A BIMODAL SOUND SOURCE MODEL FOR VEHICLE TRACKING IN TRAFFIC MONITORING

Patrick Marmaroli[1], *Jean-Marc Odobez*[2], *Xavier Falourd*[1] *and Hervé Lissek*[1]

1 : Laboratory of Electromagnetism and Acoustic (LEMA) - EPFL, Lausanne, Switzerland
2 : Idiap Research Institute, Martigny, Switzerland
email: patrick.marmaroli@epfl.ch

ABSTRACT

The paper addresses road traffic monitoring using a compact microphone array. More precisely, estimation of both speed and wheelbase distance of detected vehicles is performed. The detection algorithm is based on the comparison between theoretical and measured correlation time series using the two dimensional Bravais-Pearson correlation coefficient. The tracking step is conducted with a particle filter specifically designed to model the position-variant bimodal sound source nature of the vehicles, i.e. taking into account the sound emitted by both vehicle axles. Sensitivity and performance studies using simulations and real measurements show that the bimodal approach reduces the tracking failure risk in harsh conditions when vehicles are tracked, at the same time, in opposite directions.

1. INTRODUCTION

Most of the existing methods for mobile sources monitoring are active sensor-based technologies (e.g. radar). In the context of road traffic monitoring, in addition to active sensors, some intrusive passive ones are placed under/above the road coating (induction loop, pressure sensors) to measure traffic flow related to classes of vehicles (cars, trucks,...). In contrast, we propose to use only one compact array of omnidirectional microphones for this task, which presents the advantage of being passive, non-intrusive and which is a technology already used in standardized measurements for environmental noise. The main principle is to exploit the estimation of the Direction Of Arrival (DOA) of the sound sources to estimate the vehicle location and speed, where the DOA is given by the acoustic path difference between sensors through the well-known Generalized Cross-Correlation (GCC) technique. In addition to speed, the proposed method estimates the wheelbase length of each detected vehicle.

Previous works have addressed this problem. In 1997, Forren et al. [7] and Chen et. al [4] are among the first to show the potentiality of road monitoring through correlation time series. But at this time, no automation was proposed. In 2005, Duffner et al. [5] proposed a pattern recognition method applied to the correlation time series for automatic speed estimation of isolated vehicles moving in two opposite direction; however it fails when vehicles pass each other in front of the microphone array and the wheelbase estimation is not explored. In 2010, an automatic detection method is proposed by Barbagli et al. [1] to prevent traffic jam, but the case of several lanes of circulation is not explored.

Between 14 m/s and 33 m/s approximately, the vehicle passe-by sound emissions are dominated by those produced by the contacts between the tyres and the road. The main difficulties for tracking vehicles on roads with opposite lane of circulation are the abrupt change of the dominant sound source (axle) when the vehicle pass in front of the microphone array and the masking effect of vehicles that pass each other. An example of the correlation time series obtained in the first situation is illustrated on Fig. 1: two vehicles pass one after the other within a period of 6 seconds. The dominant sound source reversal (ahead and rear axle) is clearly observable. This may cause the loss of the vehicle during tracking, especially if a vehicle crossing occurs simultaneously. So the challenge is to track and characterize each vehicle independently using only this mixing acoustic correlation time series. For that, we designed an observation model that explicitly account for the above-mentioned bimodal property (see Fig. 1) in the GCC measurements. Such a model was implemented using a particle filtering algorithm. Experiments showed that the use of this bimodal observation model allowed a wheelbase estimation (which could further be used for vehicle classification) and increased the speed estimation robustness in harsh situations.

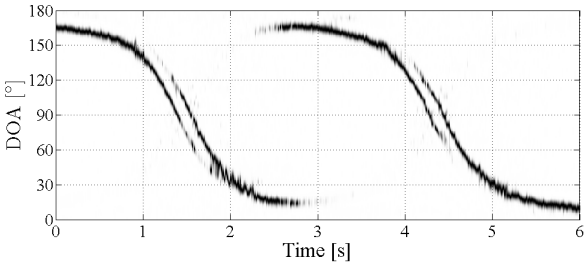

Figure 1: Example of a correlation time series in the DOA [°] vs. Time [s] plane for two vehicles passing in front of the array in clean conditions. The change of dominant axle is clearly observable.

The rest of the paper is organized as follows: in Section 2 the vehicle detection algorithm which is used for the tracking initialization is described; in Section 3, the tracking algorithm based on a bimodal sound source model is detailed; some experiments conducted on both simulated and real measurements are presented and discussed in Section 4; a final discussion about obtained results and forthcoming works is given in section 5.

2. SOUND SOURCES DETECTION

One difficulty in vehicle tracking is the automatic determination of the number of vehicles to track and their origin (left or right). Indeed, the number of sources, their initial position and their presumed speed is generally a crucial knowledge for ensuring some good tracking performance. To address this issue, we rely on a detection algorithm applied prior to the tracking phase. To achieve this under real-time constraints, the basic idea is to compare a detection score $\mathscr{D}(t)$ with a threshold Λ following the equation :

$$\mathscr{D}(t) \lessgtr_{H_1}^{H_0} \Lambda \tag{1}$$

where the hypothesis H_1 (respectively H_0) means that a vehicle is present (respectively no vehicle is present). In the outdoor traffic monitoring context, a simple sound pressure level based detector (in dB) is generally insufficient because of multiple possible noise sources. A better solution is to consider the dynamic information of sources around and check if a measured movement corresponds to a theoretical one. This is the way how $\mathscr{D}$ is developed, according to the theory explained below.

Let an array of M omnidirectional microphones be installed by the wayside at a distance d_r of the right lane and d_l of the left lane.

The road is divided in three zones: the detection zone at the right side, the tracking zone and the left detection zone (see Fig.2). For clarity in reading, only equations for vehicles which comes from right are expressed. The right detection zone is delimited by θ_1 and θ_2. Let $\tau_m(t)$ be the time of flight of the sound wave emitted by a vehicle considered as a monopolar source (the wheelbase is not considered here because of its non observability in this end-fire zone) with coordinates in the 2D plane $[x_s(t), d_r]$ and each microphone m ($1 < m \leq M$) with coordinates $[x_m, y_m]$ at time t. Then :

$$\tau_m(t) = \frac{\sqrt{(x_s(t) - x_m)^2 + (d_r - y_m)^2)}}{c} \quad (2)$$

where c is the speed of sound. Let θ_s(t) be the DOA of the vehicle at time t. The broadband nature of the friction noise between wheels and the asphalt [3] allows to estimate θ_s by taking the largest peak of the GCC function R_t between two microphones i and j parallel to the road and separated by the distance d_a [2] :

$$\theta_s(t) = Arccos\left(\frac{c}{d_a} \times \arg\max_{\tau} R_t(\tau)\right) \quad (3)$$

$$\text{with} \qquad R_t(\tau) = \sum_{f=1}^{N_f} \psi(f) X_i(f) X_j^*(f) e^{2j\pi f\tau} df \quad (4)$$

where f is the frequency bin index, $X_m(f)$ is the discrete Fourier transform with length N_f of the temporal signal $x_m(t)$ recorded by the m^{th} microphone and $\psi(f)$ is a weighting function applied upstream the correlation. The *PHAT* weighting function, presented in [9], is frequently used for its good temporal resolution:

$$\psi(f) = \frac{1}{|X_i(f) X_j^*(f)|} \quad (5)$$

Let v_0 be the presumed speed of the vehicle. The crossing of a detection zone by a vehicle lasts K time frames such that :

$$K = \frac{d_r}{\Delta T} \frac{|\tan\left(\frac{\pi}{2} - \theta_1\right) - \tan\left(\frac{\pi}{2} - \theta_2\right)|}{v_0} \quad (6)$$

where ΔT is the duration of each time frame. Let B_t be the concatenation of the last K GCC until time t (the measured correlation time series), thus $B_t = [R_{t-K}(\tau), ..., R_{t-1}(\tau), R_t(\tau)]$ with size ($N_f \times K$). Knowing the theoretical function θ_s(t) of the *DOA* evolution in time for a vehicle which crosses the detection zone at the speed v_0, a theoretical correlation time series A ($N_f \times K$) can be constructed. Then, A and B_t can be compared in real-time using a 2D distance metric so that if A and B_t are similar enough, a new vehicle is counted (cf. Eq 1).

In practice, two important precautions must be taken concerning the length of the detection zone (fully parameterized by K and θ_1). Firstly, the time of observation needs to be sufficiently large (large K) to observe and to recognize a characteristic vehicle dynamical movement while avoiding the case where two vehicles pass each other inside the zone (not too large K), which leads to an unidentifiable measurement; secondly, the detection zone needs to be sufficiently far from the array (small θ_1) to reduce the effect of the approximate speed setting v_0 while remaining close enough (not too small θ_1) because of signal to noise ratio considerations. All these real-time constraints make the detection system sensitive to noise. One way to increase the robustness to noise is to take advantage of the redundancy brought by several pairs of an M>2 microphone array. If P is the number of used pairs, P theoretical and observation matrices can be created, respectively A^p and B_t^p, p = [1,2,...,P]. A^p and B_t^p are compared using the Bravais-Pearson correlation coefficient in two dimensions $r^p \in [-1, 1]$:

$$r^p(t) = \frac{\sum_{k=1}^{K}\sum_{f=1}^{N_f}\left(A^p(f,k) - \bar{A}^p\right)\left(B_t^p(f,k) - \bar{B}_t^p\right)}{\sqrt{\sum_{k=1}^{K}\sum_{f=1}^{N_f}\left(A^p(f,k) - \bar{A}^p\right)^2 \sum_{k=1}^{K}\sum_{f=1}^{N_f}\left(B_t^p(f,k) - \bar{B}_t^p\right)^2}} \quad (7)$$

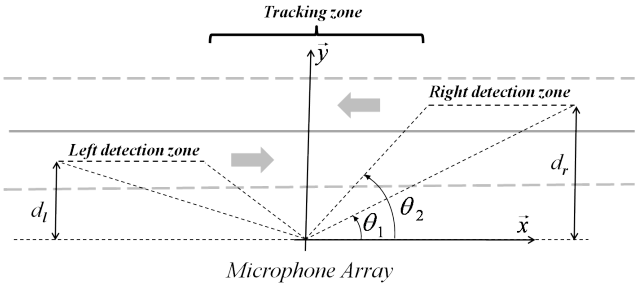

Figure 2: The road is divided in three zones : right detection zone, left detection zone and tracking zone.

where $\bar{A}^p$ (respectively $\bar{B}_t^p$) is the mean value of all values of A^p (respectively B_t^p). The final score $\mathscr{D}$ between -1 and 1 is obtained doing the product of each pair coefficient :

$$\mathscr{D}(t) = \prod_{p=1}^{P} r^p(t) \quad (8)$$

The threshold Λ of the equation (1) is finally optimized according to a ROC analysis [6], see section 4.1.

3. SOUND SOURCES TRACKING

In this section we assume that vehicles have been detected using the method described in the previous Section. Our goal is now to estimate the vehicle speed as well as its wheelbase length. For broadband sound sources, it can be best achieved in the broadside situation as demonstrated in [10]. Thus the tracking zone has been defined in front of the array, as illustrated on Fig. 2.

During the tracking zone crossing, risks of noises on the observation R_t is high and the largest peaks may not always come from sounds emitted by vehicles. The filtering theory permits to overcome this issue by making a distinction between peaks that follow a presumed dynamical movement (vehicles) and the others (noise). In addition, vehicles can not be modeled as single sound sources, as shown in Fig. 1. Not accounting for this leads to failures because the dominant sound source (axle) often changes when vehicles pass in front of the microphone array. This can further be accentuated by masking effects between vehicles. To increase the tracking robustness, we propose to explicitely model the sounds emitted by the two axles and their expected contributions in the observations R_t.

The presented method is described for the particular case when vehicles move on a straight line road of two opposite lanes of circulation in the (x,y) plane. Up to random perturbations, it is assumed that ordinates of each circulation lane are known, the abscissa speed $\dot{x}$ (m/s) is equal to a constant, and the ordinate speed $\dot{y}$ (m/s) is equal to zero.

In order to achieve the location, speed and wheelbase estimation of each vehicle which travels in the tracking zone, we rely on the Bayesian formulation of the tracking problem. In this framework, the objective is to recursively estimate the *a posteriori* distribution $p(\alpha_t|\beta_{1:t})$, known as the filtering distribution, where α_t is the state to estimate at time t and $\beta_{1:t}$ denotes the set of measurements from time 1 to time t. Under standard assumptions (first order dynamical model, conditional independance of observations given the states), the recursion consists of two steps:

$$p(\alpha_t|\beta_{1:t-1}) = \int p(\alpha_t|\alpha_{t-1}) p(\alpha_{t-1}|\beta_{1:t-1}) d\alpha_{t-1} \quad (9)$$

$$p(\alpha_t|\beta_{1:t}) \propto p(\beta_t|\alpha_t) p(\alpha_t|\beta_{1:t-1}) \quad (10)$$

In the first step (prediction), the dynamical model $p(\alpha_t|\alpha_{t-1})$ is used to propagate the filtering distribution $p(\alpha_{t-1}|\beta_{1:t-1})$ at time

$t-1$ to provide the predictive distribution $p(\alpha_t|\beta_{1:t-1})$. In step two (update), the later distribution is combined with the likelihood $p(\beta_t|\alpha_t)$ to obtain the new filtering distribution $p(\alpha_t|\beta_{1:t})$ at time t.

In non-linear, non-Gaussian cases, the system of equations (9,10) can best be solved using sampling approaches also known as particle filters (PF). The idea behind PF consists of representing the filtering distribution using a set of weighted samples (particles) $\{\alpha_t^n, w_t^n, n=1,...,N\}$ and updating this representation since new data are available. Given the particle set of the previous time step, configurations of the current step are drawn from a proposal distribution $\alpha_t \sim q(\alpha|\alpha_{t-1}^n, \beta_t)$. The weights are then computed as $w_t \propto w_{t-1}^n \frac{p(\beta_t|\alpha_t)p(\alpha_t|\alpha_{t-1}^n)}{q(\alpha_t|\alpha_{t-1}^n,\beta_t)}$. In addition, the *multinomial resampling* method [8] consisting of duplicating particles with heavier weights is conducted at each time step to avoid sample impoverishment. Four main elements are important in defining a PF:

- the state model, that is, the abstract representation of the object we are interested in;
- the dynamical model $p(\alpha_t|\alpha_{t-1})$ governing the temporal evolution of the state;
- the likelihood model $p(\beta_t|\alpha_t)$ measuring the adequacy of the data given the proposed configuration of the tracked object;
- a proposal distribution $q(\alpha|\alpha_{t-1}^n, \beta_t)$ which role is to propose new configurations in high likelihood regions of the state space.

In the current case, we use the standard bootstrap filter, in which the dynamical model is used as proposal. In the following paragraphs, we thus present the remaining elements for our vehicle tracking application[1].

State space. It is defined as $\alpha_t = [x_t, y_t, \dot{x}_t, wb_t]^T$ denoting respectively the position (x,y) of the vehicle, its speed $(\dot{x})$ and the wheelbase wb.

Dynamical model. A constant speed model along the road axis disturbed by Gaussian noise is assumed, hence:

$$p(\alpha_t|\alpha_{t-1}) = \mathcal{N}(\mathbf{F}\alpha_{t-1}, \mathbf{V}) \tag{11}$$

with the prediction matrix $\mathbf{F}$ and the noise covariance $\mathbf{V}$ given by:

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{V} = \begin{pmatrix} \sigma_x^2 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{wb}^2 \end{pmatrix}$$

where σ_x^2 (respectively σ_y^2, $\sigma_{\dot{x}}^2$ and σ_{wb}^2) are the noise variances of x (respectively y, speed $\dot{x}$ and wheelbase length wb).

Likelihood model. As it has been shown in [11], R_t can be directly used for estimating vehicle speed in the maximum likelihood sense. In essence, given the state, observations will receive a high likelihood if the time delays of the front axle (x_t^i, y_t^i) and of the rear axle $(x_t^i - wb_t^i, y_t^i)$ actually correspond to high cross-correlation measures. Let us denote by $\tau(x,y)$ the function which computes the relative time delay between microphones for a sound source located at position (x,y) on the road. Then the likelihood is defined as

$$p(\beta_t|\alpha_t^{(i)}) = \gamma(\mu_\tau)R_t(\tau(x_t^i, y_t^i)) + (1-\gamma(\mu_\tau))R_t(\tau(x_t^i - wb_t^i, y_t^i))$$

where $\gamma(.) \in [0,1]$ is a weighting function reflecting the respective contributions of each axle and defined as :

$$\gamma(\mu_\tau) = 0.5\frac{c}{d}\mu_\tau + 0.5 \tag{12}$$

where μ_τ denotes the mean relative delay predicted by the model, which can be computed from the equally weighted generated samples:

$$\mu_\tau = \frac{1}{2N}\sum_{i=1}^{N}\left[\tau(x_t^i, y_t^i) + \tau(x_t^i - wb_t^i, y_t^i)\right] \tag{13}$$

[1]The treatment is presented for vehicles coming from the left. Modifications for vehicles coming from the right are straightforward.

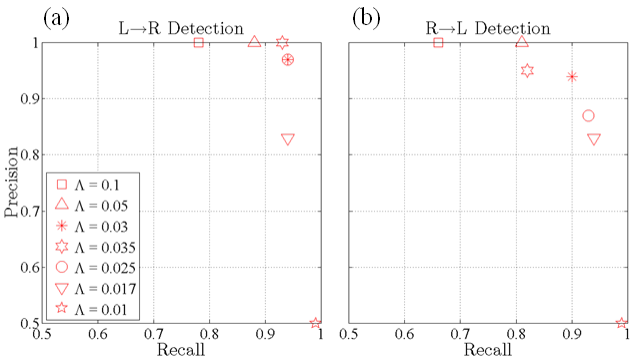

Figure 3: Results in detection based on real measurements for vehicles comes from left (a) and vehicles comes from right (b) in function of threshold Λ.

So, when the vehicle is approaching, $\gamma(\mu_\tau) \geq 0.5$, more importance in the likelihood is given to the front axle. Reversely, when the vehicle is leaving, 1-$\gamma(\mu_\tau) \geq 0.5$, more importance in the likelihood is given to the rear axle.

Initialization and track end. A track is initialized whenever a sound source is detected (cf previous Section). The set of initial particles are drawn in a physically coherent space according to:

$$\alpha_0^i \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x,0} \\ \mu_{y,0} \\ \mu_{\dot{x},0} \\ \mu_{wb,0} \end{bmatrix}, \begin{bmatrix} \sigma_{x,0}^2 & 0 & 0 & 0 \\ 0 & \sigma_{y,0}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x},0}^2 & 0 \\ 0 & 0 & 0 & \sigma_{wb,0}^2 \end{bmatrix}\right) \tag{14}$$

and their weight is set uniformly, i.e $w_0^i = 1/N$. We choose $N = 10000$. Finally, a source is not tracked anymore when 70% of its particles are beyond a fixed abscissa.

State estimation. It is defined as the expected value of the filtering distribution, which, given the sampling approximation, can be simply computed as the weighted mean of the particles.

4. RESULTS

In this section, we are interested in assessing the detection method efficiency on real measurement, studying the sensitivity of our tracking method in function of physical parameters related to in-situ measurements using simulations and comparing qualitatively the robustness of the tracking method with a standard one.

4.1 Detection performances (real data)

Real measurements have been done on a rectilinear road of two opposite lanes (Route du Simplon, St Maurice, Switzerland). 139 vehicles had been recorded in 14 minutes, 72 came from left, 67 came from right. A triangular microphone array is placed at d_r = 8.8 m of the right lane and d_l = 5.1 m of the left lane. Data processing is done on temporal frame of 2048 samples for detection and 1024 samples for tracking both without overlap, using a sampling frequency of 50 kHz.

In order to assess performances of the detection algorithm, a precision-recall graph is established. This is based on True Positive Rate (TPR or recall) and Positive Predictive Value (PPV or precision) given by the formulas [6] :

$$TPR = \frac{TP}{TP+FN}, PPV = \frac{TP}{TP+FP}$$

where TP means the number of true positive detected vehicles, FN the number of missed vehicles and FP the number of wrong detected vehicles (false alarm). Thus, a perfect detector presents

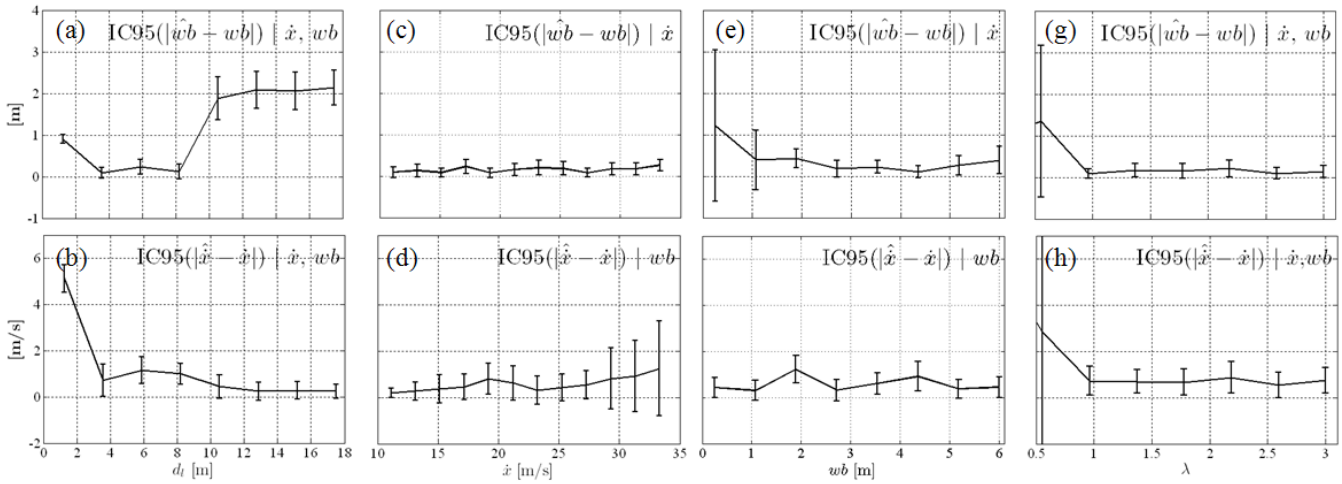

Figure 4: Sensitivity study on the wheelbase estimation (first line) and speed estimation (second line) with respect to the vehicle/array distance d_l (first column), the vehicle speed $\dot{x}$ (second column), the vehicle wheelbase wb (third column) and the ratio distance λ with a parasitic vehicle (fourth column). Mean errors μ_ε of the estimations and IC95 $\mu_\varepsilon \pm 2\sigma_\varepsilon$ are represented.

both null FN and FP and is represented by the point (1,1) on the precision-recall graph. Fig. 3 shows the detection results for 6 different threshold Λ. As expected, results are better for vehicles coming from the left than those coming from the right. This is because the main error source is when two vehicles simultaneously cross the detection zone. When this happens inside the right detection zone, the nearest vehicle masks the further one, and the correlation between the dynamical model and measurement is greatly affected. This problem is reduced in the left detection zone because of the proximity of the vehicle to detect. In both cases, an optimal threshold can be found, here $\Lambda = 0.03$ is the best threshold for left detection with $TPR = 94\%$ and $PPV = 97\%$ and for right detection too with $TPR = 90\%$ and $PPV = 94\%$.

4.2 Bimodal particle filter sensitivity (simulated data)

Simulated data of one vehicle moving from left to right and being tracked between 70° and 130° along the x axis have been generated. They consist in two moving colored noises (central frequency: 1650 Hz, rectangular bandwidth: 3250 Hz) radiating from the axles position towards a triangular microphone array. Amplitude of each sonorous source is inversely proportional to the distance with the microphones. The performances of the wheelbase estimation $\hat{wb}$ and of the speed estimation $\hat{x}$ are studied in function of four parameters :

1. the distance between the vehicle and the array (d_l)
2. the speed of the vehicle ($\dot{x}$)
3. the wheelbase of the vehicle (wb)
4. the distance ratio between the vehicle to track and a parasitic one (λ) which goes on an opposite direction

Each operation is repeated 100 times, allowing a mean error value μ_ε and a 95% confidence interval (IC95) $\mu_\varepsilon \pm 2\sigma_\varepsilon$ for each experimental parametrization, where σ_ε is the standard deviation of the answers. The simulated set-up is identical to the measurement set-up described in the previous part. Initial and noise standard deviations used for tracking are resumed in the following table:

	x [m]	y [m]	$\dot{x}$ [m/s]	wb [m]
Initial σ	$\sigma_{x,0} = 2$	$\sigma_{y,0} = 0.125$	$\sigma_{\dot{x},0} = 10$	$\sigma_{wb,0} = 1$
Noise σ	$\sigma_x = 0.05$	$\sigma_y = 0.025$	$\sigma_{\dot{x}} = 0.3$	$\sigma_{wb} = 0.05$

To study the influence of each parameter specifically, the initial mean values of the particles $\mu_{.,0}$ are set to the exact variable parameters, except for speed when speed estimation sensitivity is studied ($\dot{x}$ = 16.7 m/s and $\mu_{\dot{x},0}$ = 12.5 m/s) and wheelbase when wheelbase estimation sensitivity is studied (wb = 3 and $\mu_{wb,0} = 4$). The latter is done to check the good convergence of the algorithm. All the results are presented on Fig. 4.

Influence of the vehicle/array distance d_l: The influence of d_l on the wheelbase and speed estimation performances is presented in Fig. 4a and 4b respectively. The wheelbase is well estimated, with an IC95 error below 50 cm for a large distance range (2 m $\leq d_l \leq$ 9 m). Below and above these values, the estimation is degraded because, if the vehicle is near the array ($d_l <$ 2 m), the angular speed of both axles is very high and the duration of observation of both modes in R_t is too short to ensure a good convergence, if the vehicle is too far, no distincts peaks appear on R_t and the wheelbase length becomes unobservable. Duration of observation is the critical point for speed estimation too. Hence, speed estimation is better in far field because of longer duration of observation. Since 3 m $\leq d_l$, the mean error of speed estimation does not exceed 2 m/s.

Influence of the vehicle speed $\dot{x}$: The influence of $\dot{x}$ on the wheelbase and speed estimation performances is presented in Fig. 4c and 4d respectively. For wheelbase study, $\mu_{\dot{x},0} = \dot{x}$ and for speed study $\mu_{\dot{x},0}$ is fixed and equal to 12.5 m/s for all tests. Results show that speed has no influence on the wheelbase estimation even for fast vehicles. This proves the fast convergence of the wheelbase estimator. However the standard deviation and mean error of speed estimation increases as the real speed is different from the a priori speed. This shows the importance of a good initialization but also the relative robustness of the PF algorithm for cases with bad initialization, indeed, the mean speed estimation error is lower than 2 m/s in all tested cases.

Influence of the vehicle wheelbase wb: The influence of wb on the wheelbase and speed estimation performances is presented in Fig. 4e and 4f respectively. For the speed study, $\mu_{wb,0} = wb$, and for the wheelbase study $\mu_{wb,0}$ is fixed and equal to 4 m. The array geometry, bandwidth frequency and used generalized cross-correlation function make the wheelbase unobservable for too low values ($wb <$ 0.5 m) inducing an important variance and mean error in the answers. IC95 of the wheelbase estimation is below 50 cm for $2m \leq wb \leq 5m$. The wheelbase length has no influence on the speed estimation because only one mode is sufficient for estimating speed. This result is confirmed in Section 4.3.

Influence of the ratio of distance with a parasitic vehicle λ: Let us consider another (parasitic) identical vehicle that goes in the opposite direction (right to left) at the same speed (16.7 m/s) and at

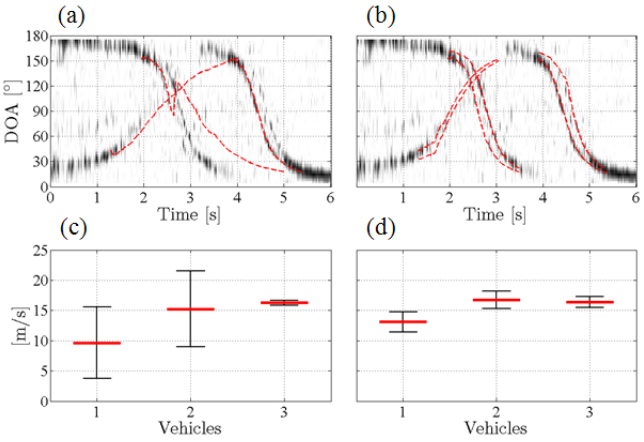

Figure 5: DOA and speed estimation of three vehicles in a real harsh situation. Raw observation and results are superimposed using (a) : a unimodal sound source PF based model, (b) a bimodal sound source PF based model. The same measurement is processed 200 times, mean and IC95 in speed estimates are represented on (c) and (d) for each vehicle.

a distance d_r from the array such that $d_r = \lambda d_l$. The influence of λ on wheelbase and speed estimation performances is presented on Fig. 4g and 4h respectively.When the parasitic vehicle is between the array and the tracked vehicle - $\lambda < 1$ - the measurement is very noisy contrary to $\lambda \geq 1$. Significant errors appear as soon as $\lambda \leq 1$. A good point is that, even for the critical case where $\lambda \approx 1$, both figures show that speed and wheelbase could be efficiently estimated with the proposed method.

4.3 Unimodal vs. Bimodal particle filter (real data)

The Fig. 5 illustrates some tracking results based on a 6 seconds measurement where two vehicles pass each other quickly followed by a third one. It clearly appears on Fig. 5a that using a unimodal (unisource) observation model (i.e. $p(\beta_t|\alpha_t^{(i)}) = R_t\left(\tau(x_t^i, y_t^i)\right)$, no wheelbase taken account) particles follows the most dominant of the two axles, and need to overcome a large gap when the dominant axle is changing, which typically happens when the vehicle is in the broadside situation. Risks of failures during this gap are accentuated when another vehicle is tracked at the same time as it is the case here. This risk is drastically reduced using the bimodal observation model, Fig. 5b, where no gap is noticed anymore and a wheelbase estimation is provided.

Both methods have been applied 200 times on the same measurement. Results in speed estimation for each case (mean and IC95) are depicted on Fig. 5c and Fig. 5d. For the unimodal PF, the IC95 of the vehicle 1 and 2 estimates cover a very large zone of around 11 m/s, showing a very large estimation variability. In contrast, for the bimodal PF, the IC95 are drastically reduced for these vehicles. Regarding the third vehicle, we can notice that both approaches lead on average to the same (correct) speed estimation. However, their variance is slightly different, reflecting the variability of the main observations source they rely on at the end of the tracking: in the unimodal case, as shown in Fig. 5a, the filter tracks the front axle sound source, which happens to be the dominant and less noisy one (in contrast to the usual case, see Fig. 1), whereas the bimodal filter is mainly driven by the rear axle observations in this tracking region (cf Eq. (12)), which are noisier.

5. CONCLUSION

We presented an acoustic method for both detection and characterization of vehicles. The detection step presents good experimental results but the remaining cases of failures come from situations where several vehicles pass each other inside the end-fire detection zone. On the other hand, it is shown that the proposed particle filter, based on a bimodal sound source model, permits a much lower variability of the results in speed estimation of vehicles that pass each other in front of the microphone array in comparison with a particle filter based on a unimodal sound source model. Moreover, the proposed method permits to estimate the wheelbase in addition to speed with a totally passive and non intrusive device. The sensitivity study of the proposed method ensures effective speed and wheelbase estimations under realistic parameters. Forthcoming works include increasing the size of real harsh situations database, improving the detection and the characterization using some smoothing methods in a post-processing way, develop a detection model for vehicle that pass each other in the end-fire detection zone and extending the algorithm to vehicles with unknown number of axles.

6. ACKNOWLEDGMENTS

The authors wish to thank *AER-Sàrl* (Lausanne, Switzerland) for their help on measurements and the provided data.

REFERENCES

[1] B. Barbagli, I. Magrini, G. Manes, A. Manes, G. Langer, and M. Bacchi. A distributed sensor network for real-time acoustic traffic monitoring and early queue detection. In *Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on*, pages 173 –178, 2010.

[2] Bendat. Interpretation and application of statistical analysis for random physical phenomena. *I.R.E Transactions on Bio-Medical Electronics*, 9(1):31–43, January 1962.

[3] V. Cevher, R. Chellappa, and J. McClellan. Vehicle speed estimation using acoustic wave patterns. *Signal Processing, IEEE Transactions on*, 57(1):30 –47, 2009.

[4] S. Chen, Z. Sun, and B. Bridge. Automatic traffic monitoring by intelligent sound detection. In *Intelligent Transportation System, 1997. ITSC '97., IEEE Conference on*, pages 171 – 176, 9-12 1997.

[5] O. Duffner, N. O'Connor, N. Murphy, A. Smeanton, and S. Marlow. Road traffic monitoring using a two-microphone array. In *Audio Engineering Society Convention 118*, 5 2005.

[6] T. Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27:861–874, June 2006.

[7] J. Forren and D. Jaarsma. Traffic monitoring by tire noise. In *Intelligent Transportation System, 1997. ITSC '97., IEEE Conference on*, pages 177 –182, Nov. 1997.

[8] F. Gustafsson. Particle filter theory and practice with positioning applications. *Aerospace and Electronic Systems Magazine, IEEE*, 25(7):53 –82, July 2010.

[9] C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(4):320–327, Aug 1976.

[10] P. Marmaroli, X. Falourd, and H. Lissek. Study of an octahedral antenna for both sound pressure level estimation and 3d localization of multiple sources. In *InterNoise 2010*, 2010.

[11] F. Pérez-Gonzàlez, R. López-Valcarce, and C. Mosquera. Road vehicle speed estimation from a two-microphone array. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 2, page II, 2002.