A comparative study of time delay estimation techniques for road vehicle tracking

Patrick Marmaroli, Xavier Falourd, Hervé Lissek

ACOUSTICS 2012

A comparative study of time delay estimation techniques for road vehicle tracking

P. Marmaroli, X. Falourd and H. Lissek

Ecole Polytechnique Fédérale de Lausanne, EPFL STI IEL LEMA, 1015 Lausanne, Switzerland
patrick.marmaroli@epfl.ch

This paper addresses road traffic monitoring using passive acoustic sensors. Recently, the feasibility of the joint speed and wheelbase length estimation of a road vehicle using particle filtering has been demonstrated. In essence, the direction of arrival of propagated tyre/road noises are estimated using a time delay estimation (TDE) technique between pairs of microphones placed near the road. The concatenation in time of these estimates play the role of an observation likelihood function which determine the particle weights and final convergence quality. In this paper, five classical TDE techniques are detailed and applied on a real road vehicle pass-by measurement. The obtained time series are used as likelihood functions in a particle filtering algorithm with same initial bias and parameters. The accuracy and precision for speed and wheelbase estimation are compared for each case.

1 Introduction

The road traffic monitoring is traditionally based on inductive loops, pneumatic road tubes, video processing and other active sensors (e.g. radar, laser, ultrasonic...). On one hand, some of them are intrusive techniques inducing high costs for their installation and their maintenance for permanent (or not) monitoring of small road areas. On the other hand active techniques are submitted to specific authorization regarding public health and they requires high robustness of the processing to daylight and climatic conditions.

In contrast, the use of acoustic sensors presents some good advantages regarding their cost and their easy installation to monitor larger road areas. The interest of using microphones is twofold, firstly because the noise generated by road vehicle is now considered as a pollutant and constrain communities to dispose of tools to measure it, and secondly because the propagated sound contains a huge amount of information on the type of vehicles, their speed, direction and their wheelbase length.

The measure of the kinematics of road vehicles using passive acoustic sensors laid out near the road was initially investigated by S. Chen *et.al* [1] and J.F. Forren *et al.* [2] which both have independently showed the usefulness of the cross-correlation functions to localize vehicles, even under bad weather conditions [3]. However, no method was proposed at this time to automatically extract the spatial parameters (trajectory, direction of arrival) from the acoustic observations. In 2005, Duffner *et al* proposes to apply a spatial filtering on the correlation time series to automatically detect and estimate speed of vehicles [4]. This proposed method is specifically designed for single lane circulation and it fails in case of crossing vehicles. In 2010, Barbagli *et al.* propose an automatic counting method for queue prevention for a lane of circulation using a linear microphone array [5]. In 2011, we proposed in [6] an automatic procedure based on the particle filtering method to estimate jointly speed and wheelbase length on a two lane road with opposite direction of circulation. As a complementary study of [6], the scope of this paper is to compare several TDE techniques to feed the tracking algorithm in order to increase the accuracy and the precision of both speed and wheelbase estimation.

The remainder of this paper proceeds as follow. A brief spectral analysis of some measured pass-by vehicle noises is provided in Section 2. Different time delay estimation techniques are theoretically detailed in Section 3. An assessment of each technique in the specific context of road vehicle tracking is proposed in Section 4. A conclusion is given in Section 5.

2 Vehicle pass-by noise

External noise radiated by a road vehicle in movement is essentially composed of the powertrain noise (including transmission and exhaust system) and the tyre/road noise. For light vehicles, the former is predominant at low speed and the latter quickly rains the upper hand for speed greater than 50 km/h [7].

To illustrate noise components measured near the road, the spectrogram of successive road vehicles at about 50-60 km/h with two lanes of opposite circulation is depicted on Figure 1. In this example, the audio signal is acquired with one microphone located at 7.5 m of the middle of the road at 1.5 m height. It is clear that the spectrum of the radiated signals enter in the category of broadband noise in the [100 Hz - 3000 Hz] band of frequency, especially when the vehicle is close to the microphone.
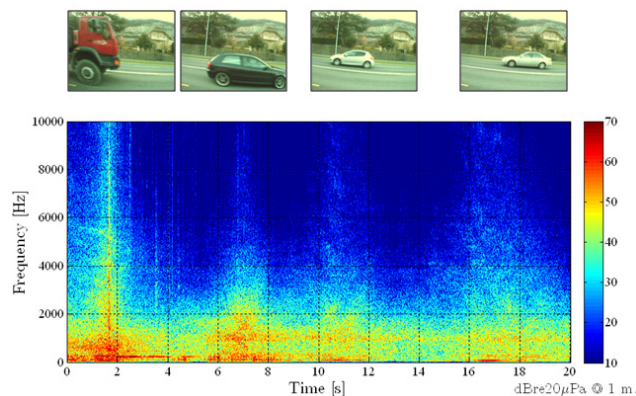

Figure 1: Temporal and spectro-temporal representation of several vehicle passages.

3 Time delay estimation techniques

Passive acoustic array processing based systems aims at extracting temporal and spatial parameters from the analysis of the sound. So as to deliver the number of vehicles and their kinematics, the objective of the proposed technique is the estimation of speeds and wheelbase lengths of passing vehicles by localizing and tracking each axle independently using the following signal model.

3.1 Signal model

Let $x_1[k]$ and $x_2[k]$ be the samples acquired by two spatially separated sensors placed near the road and parallel to the vehicle trajectory at instant kT_s where T_s is the inverse of the sample rate. Let $s[k]$ be the sound signal emitted by one

axle. Assuming a no reverberant medium of propagation, the received signals can be mathematically expressed using the ideal sound propagation model:

$$x_1[k] = s[k] + n_1[k] \quad (1)$$
$$x_2[k] = \alpha s[k-D] + n_2[k] \quad (2)$$

where D and α are respectively the relative time delay of propagation and relative attenuation factor between both microphones. n_j is one additive noise due to measurement devices, assumed as a wide-sense stationary random process, equidistributed among each sensor j and uncorrelated both with the source signals and the noise observed at other sensors. In the following, we assume that $\alpha = 1$ because of the small dimension of the microphone array in comparison with the distance to the vehicle. The relative time delay D is nonlinearly related to the Direction Of Arrival (DOA) θ of the road vehicle through the relation:

$$D = \frac{d}{c} sin(\theta) \quad (3)$$

where d is the inter-sensor distance and c is the speed of sound. In this paper, the *broadside* direction is the angle for which D is null, that is $\theta = \pi/2$ and the *endfire* direction are the angle for which $D = \pm d/c$, that is $\theta = 0°$ or $\theta = \pi$.

As expressed in (3), the localization problem through the observation of DOA is related to the estimation of the time delay D between microphones. In the following, some common TDE techniques dedicated to the broadband context are detailed.

3.2 The cross-correlation

The more straightforward and earliest method to achieve a TDE is the cross-correlation function (CC). Particularly well adapted in case of Constant Delay, Stationary Processes and Long Observation Interval (CDSPLOT) [8], the CC is mathematically expressed by:

$$R_{\mathbf{x}_1\mathbf{x}_2}[m] = \mathbb{E}\left[\mathbf{x}_1[l]\mathbf{x}_2[l+m]\right] \quad (4)$$

where $\mathbb{E}[.]$ stands for the mathematical expectation and $\mathbf{x}_j[k]$ is a block of L samples acquired by the sensor j at instant k such that:

$$\mathbf{x}_j[k] = \left[x_j[k], x_j[k+1], ..., x_j[k+L]\right]^T, \ j \in [1,2] \quad (5)$$

(4) can be viewed as an inner product in a vector space between two non-collinear vectors (i.e. delayed signals with different amplitudes). A maximum result is obtained for identical signals into phase. An estimation of D is given by the lag time for which the cross-correlation reaches its maximum:

$$\hat{D} = \arg\max_m \hat{R}_{\mathbf{x}_1\mathbf{x}_2}[m] \quad (6)$$

where $m \in [-M, M]$ and M is the maximum observable delay. For finite observations, only an estimation of $R_{\mathbf{x}_1\mathbf{x}_2}[m]$ can be achieved:

$$\hat{R}_{\mathbf{x}_1\mathbf{x}_2}[m] = \frac{1}{L-m} \sum_{m=-M}^{M} \mathbf{x}_1[l]\mathbf{x}_2[l+m] \quad (7)$$

After Wiener-Khinchine theorem, the cross-correlation $R_{\mathbf{x}_1\mathbf{x}_2}$ is relied to the cross power spectrum $S_{\mathbf{x}_1\mathbf{x}_2}$:

$$\hat{R}_{\mathbf{x}_1\mathbf{x}_2}[m] = \sum_{\omega=-\infty}^{+\infty} S_{\mathbf{x}_1\mathbf{x}_2}[\omega]e^{j\omega m} \quad (8)$$

where

$$S_{\mathbf{x}_1\mathbf{x}_2}[\omega] = \mathbf{X}_1[\omega]\mathbf{X}_2^*[\omega] \quad (9)$$

and where $(.)^*$ stands for the complex conjugate operator and $\mathbf{X}_j[\omega]$ is the Fourier transform of $\mathbf{x}_j[m]$. From (8), the form of the cross-correlation depends of the spectral contents of the original signals. A narrow spectrum will give a broad cross-correlation and vice-versa. For the extreme case where the original signal is a white noise, the auto-correlation is a delta function with its singular point at the relative time delay D.

3.3 The Generalized Cross-Correlations

The generalized cross-correlations functions (GCC) aim at accentuating the CC peak by filtering the signals upstream the correlation. The general expression of a GCC is given by:

$$\hat{R}_{\mathbf{x}_1\mathbf{x}_2}^{GCC}[m] = \sum_{\omega=-\infty}^{+\infty} \psi[\omega]S_{\mathbf{x}_1\mathbf{x}_2}[\omega]e^{j\omega m} \quad (10)$$

The element which differs from (8) is the weighting function ψ. Different weighting functions are described in the literature: Roth [9], Smoothed COherence Transform (SCOT) [10], Eckart, Hassab-Boucher [11] to cite a few. Two extremely popular weighting functions are the PHAse Transform (PHAT), heuristically developed by Knapp and Carter in 1976 [12], and the Maximum Likelihood transform derived by Hannan and Thomson in 1971 [13]. They are expressed by:

$$\psi_{phat}[\omega] = \frac{1}{|S_{\mathbf{x}_1\mathbf{x}_2}[\omega]|} \quad (11)$$

$$\psi_{ml}[\omega] = \frac{1}{|S_{\mathbf{x}_1\mathbf{x}_2}[\omega]|} \frac{C_{\mathbf{x}_1\mathbf{x}_2}[\omega]}{1-C_{\mathbf{x}_1\mathbf{x}_2}[\omega]} \quad (12)$$

where $C_{\mathbf{x}_1\mathbf{x}_2}[\omega]$ is the *magnitude-squared coherence* defined by:

$$C_{\mathbf{x}_1\mathbf{x}_2}[\omega] = \left|\frac{S_{\mathbf{x}_1\mathbf{x}_2}[\omega]}{\sqrt{S_{\mathbf{x}_1\mathbf{x}_2}[\omega]S_{x_2x_2}[\omega]}}\right|^2, \ 0 \le C_{\mathbf{x}_1\mathbf{x}_2}[\omega] \le 1 \ \forall \omega \quad (13)$$

and |.| denotes the absolute value.

Regarding at the Eq. 12, one can show that the ML processor assigns greater weight in regions of the frequency domain where the coherence is large. From a statistical point of view, the ML processor is the optimal time delay estimation technique in the sense that its variance can achieve the Cramèr-Rao lower bound under CDSPLOT conditions, without reverberation and assuming signal and noise gaussianity distribution. However, the GCC-PHAT seems to be the preferred weighting filter for broadband sound source localization systems. It is known that GCC-PHAT performs more consistently than some other GCC members when the characteristics of the source change over the time [14]; for this latter

advantage it presents good advantages to be used in road traffic monitoring context regarding changes in tyre/road sounds during the pass-by as figured out in Section 2.

3.4 High-Order Statistics

The high-order statistics (HOS) based method are effective under assumption that signal and noise are respectively non-Gaussian and Gaussian processes. HOS exploits the fact that, for Gaussian processes, moments and cumulants of order greater than 2 are null, it therefore can be an advantage to estimate parameters in the higher spectrum domain in case of Gaussian noise (correlated or not). By considering (1) and (2), one can show that [15]:

$$x_2[k] = x_1[k - D] + n_2[k] - n_1[k] \qquad (14)$$

$$= \sum_{q=-Q}^{Q} a[q]x_1[k - q] + n_2[k] - n_1[k - D] \quad (15)$$

where $a[q] = 1$ for $q = D$, 0 elsewhere, and Q is the maximum expected delay defined by the user, $Q \leq M$. Considering the third-order cumulants,

$$C_{x_2 x_1 x_1}[\tau, \rho] = E[x_2[k]x_1[k + \tau]x_1[k + \rho]] \quad (16)$$

$$C_{x_1 x_1 x_1}[\tau, \rho] = E[x_1[k]x_1[k + \tau]x_1[k + \rho]] \quad (17)$$

and substituting (14) into (16) gives:

$$C_{x_2 x_1 x_1}[\tau, \rho] = \sum_{q=-Q}^{Q} a[q]C_{x_1 x_1 x_1}[\tau + q, \rho + q] \qquad (18)$$

Selecting many values of τ and ρ will give an overdetermined system of linear equations in the $a[q]$'s:

$$\mathbf{C}_{x_1 x_1 x_1} \mathbf{a} = \mathbf{C}_{x_2 x_1 x_1} \qquad (19)$$

The estimated delay is the index n which maximizes $|a(n)|$.

3.5 Last Mean Square Algorithm

The Last-Mean Square (LMS) adaptive filter for TDE was proposed by Reed et al. in 1981 [16]. The LMS algorithm is characterized by the update equation:

$$\mathbf{h}[k + 1] = \mathbf{h}[k] + \beta e[k]\mathbf{x}_1[k], \qquad (20)$$

where

$\mathbf{h}[k]$ is a FIR filter of length $2Q + 1$ as:

$$\mathbf{h}[k] = [h_0, h_1, ..., h_{2Q}]^T \qquad (21)$$

$\mathbf{x}_1[k]$ is an input data vector of length $2Q + 1$ as:

$$\mathbf{x}_1[k] = [x_1[k - Q], x_1[k - Q + 1], ..., x_1[k + Q]]^T \qquad (22)$$

e[k] is the error value between the input and the filter output $y[k] = \mathbf{h}^T[k]\mathbf{x}_1[k]$:

$$e[k] = x_2[k] - y[k] \qquad (23)$$

β is the adaptation constant, and Q is the filter length corresponding to the maximal expected delay, $Q \leq M$. An estimation of the time delay is obtained by minimizing the mean-square error between a desired signal $x_1(t)$ and the FIR filter output $\tilde{x}_2(t)$, the lag time associated with the largest component of the FIR filter gives the delay estimate.

4 Comparison

In this section, previously detailed TDE techniques are compared in term of their temporal resolution and their ability to feed conveniently one particle filtering algorithm for the tracking of road vehicles. The calculations are conducted on one audio sequence of 4 seconds corresponding to a vehicle passage from left to right at about 50 km/h. Signals are synchronously acquired on three sensors laid out near the road (5.1 m to the middle of the lane, 1.5 m height). The three sensors form an equilateral triangle of side 25 cm. The sample rate is 50 kHz and the processing is done on successive frames of 41 ms duration with a 50 % overlap.

4.1 Delay time series

Let call a delay time serie (DTS) the concatenation of the successive time delay estimates in time. In this experiment, one DTS per available pair of sensor is computed and coherently projected on the pair parallel to the road. This give a final diagram with the absolute time on the x-axis and the relative time delay on the y-axis. In the following, results are expressed in ms but the conversion in degree (DOA) is straightforward using relation (3).
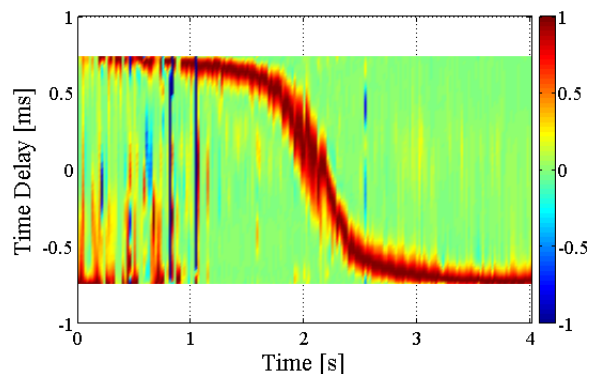

Figure 2: Cross-Correlation

DTS using the classical cross-correlation is depicted on Fig. 2. Coherently with the signal model of Section 3.1, the highest time delays (in absolute value) are observed at the beginning and at the end of the audio excerpt, i.e. when the vehicle in the endfire DOA, and the lowest relative time delay appears at the middle of the sequence, i.e. when the vehicle is in the broadside DOA. Note that some noise is observed at the beginning of the DTS and this last is related to a previous vehicle pass-by.

As expected regarding the Section 3.3, a higher temporal resolution is obtained using the generalized cross-correlation as depicted in Fig. 3 and Fig. 4. Both DTS illustrate that these GCC techniques permit to discriminate each tyre noise of the vehicle in the broadside area for azimuthal range of
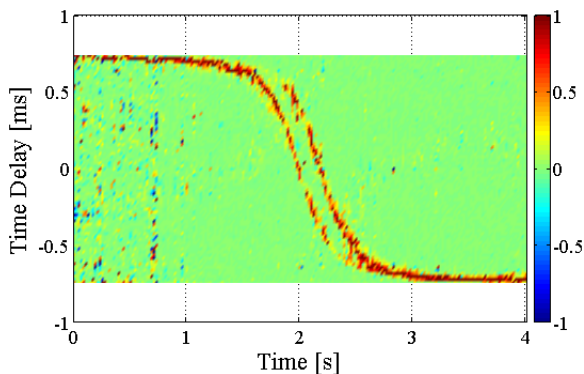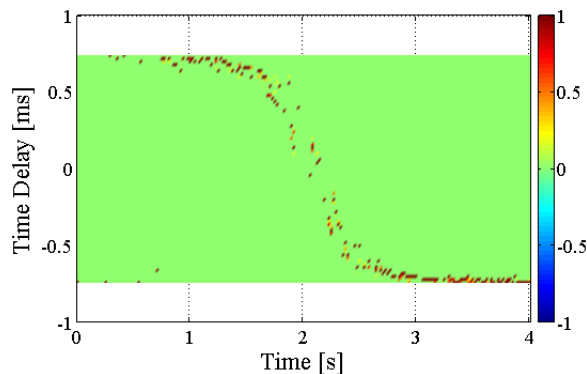
Figure 3: GCC-PHAT

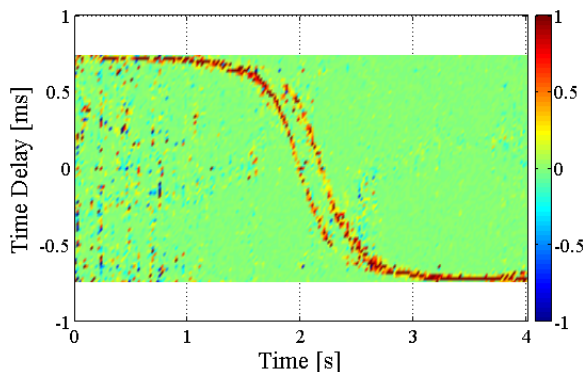
Figure 4: GCC-ML

about 70°. Hence the distance between the two traces is directly related to the wheelbase length of the vehicle. The signal to noise ratio seems to be gently higher for the GCC-PHAT than for the GCC-ML.
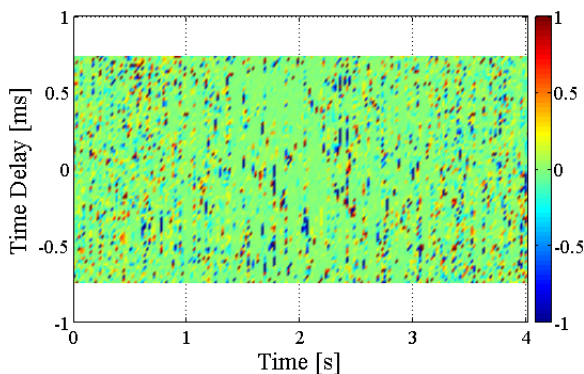

Figure 5: High Order Statistics

In comparison with GCC techniques, DTS extracted from High Order statistics analysis do not permit to detect passing-by vehicle (Fig. 5) . Further investigations, not detailed in this paper, confirm that tyre/road noise can effectively be regarded as a Gaussian noise discarding this method. This explains the total disparition of the signal in the corresponding DTS.

DTS provided by the last mean square algorithm is depicted on Fig. 6. Results are more sparse in the broadside area than in the endfire areas. This is because we defined the number of iterations as a constant and not with respect to an error threshold. Consequently, in regions with high angu-

Figure 6: Last Mean Square

lar speed, the converging procedure does not give as stable results as for regions with lower sound variations. A fixed error threshold may be chosen as a stopping criterion to overcome this problem but it will drastically increase the time of computation. However, a very good signal to noise ratio is provided by the LMS algorithm and can be preferred for low speed vehicle monitoring.

4.2 Assessment with tracking

A particle filtering (PF) algorithm has been launched on DTS previously obtained. The PF - also called Sequential Monte Carlo Method - is a very popular class of algorithms aiming at estimate a hidden state of one non-linear non-Gaussian dynamical system. Recursively, each particle, or hypothesis, is propagated by following a motion model, then weighted according to a likelihood function and finally the space of potential source states is resampled according to the weights of particles [17]. Details of PF implementation for the specific application of road vehicle tracking with two axles are not recalled here but can be found in our previous study [6], for which same notations are used in the following.

In this experiment, $N = 10000$ particles track four states: abscissa x [m], ordinate y [m], speed v [km/h] and wheelbase wb [m]. We voluntary initialized the PF with biased speed and wheelbase initial values in order to determine which kind of TDE technique permit to compensate at best the lack of *a-priori* knowledge on these parameters. Initial wheelbase is set to wb_0 = 3.7 m (real value = 2.7 m) and initial speed is set to different values from v_0 = 0 km/h to v_0 = 100 km/h with 25 km/h step (real value = 50 km/h). Initial abscissa and ordinate are set to $x_0 = -28$ m and $y_0 = 4.2$ m, they are assumed close to the reality. Each DTS is tested 500 times, allowing the computation of the mean and 95% confidence interval values for speed and wheelbase estimates.

From Fig. 7, the *LMS+PF* algorithm provides the worst precision in the estimation for both speed and wheelbase. This is probably due to its strong sparsity so that particle easily diverge. The *GCC-ML+PF* algorithm provides accurate speed estimation but not so accurate result for wheelbase estimation. Both precisions are not so good. The *CC+PF* algorithm which provides an excellent speed estimation with low error and good repeatability but the performance on wheelbase estimation is difficult to judge because of the weak observability of this parameter in the time series. Finally, *GCC-PHAT+PF* provides the best result in term of robustness to biased initialisation for both speed and wheelbase estimation.
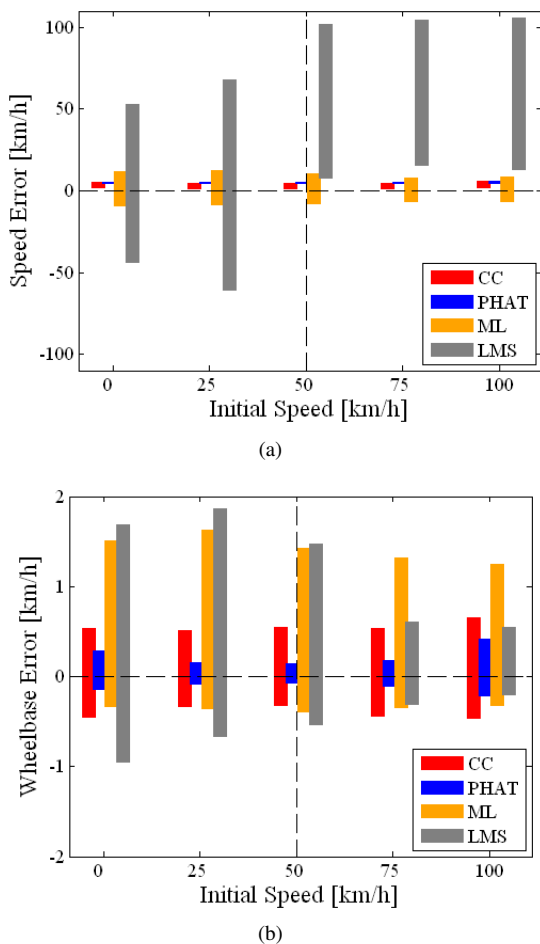
(a)

(b)

Figure 7: Influence of each TDE method on speed and wheelbase estimation performances

5 Conclusion

Different time delay estimation techniques were assessed in the specific context of acoustic road vehicle tracking using a microphone array. Especially, we were interested in finding which method allows the more accurate and precise joint speed and wheelbase estimation from vehicle pass-by noise. Among presented techniques, the GCC-PHAT provided the best observation in term of temporal resolution and convergence quality when associated to a particle filtering algorithm[1].

References

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

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

[3] S. Chen, Z. Sun, and B. Bridge, "Traffic monitoring using digital sound field mapping," *Vehicular Technology,*

IEEE Transactions on, vol. 50, pp. 1582 –1589, Nov. 2001.

[4] 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.

[5] 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*, pp. 173 –178, 2010.

[6] P. Marmaroli, J.-M. Odobez, X. Falourd, and H. Lissek, "A bimodal sound source model for vehicle tracking in traffic monitoring," in *19th European Signal Processing Conference (EUSIPCO)*, 2011.

[7] J.-F. Hamet, "Les mécanismes de génrération du bruit de roulement et l'influence des caractéristiques de chaussée," *Acoustique & Techniques de l'Ingénieur*, no. 32, 2003.

[8] J. Stuller, "Maximum-likelihood estimation of time-varying delay–part i," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, pp. 300 – 313, mar 1987.

[9] P. R. Roth, "Effective measurements using digital signal analysis," *Spectrum, IEEE*, vol. 8, pp. 62–70, April 1971.

[10] G. Carter, A. Nuttall, and P. Cable, "The smoothed coherence transform," *Proceedings of the IEEE*, vol. 61, no. 10, pp. 1497 – 1498, 1973.

[11] J. Hassab and R. Boucher, "Optimum estimation of time delay by a generalized correlator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 27, pp. 373 – 380, aug 1979.

[12] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, pp. 320–327, Aug 1976.

[13] E. J. Hannan and P. J. Thomson, "The estimation of coherence and group delay," *Biometrika*, vol. 58, no. 3, pp. pp. 469–481, 1971.

[14] J. Chen, Y. Huang, and J. Benesty, "Time delay estimation," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems* (Y. Huang and J. Benesty, eds.), pp. 197–227, Springer US, 2004.

[15] C. Nikias and R. Pan, "Time delay estimation in unknown gaussian spatially correlated noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, pp. 1706 –1714, nov 1988.

[16] F. Reed, P. Feintuch, and N. Bershad, "Time delay estimation using the lms adaptive filter - static behavior," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, pp. 561 – 571, jun 1981.

[17] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: fifteen years later," 2008.

---

[1]A video illustration of one tracking trial is available at `http://youtu.be/ulNWOgAd-ig`