviation; (c) Indirect approach, bias; (d) Indirect approach, standard deviation.

more effective noise averaging. Decreasing T is seen to have a beneficial effect in the bias of both estimates, while it does not substantially affect the variance behavior of the indirect approach. As in the narrowband case, the variance of the proposed estimate increases for low speeds when T is reduced but decreases for high speeds (this effect is seen to become more pronounced with wideband signals).

VII. EXPERIMENTAL RESULTS

We have tested the estimation algorithm on acoustic signals recorded from real traffic data. Two omnidirectional microphones were set up as in Fig. 1, separated $2b = 0.9$ m and mounted on a 6.5 m pole whose base was 13 and 16 m from the center of the two road lanes, yielding $D \approx 14.5$ m for the close lane and 17.3 m for the far one. The sampling rate was $f_s = 14.7$ kHz, and the signals were recorded with 16 bit precision. A videocamera was also mounted in order to have an alternative means to determine the parameters of the traffic flow. The signals are available at <http://www.gts.tsc.uvigo.es/~valcarce/traffic.html>.

Fig. 14 shows the waveform and the spectrogram of the signal produced by a bus traveling along the close lane at a speed of approximately 40 km/h, as determined from the video recording. Near $t = 0.86$, 2.36 and 3.36 s, and for unknown reasons,

the recording equipment zeroed out the output signals during approximately 20 ms. However, the estimator is expected to be robust to such time-localized effects since it is based on (modified) crosscorrelations over the whole observation window.

We computed the function $\psi(v)$ using different highpass filtered versions of the recorded signals. The CPA was taken as $t \approx 2.21$ s, determined from the position of the peak of the short-time autocorrelation of the signals using a 2048-sample (0.14 s) sliding window. Fig. 15 shows the results obtained with observation intervals of $T = 1$ and 2 s, using highpass filters with cutoff frequencies $f_c = 0$ (no filtering), 60, 125 and 250 Hz. For each case, $\psi(v)$ was computed for a range of speeds ($v < 0$ corresponding to a vehicle approaching the array from the right, in the notation of Fig. 1) and normalized by its peak value. The estimated speed was $\hat{v}_0 = 41$ km/h. It can be observed that highpass filtering becomes necessary in order to ‘sharpen’ the lobe associated to the true speed v_0 .

In a second experiment we used the signals from a compact car moving along the close lane at 50 km/h according to the video data. The waveform and spectrogram of $r_1(t)$ are shown in Fig. 16. The corresponding score functions are depicted in Fig. 17 for a CPA of $t = 1.55$ s. The estimate obtained with $T = 2$ s is $\hat{v}_0 = 53$ km/h. The beneficial effect of removing low frequency content is noted again.

Fig. 18 shows the waveform and spectrogram corresponding to a sedan traveling at -80 km/h along the far lane. CPA was taken at $t = 3.75$ s. The score functions are depicted in Fig. 19. The estimate using $T = 2$ s is $\hat{v}_0 = -72$ km/h. Conditions were quite windy (notice the gust toward the end of the record), but fortunately it was found that in most cases the effect of wind is concentrated in the low frequency region and can be effectively suppressed by highpass filtering.

We must mention that, although we attempted to use the DTD estimation based indirect approach with these recorded signals, in all of the cases and for a variety of the crosscorrelation window size, the differential time delay estimate $\Delta\hat{\tau}$ exhibited a highly irregular behavior, not resembling the expected S-shape of Fig. 2. This could be due to the sensitivity of short-time DTD estimation to noise as well as time localized (i.e. short duration) disturbances present in the records. Under these conditions, this method was unable to produce a usable speed estimate: the use of directional microphones as in [5] may be required for this approach to work.

VIII. CONCLUSIONS

The proposed approximate ML estimate is easily implemented, and its application is quite general. Its main advantage is the ability to estimate car speed directly without requiring a model for the emitted signal. Thus, intermediate delay estimation steps and source modeling, which may be problematic, are avoided altogether. The estimate is reasonably robust to noise and time localized disturbances since the crosscorrelations involved are computed over the whole observation window. It is expected as well to be robust to small uncertainties in the values of parameters such

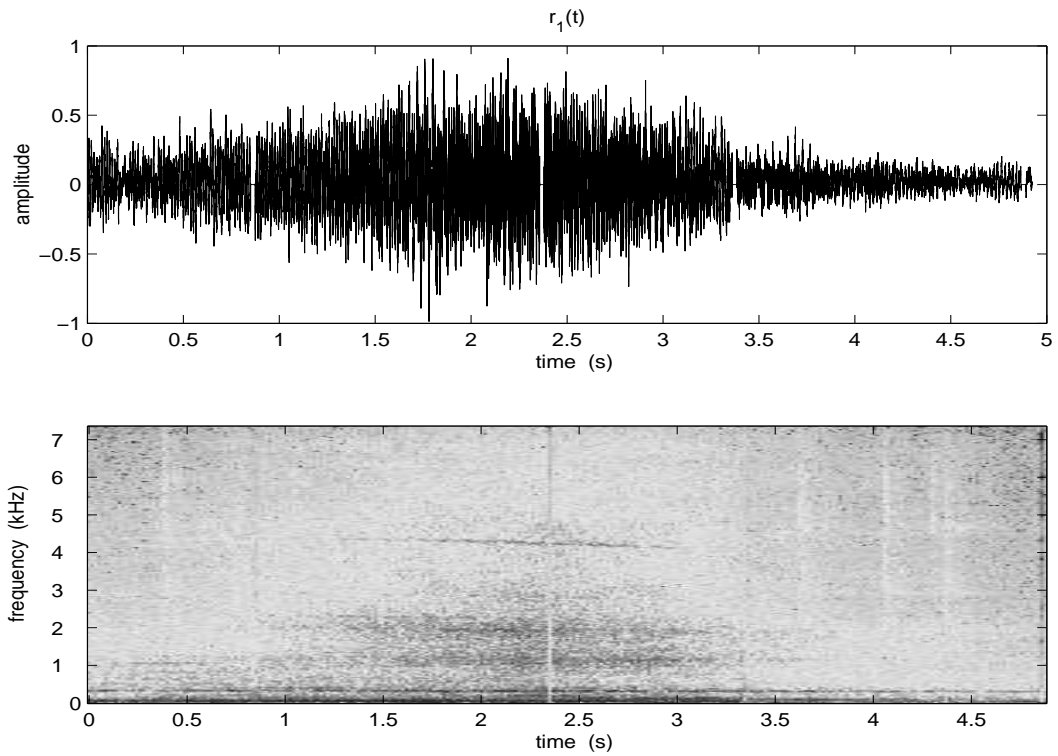


Fig. 14. Waveform and spectrogram of the acoustic signature of a passing bus.

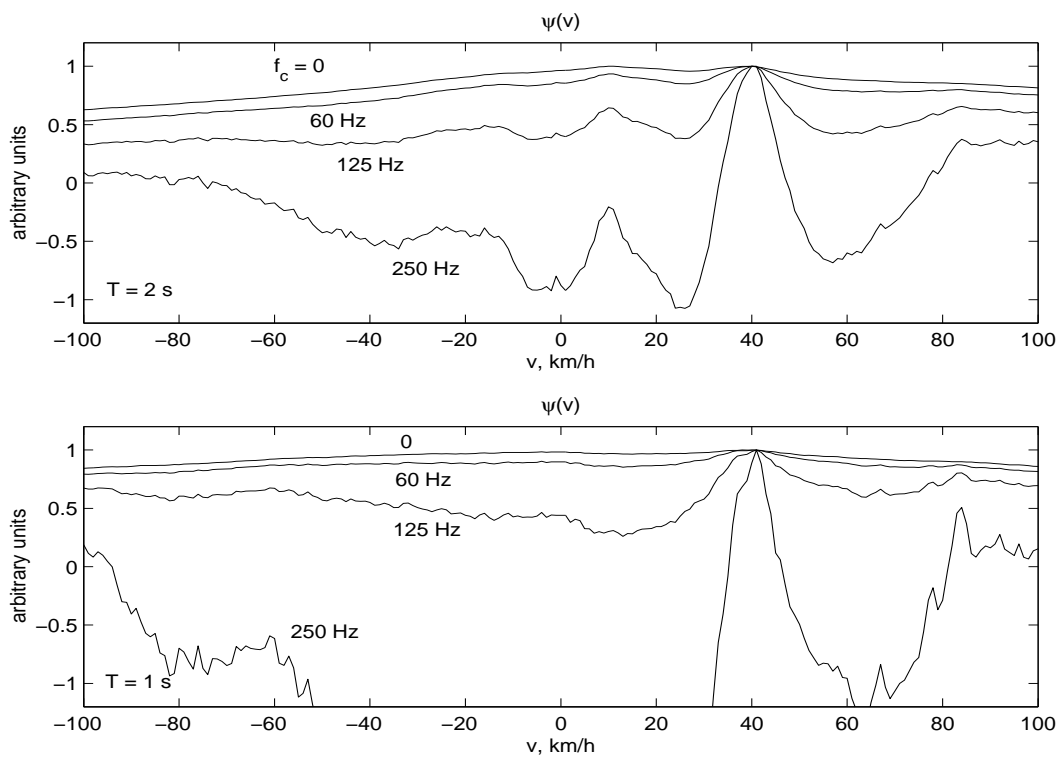


Fig. 15. The score function computed for a passing bus. $T = 2$ s (top) and 1 s (bottom).

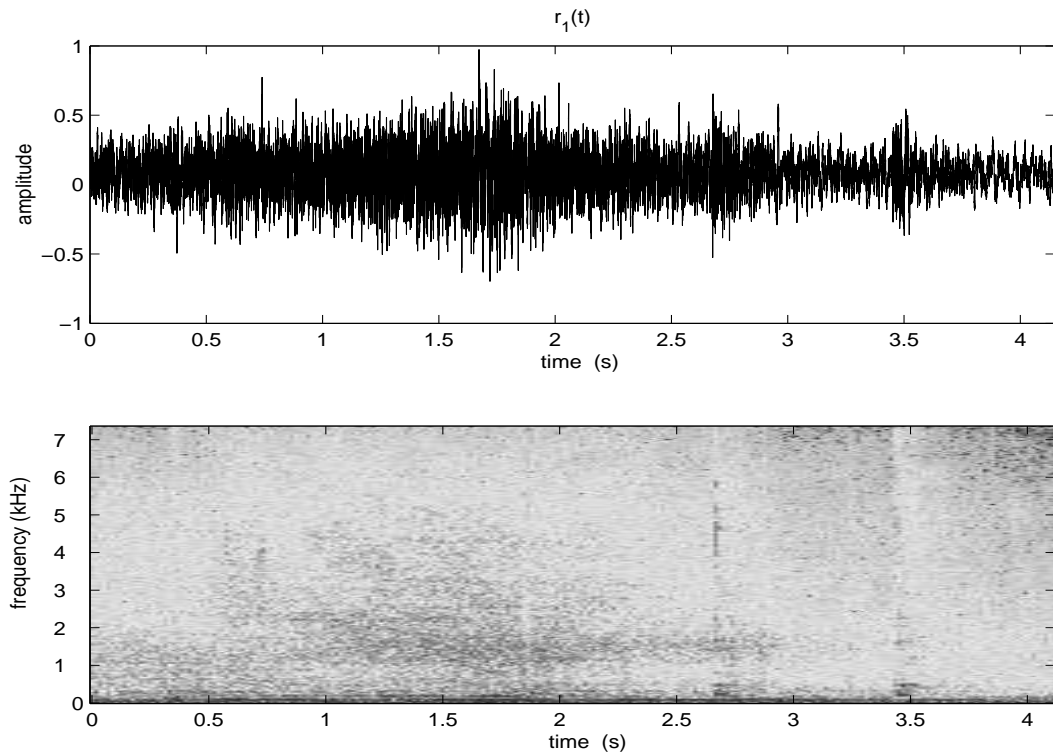


Fig. 16. Waveform and spectrogram of the acoustic signature of a passing car.

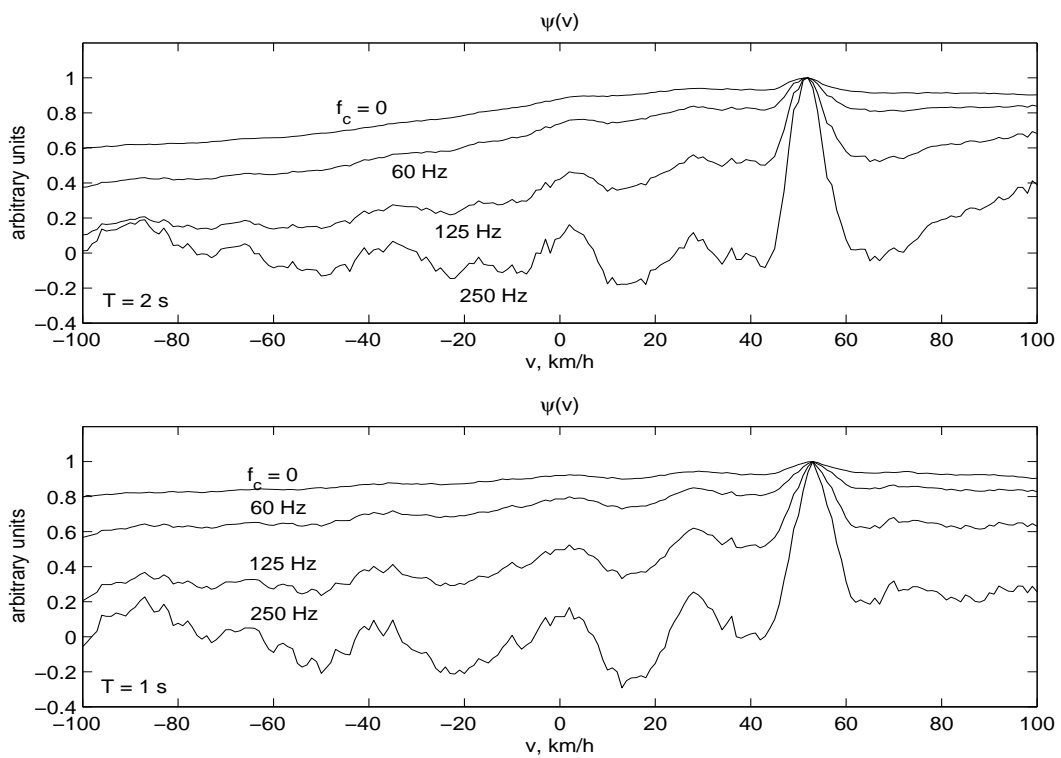


Fig. 17. The score function computed for a passing car. $T = 2$ s (top) and 1 s (bottom).

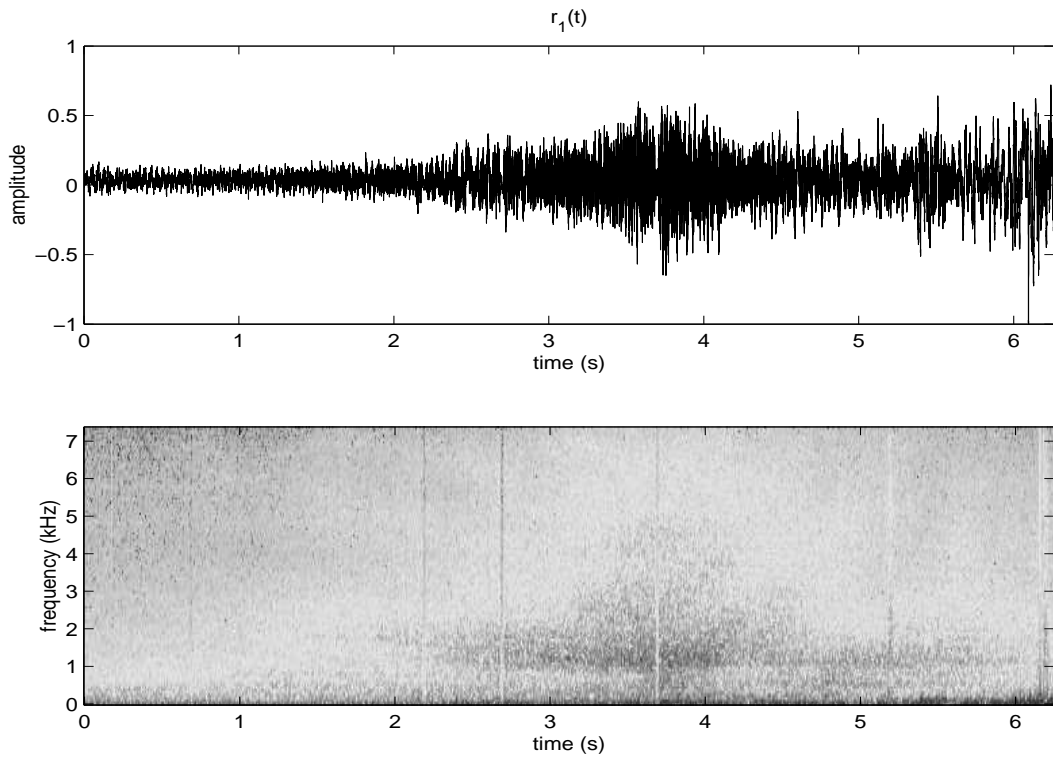


Fig. 18. Waveform and spectrogram of the acoustic signature of a passing car.

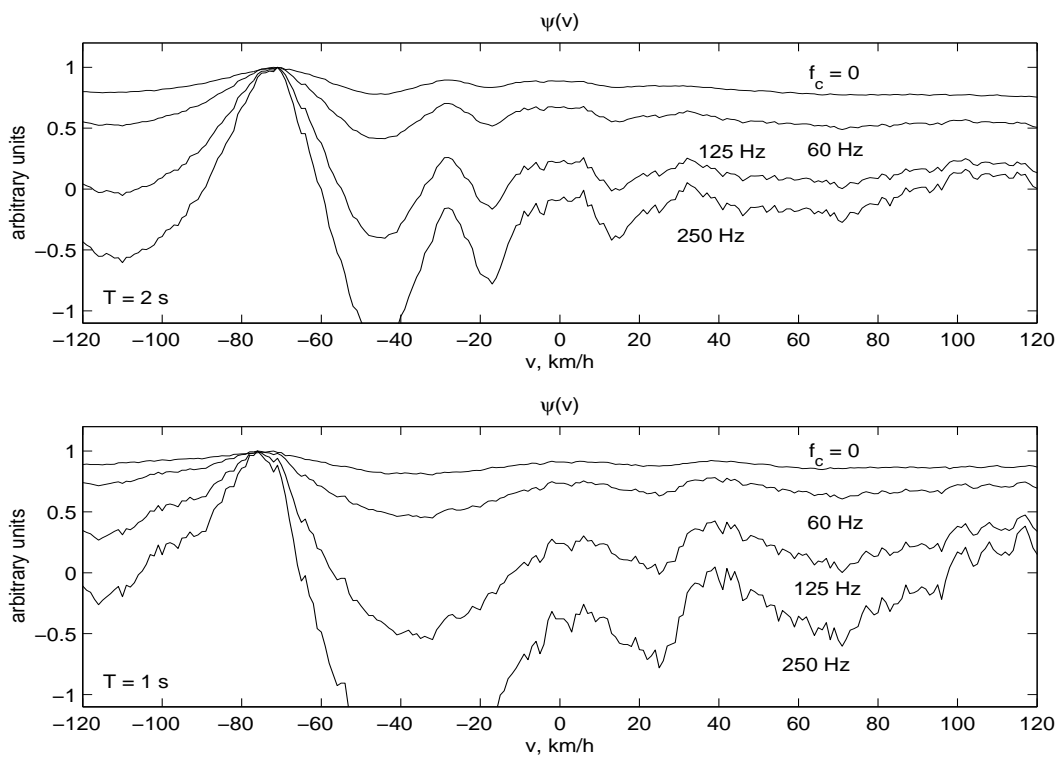


Fig. 19. The score function computed for a passing car. $T = 2$ s (top) and 1 s (bottom).

as the speed of sound c and the array to road distance D .

Our analysis reveals the impact of the system parameters in the accuracy of the estimate. Perhaps the most dramatic one is the harmful effect of low frequency signal components, which has been confirmed by the experiments. Ongoing work will try to determine the most adequate frequency band, taking into account the spectral characteristics of road vehicles.

The presence of multiple vehicles within the observation window should be resolvable as long as their corresponding CPAs are sufficiently apart in time. In practice, the location of the CPA has to be estimated. This problem is currently being investigated, as well as the robustness of the proposed estimate to uncertainties in CPA determination. More extensive field tests of the algorithm are also under way. Other open issues are the determination of the time window and sampling frequency as a trade-off between complexity and performance.

ACKNOWLEDGEMENT

The authors would like to thank Serafín A. Martínez for his help in collecting the acoustic data.

APPENDIX A: MEAN SCORE FUNCTION IN THE NARROWBAND CASE

With $s_i(t)$ given by (4), one finds that the time-shifted value of $s_1(t)$ in (24) is

$$s_1(t - \Delta\tau(t; v)) = \frac{s(t - \Delta\tau(t; v) - \tau_1(t - \Delta\tau(t; v); v_0))}{d(t - \Delta\tau(t; v); v_0)}. \quad (45)$$

In the denominator of (45), we can make the approximation $d(t - \Delta\tau(t; v); v_0) \approx d(t; v_0)$. However, we must be more accurate with the analogous term appearing in the argument of $s(\cdot)$. For $s(t) = A \sin \omega t$, one has

$$s_1(t - \Delta\tau(t; v)) \approx A \frac{\sin[\omega(t - \Delta\tau(t; v) - \tau_1(t; v_0) - \xi(t; v_0, v))]}{d(t; v_0)} \quad (46)$$

with $\xi(t; v_0, v)$ defined in (22). Therefore, the product of (46) with $s_2(t)$ becomes

$$s_1(t - \Delta\tau(t; v))s_2(t) \approx \frac{A^2/2}{d^2(t; v_0)} \times \{ \cos[\omega(\Delta^2\tau(t; v_0, v) - \xi(t; v_0, v))] - \cos[\omega(2t - \tau_+(t; v_0, v) - \xi(t; v_0, v))] \}, \quad (47)$$

where

$$\Delta^2\tau(t; v_0, v) \triangleq \Delta\tau(t; v_0) - \Delta\tau(t; v), \quad (48)$$

$$\tau_+(t; v_0, v) \triangleq \Delta\tau(t; v) - [\tau_1(t; v_0) + \tau_2(t; v_0)]. \quad (49)$$

When integrating (47), the contribution of the ‘double-frequency’ term is small compared to that of the second term in the right hand side of (47), so it can be neglected. On the other hand,

$$\begin{aligned} \cos[\omega(\Delta^2\tau(t; v_0, v) - \xi(t; v_0, v))] &= \cos[\omega \Delta^2\tau(t; v_0, v)] \cos[\omega \xi(t; v_0, v)] \\ &\quad + \sin[\omega \Delta^2\tau(t; v_0, v)] \sin[\omega \xi(t; v_0, v)]. \end{aligned} \quad (50)$$

At this point we need an approximation for the terms involving $\Delta^2\tau(t; v_0, v)$. Note that in view of the ‘far field’ approximation (7), one has

$$\Delta^2\tau(t; v_0, v) \approx \frac{2b}{c}[\sin\alpha(t; v) - \sin\alpha(t; v_0)]. \quad (51)$$

Although (51) is accurate, it still is too complicated for our purposes. Nevertheless, by visual inspection of $\Delta^2\tau$, the following approximation seems well suited:

$$\Delta^2\tau(t; v_0, v) \approx R \sin[2 \operatorname{atan}(zt)]. \quad (52)$$

The values of R and z can be selected by imposing that the two sides of (52) have the same slope at $t = 0$, and that they peak at the same time instants. The first condition reads as $Rz = b(v - v_0)/Dc$. On the other hand, after some algebra one finds that the extrema of the right hand side of (51) are approximately located at $t \approx \pm D/\sqrt{2v_0v}$, while those of (52) take place at $t = \pm 1/z$. Hence

$$R = \frac{b(v - v_0)}{c\sqrt{2v_0v}}, \quad z = \frac{\sqrt{2v_0v}}{D}. \quad (53)$$

The advantage of (52) resides in that it allows to expand the sine and cosine terms in (50), in view of the Fourier series

$$\cos(r \sin x) = \sum_{k=-\infty}^{\infty} J_k(r) \cos(kx), \quad \sin(r \sin x) = \sum_{k=-\infty}^{\infty} J_k(r) \sin(kx), \quad (54)$$

(see e.g. [1]), where J_k is the k th order Bessel function of the first kind. Hence, after neglecting the double-frequency term, we have

$$\begin{aligned} E[\psi(v)] &\approx \frac{A^2}{2} \sum_{k=-\infty}^{\infty} J_k(\omega R) \int_{-T/2}^{T/2} \frac{\cos[2k \operatorname{atan}(zt)]}{d^2(t; v_0)} \cos[\omega\xi(t; v_0, v)] dt \\ &\quad + \frac{A^2}{2} \sum_{k=-\infty}^{\infty} J_k(\omega R) \int_{-T/2}^{T/2} \frac{\sin[2k \operatorname{atan}(zt)]}{d^2(t; v_0)} \sin[\omega\xi(t; v_0, v)] dt. \end{aligned} \quad (55)$$

This sum is dominated by the $k = 0$ term, so that (25) follows. ■

APPENDIX B: DERIVATION OF THE CRB

Since the pdf of the observations conditioned on $s(t)$ is Gaussian, the CRB for the estimation of the source speed v is given by

$$\sigma_{\text{CR}}^2 = \frac{\sigma^2}{\left\| \frac{\partial \mathbf{s}_1(v)}{\partial v} \right\|^2 + \left\| \frac{\partial \mathbf{s}_2(v)}{\partial v} \right\|^2}, \quad (56)$$

where $\mathbf{s}_i(v)$ are the vectors of samples of the noiseless signals $s_i(t)$ impinging on the microphones. With $s_i(t)$ defined in (4), and with $s'(t) \triangleq \partial s(t)/\partial t$, one has

$$\frac{\partial s_i(t)}{\partial v} = -\frac{(vt \pm b)t}{c^2\tau_i^2(t; v)} \left[\frac{1}{c} s'(t - \tau_i(t; v)) + \frac{s(t - \tau_i(t; v))}{c\tau_i(t; v)} \right]. \quad (57)$$

Let $S(\omega)$ be the spectrum of $s(t)$. Then (57) can be written in terms of $S(\omega)$ as

$$\begin{aligned}\frac{\partial s_i(t)}{\partial v} &= -\frac{(vt \pm b)t}{d_i^2(t; v)} \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) \left[\frac{1}{d_i(t; v)} + j \frac{\omega}{c} \right] e^{j\omega[t - \tau_i(t; v)]} d\omega \\ &\approx -\frac{(vt \pm b)t}{d_i^2(t; v)} \frac{j}{2\pi c} \int_{-\infty}^{\infty} \omega S(\omega) e^{j\omega[t - \tau_i(t; v)]} d\omega\end{aligned}\quad (58)$$

where the last approximation follows from $1/d_i(t; v) \leq 1/D \ll \omega/c$. Therefore,

$$\begin{aligned}\left\| \frac{\partial \mathbf{s}_i(v)}{\partial v} \right\|^2 &\approx f_s \int_{-T/2}^{T/2} \left[\frac{\partial s_i(t)}{\partial v} \right]^2 dt \\ &\approx \frac{f_s}{(2\pi c)^2} \iint_{-\infty}^{\infty} \omega_1 \omega_2 S(\omega_1) S^*(\omega_2) \Gamma_i(\omega_1, \omega_2; v) d\omega_1 d\omega_2,\end{aligned}\quad (59)$$

where we have introduced the functions

$$\Gamma_i(\omega_1, \omega_2; v) \triangleq \int_{-T/2}^{T/2} \left[\frac{(vt \pm b)t}{d_i^2(t; v)} \right]^2 e^{j(\omega_1 - \omega_2)[t - \tau_i(t; v)]} dt, \quad (60)$$

which can be seen as the Fourier transform of the bracketed term, for T large enough. This bracketed term is a slowly varying function of time, so we can approximate $\Gamma_i(\omega_1, \omega_2; v) \approx \delta(\omega_1 - \omega_2) \beta_i(v)$, where

$$\beta_i(v) \triangleq \frac{1}{T} \int_{-T/2}^{T/2} \left[\frac{(vt \pm b)t}{d_i^2(t; v)} \right]^2 dt. \quad (61)$$

If $b/D \ll 1$ then one finds that $\beta_1(v) + \beta_2(v) \approx 4DG_0(\alpha_0)/(Tv^3)$, with α_0 and G_0 defined in (3) and (28) respectively. Substituting this into (59), we obtain

$$\left\| \frac{\partial \mathbf{s}_2(v)}{\partial v} \right\|^2 + \left\| \frac{\partial \mathbf{s}_1(v)}{\partial v} \right\|^2 \approx \frac{4DG_0(\alpha_0)f_s}{(2\pi c)^2 v^3} \left[\frac{1}{T} \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega \right], \quad (62)$$

and then (43) follows. For a narrowband source $s(t) = A \sin \omega_0 t$, the bracketed term in (62) equals $(2\pi)^2 A^2 \omega_0^2 / 2$, so that (27) is obtained. \blacksquare

APPENDIX C: ERROR ANALYSIS

Let $\xi(t) \triangleq \xi(t; v_0, v_0)$ [see (22)], and define the functions

$$\gamma_1(t) \triangleq \left. \frac{\partial}{\partial v} \Delta \tau(t; v) \right|_{v=v_0} \approx -\frac{2b}{v_0 c} \sin \alpha \cos^2 \alpha, \quad (63)$$

$$\gamma_2(t) \triangleq \left. \frac{\partial^2}{\partial v^2} \Delta \tau(t; v) \right|_{v=v_0} \approx -\frac{2b}{v_0^2 c} \sin^3 \alpha \cos^2 \alpha, \quad (64)$$

$$g(t) \triangleq 1 - \frac{\partial}{\partial t} \tau_1(t; v_0) \approx 1 - \frac{v_0}{c} \sin \alpha. \quad (65)$$

which have been written in terms of the angle $\alpha = \alpha(t; v_0)$. Then the functions $p_k(t)$, $k = 1, 2$ in (32) can be computed as follows:

$$p_1(t) \approx -\omega\gamma_1(t)g(t - \Delta\tau(t; v_0))\frac{A\cos[\omega(t - \tau_2(t; v_0) - \xi(t))]}{d(t; v_0)} \quad (66)$$

$$p_2(t) \approx -\omega\gamma_2(t)\frac{A\cos[\omega(t - \tau_2(t; v_0) - \xi(t))]}{d(t; v_0)} - \omega^2\gamma_1^2(t)g^2(t - \Delta\tau(t; v_0))\frac{A\sin[\omega(t - \tau_2(t; v_0) - \xi(t))]}{d(t; v_0)}. \quad (67)$$

The first term in the right hand side of (67) is much smaller than the second. Using these and $g(t - \Delta\tau) \approx g(t)$, the constants ρ_i in (34) become approximately

$$\rho_1 \approx -\frac{\omega}{2} \int_{-T/2}^{T/2} \frac{\gamma_1(t)g(t)}{d^2(t; v_0)} \sin[\omega\xi(t)] dt, \quad (68)$$

$$\rho_2 \approx -\frac{\omega^2}{2} \int_{-T/2}^{T/2} \frac{\gamma_1^2(t)g^2(t)}{d^2(t; v_0)} \cos[\omega\xi(t)] dt. \quad (69)$$

In (69) it is possible to make $g^2(t) \approx 1$. Hence, after changing variables ($\tan \alpha = v_0 t/D$), these lead to (36) and (37).

In order to compute $\sigma_1^2 = \text{var}[N_1]$, write $N_1 = N_{11} + N_{12} + N_{13}$ with

$$N_{11} = \frac{1}{A^2} \int_{-T/2}^{T/2} p_1(t)w_2(t) dt, \quad N_{12} = \frac{1}{A^2} \int_{-T/2}^{T/2} q_1(t)s_2(t) dt, \quad (70)$$

$$N_{13} = \frac{1}{A^2} \int_{-T/2}^{T/2} q_1(t)w_2(t) dt. \quad (71)$$

N_{1i} , $i = 1, 2, 3$ are zero-mean, uncorrelated random variables with variances σ_{1i}^2 ; hence, $\sigma_1^2 = \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2$. From (32), the stochastic process $q_1(t)$ is given by

$$q_1(t) = -\gamma_1(t)w_1'(t - \Delta\tau(t; v_0)). \quad (72)$$

Under the approximation $t_1 - t_2 + \Delta\tau(t_2; v_0) - \Delta\tau(t_1; v_0) \approx t_1 - t_2$, and in view of (5), its autocorrelation is found to be

$$\begin{aligned} E\{q_1(t_1)q_1(t_2)\} &\approx \gamma_1(t_1)\gamma_1(t_2)R_{w'}(t_1 - t_2) \\ &= -\frac{N_0 f_s^3}{2} \gamma_1(t_1)\gamma_1(t_2) \text{sinc}''(f_s(t_1 - t_2)) \end{aligned} \quad (73)$$

where we have used the fact that $R_{w'}(\tau) = -R_w''(\tau)$ [14], so that

$$S_{w'}(f) = \begin{cases} (2\pi f)^2 N_0/2, & |f| < f_s/2 \\ 0, & \text{otherwise} \end{cases} \quad (74)$$

With this, and since the signals of concern are narrowband, for a sufficiently high

sampling frequency f_s we can make the following approximations:

$$\begin{aligned}\sigma_{11}^2 &= \frac{1}{A^4} \int_{-T/2}^{T/2} p_1(t) \int_{-T/2}^{T/2} p_1(\tau) R_w(t-\tau) d\tau dt \\ &\approx \frac{N_0}{2A^4} \int_{-T/2}^{T/2} p_1^2(t) dt \\ &\approx \frac{\omega^2}{2f_s[A^2/\sigma^2]} \int_{-T/2}^{T/2} \frac{\gamma_1^2(t)}{d^2(t; v_0)} dt,\end{aligned}\quad (75)$$

$$\begin{aligned}\sigma_{12}^2 &= \frac{1}{A^4} \int_{-T/2}^{T/2} s_2(t) \gamma_1(t) \int_{-T/2}^{T/2} s_2(\tau) \gamma_1(\tau) R_w(t-\tau) d\tau dt \\ &\approx \frac{N_0 \omega^2}{2A^4} \int_{-T/2}^{T/2} [s_2(t) \gamma_1(t)]^2 dt \\ &\approx \frac{\omega^2}{2f_s[A^2/\sigma^2]} \int_{-T/2}^{T/2} \frac{\gamma_1^2(t)}{d^2(t; v_0)} dt,\end{aligned}\quad (76)$$

$$\begin{aligned}\sigma_{13}^2 &= \frac{1}{A^4} \int_{-T/2}^{T/2} \gamma_1(t) \int_{-T/2}^{T/2} \gamma_1(\tau) R_w(t-\tau) R_w(t-\tau) d\tau dt \\ &\approx \frac{N_0^2 f_s^3 \pi^2}{12A^4} \int_{-T/2}^{T/2} \gamma_1^2(t) dt \\ &= \frac{\pi^2 f_s}{3D^2[A^2/\sigma^2]^2} \int_{-T/2}^{T/2} \gamma_1^2(t) dt.\end{aligned}\quad (77)$$

One has $\sigma_{11}^2 \approx \sigma_{12}^2 \ll \sigma_{13}^2$. Hence, after integrating (77), we obtain (38).

The same approach can now be used in order to obtain $\sigma_2^2 = \text{var}[N_2]$. At the end of the process, one finds that

$$\sigma_2^2 \approx \frac{\pi^4 f_s^3}{5[A^2/\sigma^2]^2} \int_{-T/2}^{T/2} \gamma_1^4(t) dt + \frac{\pi^2 f_s}{3[A^2/\sigma^2]^2} \int_{-T/2}^{T/2} \gamma_2^2(t) dt. \quad (78)$$

For high values of the signal-to-noise ratio A^2/σ^2 , $\sigma_2 \ll \rho_2$ in (69). Therefore, $v_0 - \hat{v}_0 \approx (\rho_1 + N_1)/\rho_2$, from which both the bias and variance in (35) follow. ■

APPENDIX D: MEAN SCORE FUNCTION IN THE BROADBAND CASE

We can write $s_1(t - \Delta\tau(t; v))$, $s_2(t)$ in terms of the Fourier transform $S(\omega)$ as

$$s_1(t - \Delta\tau(t; v)) \approx \frac{(1/2\pi)}{d(t; v_0)} \int_{-\infty}^{\infty} S(\omega_1) e^{j\omega_1[t - \tau_1(t; v_0) - \Delta\tau(t; v)]} d\omega_1, \quad (79)$$

$$s_2(t) \approx \frac{(1/2\pi)}{d(t; v_0)} \int_{-\infty}^{\infty} S^*(\omega_2) e^{-j\omega_2[t - \tau_2(t; v_0)]} d\omega_2. \quad (80)$$

In (79) we have neglected the delay error term $\xi(t; v_0, v)$; thus, the analysis applies only to low speeds. With these, the expected value of $\psi(v)$ becomes

$$E[\psi(v)] \approx \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} S(\omega_1) S^*(\omega_2) \zeta(\omega_1, \omega_2; v_0, v) d\omega_1 d\omega_2, \quad (81)$$

where we have introduced

$$\zeta(\omega_1, \omega_2; v_0, v) \triangleq \int_{-T/2}^{T/2} \left[\frac{1}{d^2(t; v_0)} e^{j[\omega_2 \tau_2(t; v_0) - \omega_1 \tau_1(t; v_0)]} e^{-j\omega_1 \Delta \tau(t; v)} \right] e^{j(\omega_1 - \omega_2)t} dt. \quad (82)$$

Note that the right hand side of (82) is (approximately) the Fourier transform of the bracketed term evaluated at $\omega_2 - \omega_1$. This bracketed term is a slowly varying function of time, so that we can approximate its Fourier transform by an impulse at the zero frequency, weighted by the mean integral of the signal:

$$\zeta(\omega_1, \omega_2; v_0, v) \approx \delta(\omega_1 - \omega_2) \underbrace{\frac{1}{T} \int_{-T/2}^{T/2} e^{j\omega_1 \Delta^2 \tau(t; v_0, v)} \frac{dt}{d^2(t; v_0)}}_{\triangleq \kappa(\omega_1; v_0, v)}, \quad (83)$$

with $\Delta^2 \tau$ as in (48). With this, (81) becomes

$$E[\psi(v)] \approx \frac{1}{(2\pi)^2 T} \int_{-\infty}^{\infty} |S(\omega)|^2 \kappa(\omega; v_0, v) d\omega. \quad (84)$$

Using the approximation (52) and the development following it, we can make

$$\kappa(\omega; v_0, v) \approx J_0 \left(\frac{\omega b(v - v_0)}{c\sqrt{2v_0 v}} \right) \frac{1}{T} \int_{-T/2}^{T/2} \frac{dt}{d^2(t; v_0)}, \quad (85)$$

which leads to (42). ■

REFERENCES

- [1] M. Abramowitz and I. Stegun (eds.), *Handbook of mathematical functions*, New York: Dover, 1964.
- [2] Y. Ban, H. Banno, K. Takeda, F. Itakura, "Synthesis of car noise based on a composition of engine noise and friction noise", *Proc. IEEE ICASSP*, vol. 2, pp. 2105-2108, 2002.
- [3] J. W. Betz, "Comparison of the deskewed short-time correlator and the maximum likelihood correlator", *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 32, pp. 285-294, Apr. 1984.
- [4] J. W. Betz, "Effects of uncompensated relative time companding on a broad-band cross correlator", *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 33, pp. 505-510, Jun. 1985.
- [5] S. Chen, Z. Sun and B. Bridge, "Traffic monitoring using digital sound field mapping", *IEEE Trans. Vehicular Tech.*, vol. 50, pp. 1582-1589, Nov. 2001.
- [6] J.C. Chen, K. Yao and R.E. Hudson, "Source localization and beamforming", *IEEE Signal Proc. Mag.*, vol. 19, pp. 30-39, Mar. 2002.
- [7] "Collaborative information processing", Feature topic, *IEEE Signal Proc. Mag.*, vol. 19, Mar. 2002.
- [8] C. Couvreur and Y. Bresler, "Doppler-based motion estimation for wide-band sources from single passive sensor measurements", *Proc. IEEE ICASSP*, vol. 5, pp. 3537-3540, 1997.

- [9] G.L. Duckworth, M.L. Frey, C.E. Remer, S. Ritter and G. Vidaver, "Comparative study of non-intrusive traffic monitoring sensors", *Proc. SPIE, Intelligent Vehicle Highway Syst.*, vol. 2344, pp. 16-29, Apr. 1994.
- [10] J.F. Forren and D. Jaarsma, "Traffic monitoring by tire noise", *IEEE Conf. Intelligent Transp. Syst.*, pp. 177-182, 1997.
- [11] H. Krim and M. Viberg, "Two decades of array signal processing research", *IEEE Signal Proc. Mag.*, vol. 13, pp. 67-94, Jul. 1996.
- [12] D. Li, K.D. Wong, Y.H. Hu and A.M. Sayeed, "Detection, classification and tracking of targets in distributed sensor networks", *IEEE Signal Proc. Mag.*, vol. 19, pp. 17-29, Mar. 2002.
- [13] K.W. Lo and B.G. Ferguson, "Broadband passive acoustic technique for target motion parameter estimation", *IEEE Trans. Aerosp. Elect. Syst.*, vol. 36, pp. 163-175, Jan. 2000.
- [14] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, New York: McGraw-Hill, 1991.
- [15] Special issue on time delay estimation, *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 29, Jun. 1981.
- [16] J.A. Stuller, "Maximum likelihood estimation of time-varying delay-Part I", *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 35, pp. 300-313, Mar. 1987.
- [17] J.A. Stuller and N. Hubing, "New perspectives for maximum likelihood time-delay estimation", *IEEE Trans. on Signal Proc.*, vol. 45, pp. 513-525, Mar. 1997.
- [18] J. J. Towers and Y. T. Chan, "Passive localization of an emitting source by parametric means", *Proc. IEEE ICASSP*, vol. 5, pp. 2791-2794, 1990.