



# Direction of arrival estimation of partial sound sources of vehicles with a two-microphone array

Gabriela Dantas Rocha<sup>1</sup>, Julio Cesar B. Torres<sup>1</sup>, Mariane Rembold Petraglia<sup>1</sup>, and Michael Vorländer<sup>2,\*</sup>

<sup>1</sup>Federal University of Rio de Janeiro, Electrical Engineering Program, 21941-901 Rio de Janeiro, Brazil

<sup>2</sup>RWTH Aachen University, Institute of Technical Acoustics, 52062 Aachen, Germany

Received 31 December 2020, Accepted 15 March 2021

**Abstract** – The generalized cross-correlation with phase transform (GCC-PHAT) algorithm has proved to be useful for blindly estimating the direction of arrival of compact sound sources from microphone array recordings. In applications with distributions of partial sources, such as the tires of vehicles in urban environments, the GCC-PHAT needs to be improved, otherwise the detected sound directions change values between directions of the main sources or correspond to an intermediate value between these directions. This paper presents an extension of the GCC-PHAT, based on post-processing of the output delay matrix and on image processing techniques, in order to separately identify directions of the sound produced by the front and rear tires of moving vehicles. The proposed approach can be extended to identify the tire noise directions produced by vehicles with multiple axles. The algorithm performance is analyzed using pass-by measurements of two-axle vehicles, acquired by a two-microphone array. The experiments were conducted with passenger vehicles of four distinct models, running at different speeds. The experimental results show that the proposed method is able to estimate the vehicle speed with an average error of 10.8 km/h and the vehicle wheelbase with 26 cm on average. A possible application is multiple source characterization for parametric vehicle sound synthesis in auralization.

**Keywords:** Vehicle pass-by, Partial source separation, Array measurement, Source characterization, Traffic noise auralization

## 1 Introduction

Noise maps are the main tool for assessing the sound distribution in urban environments. However, it does not portray the real auditory perception of a given urban area, because it is essentially a visual tool displaying long-term sound level averages. An additional approach is the use of auralization [1] in order to extend the noise assessment into audible sound, using recordings and/or simulated data [2–5]. The simulation of an urban acoustic scene must be based on an accurate model for the sound signals from the main sources and for propagation effects due to multiple paths of reflections and diffractions. In the context of urban noise, the most relevant sound sources in road traffic are produced by light and heavy vehicles [6–9].

Car noise emissions contain contributions from various separate sources, whose spectral content are distinct and spatially distributed throughout the vehicle. Among these sources are the tires, the engine and the exhaust system, the first one being dominant for vehicle speeds above 30 km/h [10–12]. This work aims at developing a signal processing method in order to track and separate noise

contributions from the main vehicle sound sources, which, as observed in previous recordings, are produced by the front and rear tires on two-axle vehicles. The sound of each individual source, corresponding to tires of one of the axles, is obtained by modifying time delay estimation methods originally designed for a single sound source.

In [13], generalized cross-correlations associated to particle filtering were employed for tracking vehicle sounds. The vehicle noise was modeled as bi-modal source signal to account for the sound coming from both axles. From this model, more accurate estimates of car speed and wheelbase were produced, compared to those obtained with the uni-modal source model. In a subsequent work [14], the authors proposed a methodology for choosing the appropriate inter-sensor distance to improve the accuracy of wheelbase estimations from pass-by observations, and addressed the problem of tracking road vehicles. However, the separate identification of sound directions coming from front and rear axles has not been addressed in [14], nor, as far as we know, in any other publication.

It is relevant to consider complex partial sound sources in their spatially specific components. In far-field conditions, a sound source may well be characterized by a concentrated source with a directional pattern. This might

\*Corresponding author: [mvo@akustik.rwth-aachen.de](mailto:mvo@akustik.rwth-aachen.de)

apply to street vehicles or airplanes in very large distance. When it comes to the simulation of dynamic scenes with moving sound sources and moving listeners, near-field conditions are also relevant, so that the vehicle source is perceived as a spatially extended source. This is obvious for trains and trams. However, for street vehicles such as cars there is no comprehensive approach to this aspect. In this paper, we develop a technique to extract data for extended sound sources. This data can be used for physically-based synthesis of virtual sources in virtual acoustic environments.

The estimations of speed and wheelbase from acoustic signals have been reported in the literature, with larger predominance of the former [13, 15–18] over the latter [14, 15]. In [16], for instance, a maximum likelihood approach is proposed for speed estimation, which resulted in more accurate and robust estimates when compared to the conventional analysis of sequential short-time cross-correlations.

The direction of arrival (DoA) method proposed in this paper combines time difference of arrival (TDoA) estimates of audio signals in a pair of microphones during a determined observation interval. From the results presented in [19], which compared several DoA estimation algorithms for car pass-by recordings, it was concluded that the best TDoA estimates were obtained with the generalized cross-correlation with phase transform (GCC-PHAT) method [20, 21]. In light of this result, in this work the GCC-PHAT algorithm was employed with the purpose of generating cross-correlation estimates of the signals simultaneously acquired by the two-microphones, for different time-lags. A matrix containing the cross-correlation values of successive time intervals, while the car is close to the microphone array, is formed. By applying image processing techniques to this matrix, the time arrival differences of the predominant source signals are emphasized. Then, using a curve fitting procedure, which employs a dynamical model for the vehicle sound signal, the directions of sound arrival from the front and rear tires are obtained. From parameters extracted from these curves, car speed and wheelbase estimates can be obtained. Since information from the entire pass-by time range is employed to generate such estimates, they are more accurate than those based on data from short periods.

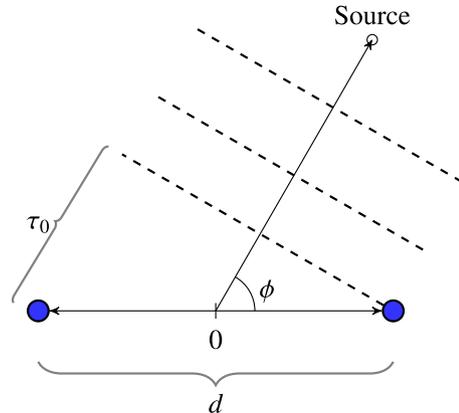
## 2 Direction of arrival estimation

Acoustic source localization might be achieved by time-delay estimation when more than one input channel is available. Let us consider the setup depicted in [Figure 1](#). A two-microphone array with inter-sensor distance  $d$  is placed parallel to a road lane. If a single source contribution is assumed, microphone signals are modeled as,

$$x_1(t) = s(t) + n_1(t), \quad (1)$$

$$x_2(t) = s(t - \tau_0) + n_2(t), \quad (2)$$

where  $s(t)$  is the source signal and  $n_1(t)$ ,  $n_2(t)$  are the noise components.



**Figure 1.** Two-microphone setup for delay  $\tau_0$  and direction of arrival  $\phi$  estimation.

Assuming free-field conditions, the DoA needed for tracking the source and represented by azimuth angle  $\phi$ , might be estimated by,

$$\phi = \arccos\left(\frac{c\tau_0}{d}\right), \quad (3)$$

where  $c$  is the sound speed and  $\tau_0$  denotes the TDoA between the two microphone signals.

The estimator presented next was tested for DoA estimation of vehicle sound sources in a previous work [19], in which five algorithms were compared when executing such task. The results indicated two algorithms as suitable choices, as shown next.

### 2.1 Generalized cross-correlation method

The cross correlation is a measure of similarity between two signals, which is expressed as a function of the delay  $\tau_0$  between them. For white noise signals that differ only by a time delay  $\tau_0$ , the cross-correlation function presents a well-defined peak, with the maximum value occurring for a lag  $\tau$  equal to  $\tau_0$ . For colored noise, such as present in tire noise, the generalized cross-correlation method applies a normalization function to the cross-power spectrum of the two-microphone signals in order to obtain a more prominent peak in the cross-correlation function [20], making it easier to estimate the TDoA.

The cross-power spectrum of the two-microphone discrete-time signals,  $x_1(n)$  and  $x_2(n)$ , can be recursively estimated by [22],

$$\hat{S}_{x_1x_2}(m, k) = \alpha \hat{S}_{x_1x_2}(m-1, k) + (1 - \alpha) X_1(m, k) X_2^*(m, k), \quad (4)$$

where  $X_1(m, k)$  and  $X_2(m, k)$  are, respectively, the  $N$ -point discrete Fourier transforms (DFT) of the windowed microphone signals  $w(n - mJ)x_1(n)$  and  $w(n - mJ)x_2(n)$ , with  $m$  being the frame index and  $k \in \{0, \dots, N-1\}$  the discrete frequency index. The  $N$ -length sequence  $w(n)$  usually employed in audio applications is the Hamming window with shift-size  $J = N/2$ . The exponential weighting coefficient  $\alpha$  is empirically set to  $\alpha = 0.7$ .

The generalized cross-correlation function for frame  $m$  is then calculated as,

$$\hat{R}_{x_1x_2}(m, n) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\hat{S}_{x_1x_2}(m, k)}{|\hat{S}_{x_1x_2}(m, k)|} e^{\frac{2\pi}{N}nk}, \quad (5)$$

with  $n \in \{0, \dots, N-1\}$ . The power spectrum normalization factor employed in Equation (5) is the cross-power estimate magnitude, resulting in a spectral function known as phase transform (PHAT) of  $\hat{S}_{x_1x_2}(m, k)$ . The inverse DFT of such function produces the generalized cross-correlation function  $\hat{R}_{x_1x_2}(m, n)$ , which presents sharper peaks for most audio signals and gives rise to the GCC-PHAT method [20].

Finally, the TDoA for frame  $m$  is estimated from  $\hat{R}_{x_1x_2}(m, n)$  as follows:

$$\frac{\tau_{0m}}{T} \approx n_0(m) = \arg \max_n \hat{R}_{x_1x_2}(m, n), \quad (6)$$

where  $T$  is the sampling period of the discrete-time microphone signals  $x_1(n)$  and  $x_2(n)$ .

### 3 Two-axle vehicle tracking

The TDoA estimation of Equation (6) provides a unique value for each time window, corresponding to the sound of the dominant source captured by the microphone signals. Therefore, one can conclude that the GCC-PHAT algorithm, derived for a single sound source, is not suitable for detecting separately the various noises produced by vehicles. The generalized cross-correlation function of Equation (5) calculated from the signals of a car pass-by acquired by a two-microphone horizontal array is shown in Figure 2, where the darker the gray level, the higher the generalized cross-correlation value. The GCC-PHAT estimates, obtained from the dominant GCC peaks by Equation (6), are shown by the blue curve. It can be observed that when the car passes in front of the microphone array, the estimated TDoA alternates between the directions of the front and rear vehicle axles.

To overcome this problem, we propose the two-source DoA estimation system depicted in Figure 3, where image processing techniques followed by a curve fitting method are appended to the single-source GCC-PHAT algorithm. Although GCC-PHAT is used throughout this work, the system was developed to allow changing and testing different single source TDoA estimators, such as those examined in [19]. Optional algorithms, such as maximum-variance distortionless response (MVDR) [23, 24] and least-mean square (LMS) [25, 26] algorithms, can obtain the TDoA estimations using different criteria, but they must have essential characteristics that are exploited in our system. All single-source estimation algorithms must use two-channel audio recordings and provide as output a matrix  $A(t, \tau)$ , which is a function of the time index  $t$  and of the possible discrete time delays  $\tau$  between the microphone signals. In GCC-PHAT algorithm, this matrix corresponds to auto-correlation matrix estimate  $\hat{R}_{x_1x_2}(m, n)$ , since the delays are obtained from Equation (6).

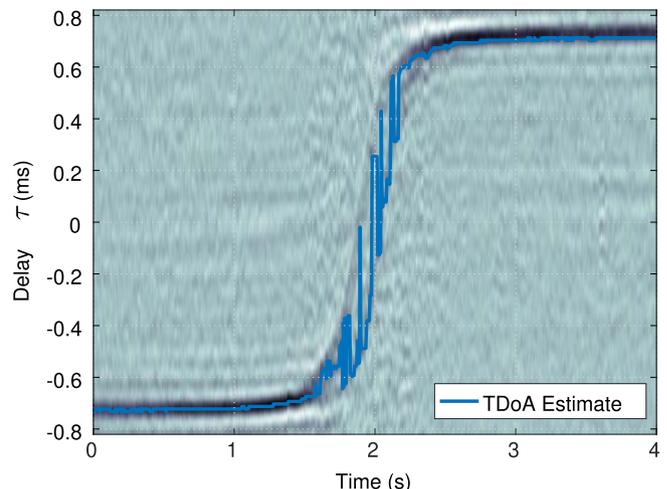


Figure 2. TDoA estimate obtained by the GCC-PHAT algorithm.

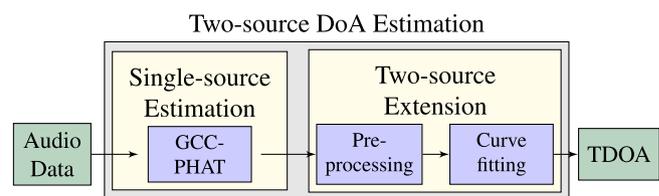


Figure 3. Schematic diagram of the proposed two-source DoA estimation system.

Single-source estimators in their original formulation provide a single DoA estimated curve as output, formed by the maximum value estimations for the various time frames. Such single-source estimate stage is ignored in the proposed system and is replaced by the processing which accounts for two sources. The matrix  $A(t, \tau)$  is used as an input data for the image processing stage, which handles the information of all relevant time frames together. This is possible for offline applications such as the one aimed in this project.

Besides the data from the single-source TDoA estimator, the curve fitting algorithm must be fed with a curve model. A rough model for TDoA curves is derived in agreement to pass-by dynamic and geometric characteristics, as shown in Section 3.1. Moreover, the steps comprised in the “Two-source Extension” block are explained in Sections 3.2 and 3.3. Data pre-processing stage aims at cleaning and adjusting input matrix  $A(t, \tau)$  by removing spurious and irrelevant data. This is carried out in order to improve the curve fitting performance.

### 3.1 Theoretical TDoA model

The theoretical evolution of TDoA over time is obtained by calculating the difference in sound path between the source and the two microphones separated by a distance  $d$ .

In the following derivation, it is assumed that: the source is at ground level, in the  $z = 0$  plane; the microphone

array is in the  $y = 0$  plane, with its axis parallel to the  $z = 0$  plane and at the height  $h$ , measured from the floor up to the array center; the vehicle velocity  $\mathbf{v}$  is parallel to the  $x$ -axis and has constant magnitude  $v_x$ .<sup>1</sup> From simple trigonometric relations and assuming that the vehicle speed is much smaller than the sound speed, it can be demonstrated that the time difference of arrival of the source sound in the two sensors in a given snapshot  $t$  is given by,

$$\tau \approx \frac{l_2(t) - l_1(t)}{v_x}, \quad (7)$$

where,

$$l_1^2(t) = h^2 + s_y^2 + (s_x(t) - d/2)^2, \quad (8)$$

$$l_2^2(t) = h^2 + s_y^2 + (s_x(t) + d/2)^2, \quad (9)$$

with  $s_x$  and  $s_y$  equal to the distances between the source and the  $x = 0$  and  $y = 0$  planes, respectively.

An illustration of the theoretical TDoA behavior for constant speed is shown in Figure 4, where  $\tau$  was obtained from Equations (7)–(9) with  $v_x = 60$ ,  $s_y = 3$ ,  $d = 0.25$ , and,

$$s_x(t) = (t - t_0)v_x, \quad (10)$$

where  $t_0 = 1.5$  s is the time instant in which the vehicle passes exactly in front of the microphone array.

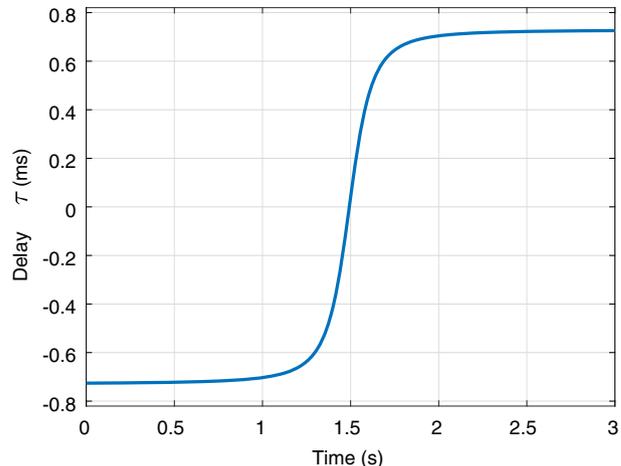
### 3.2 Data pre-processing

The data provided by the single-source DoA estimation algorithm goes through a pre-processing stage consisting of level scaling and noise reduction to adjust it to the curve fitting algorithm. Using the GCC-PHAT, the matrix  $A(t, \tau)$  contains the generalized cross-correlation between the two-microphone signals for different time windows and lag values. Alternative DoA estimation algorithms can provide metrics other than the cross-correlation. The scaling of the input data is, therefore, needed in order to adjust the two-dimensional data contained in  $A(t, \tau)$ , so that it can be treated as a digital image representation, with values in the grayscale range. This image will be further processed to reduce noise by two-dimensional filters.

Due to the decreasing amplitude of the sound propagating over long distances between the source and the microphones, the relevant audio data is concentrated around  $t = t_0$ , when source-receiver distance is minimum and signal-to-noise ratio is maximum. Thus, the data scaling is followed by cropping, which aims to discard irrelevant audio recordings, acquired when the vehicle is far away from the microphones. A 3-length data window is selected around the sample corresponding to maximum signal power ( $t = t_0$ ) and the rest is ignored.

In a real pass-by scenario, the interfering noise generated by other sources, instead of the vehicle, can corrupt the recordings. Depending on noise source position, intensity and spectral content, this interference can seriously impair

<sup>1</sup> Similar TDoA derivation can be performed for other configurations.



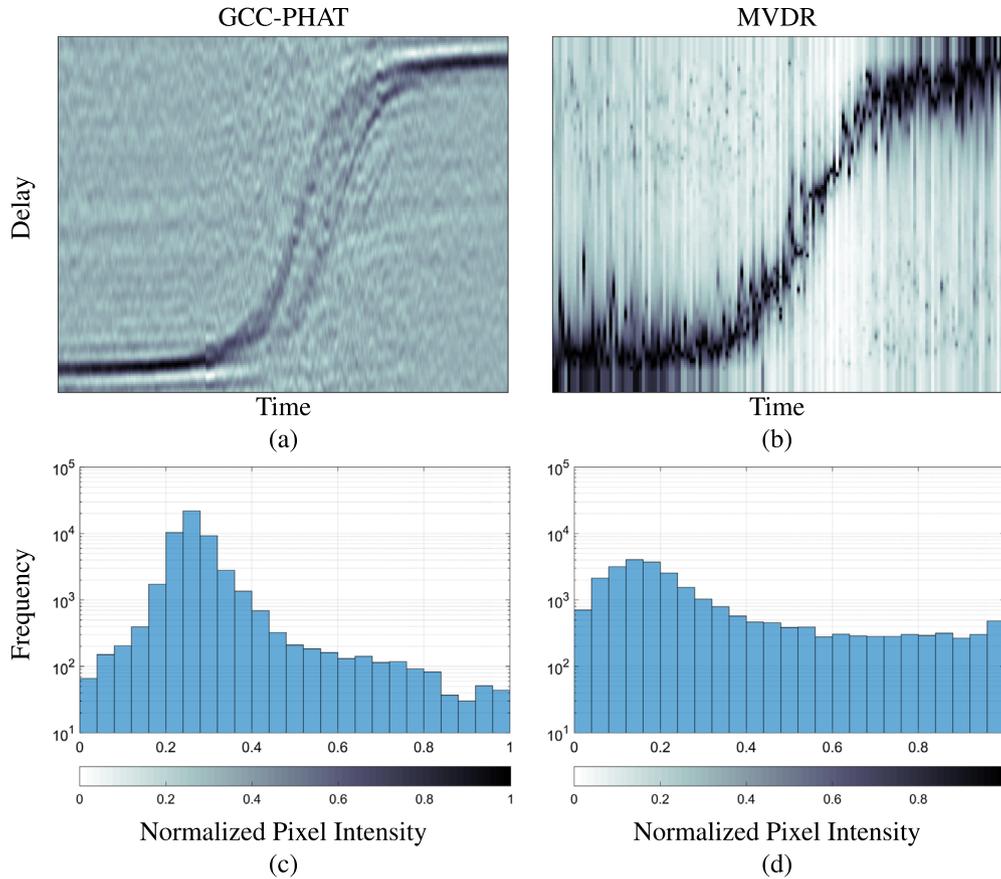
**Figure 4.** Theoretical TDoA curve for a 60 km/h moving vehicle.

the performance of the curve fit method. For this reason, the input data is appropriately filtered, in an attempt to reduce noise.

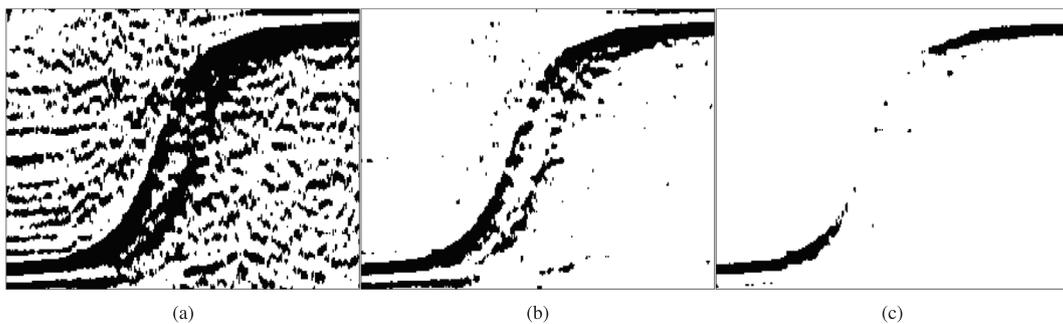
First, the grayscale image is converted into a binary image using a threshold value, selected based on empirical tests. A binary image, composed of either zero or one values, is required by signal processing algorithms to perform morphological operations, such as dilation and erosion. Such operations are used to reduce noise and to provide more precise edge detection to fit the DoA estimation. This empirical optimization leads to different threshold values for the alternative single-source DoA estimators, as a consequence of the diversity in the pixel distribution of the generated images. To illustrate this effect, the images generated by GCC-PHAT and MVDR are shown in Figure 5, together with the respective histogram plots. The histogram graphs indicate how pixel gray-levels are distributed across intensity levels. While for GCC-PHAT pixels are highly concentrated around the 0.3 level, for the MVDR they are more evenly distributed across the different levels. This clear difference between images indicates the relevance of choosing an appropriate threshold value. Binary images generated for different thresholds are depicted in Figure 6. A trade-off between removing noise and maintaining sufficient relevant data is observed.

Next, we describe the image processing approach proposed for noise reduction, with the illustrative results of the main steps shown in Figure 7. The morphological opening operation (dilatation followed by erosion [27, 28]) is performed on the binary image using a square structuring element, which eliminates isolated black pixels. Most noise is removed after the opening operation, as can be observed in Figure 7c.

In this image, two main parallel curves are highlighted, indicating the presence of two dominant noise sources. The time shift between them is in accordance with the sounds emitted by sources whose distance is the average wheelbase in passenger cars. Therefore, it can be concluded that these sound components resulted from the noise



**Figure 5.** Images obtained with (a) GCC-PHAT and (b) MVDR single-source estimators, with respective pixel histograms in (c) and (d).

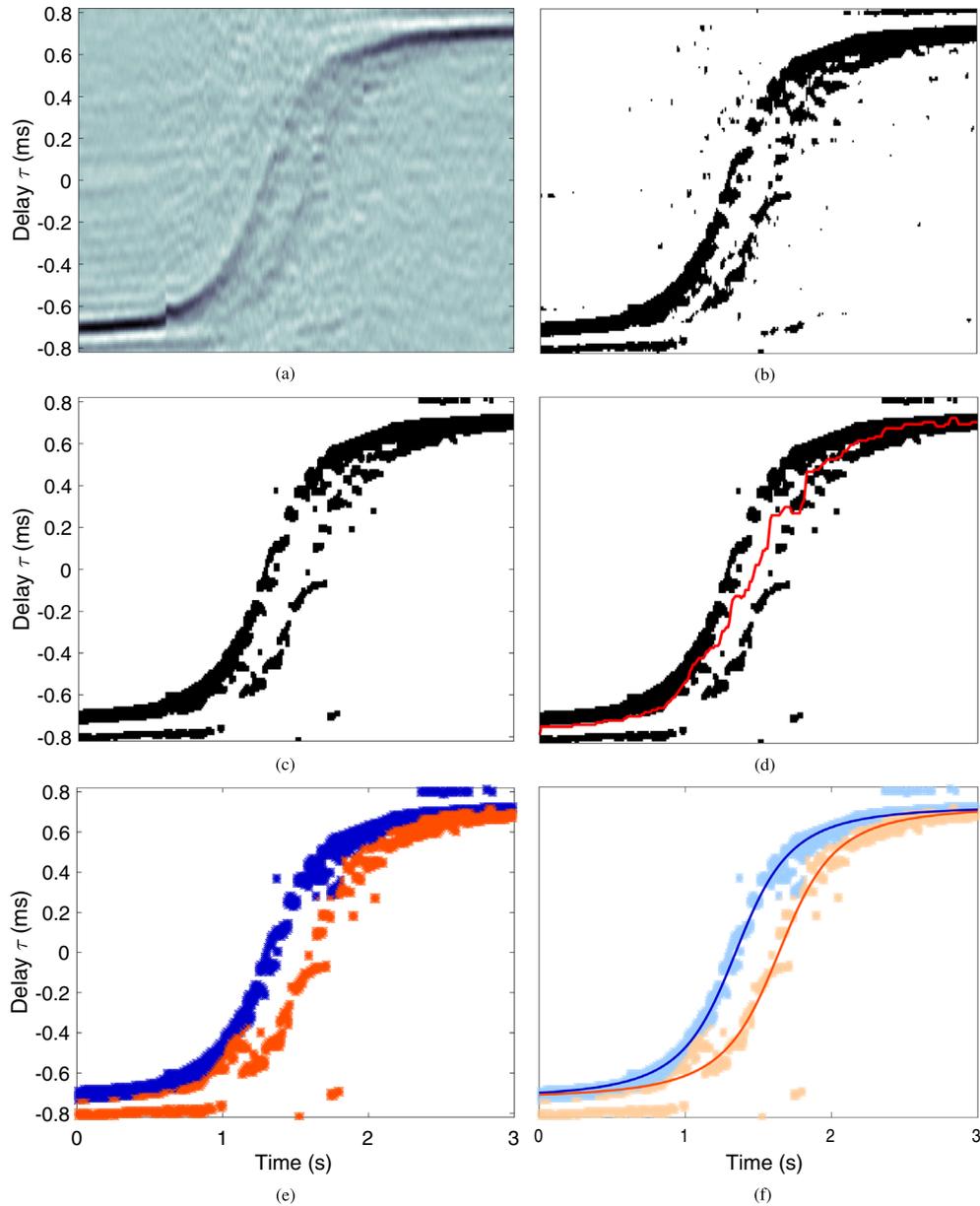


**Figure 6.** Binary images after applying the following threshold values: (a) 0.025, (b) 0.1 and (c) 0.25.

generated by the vehicle’s front and rear axles. This assumption is supported by the work presented in [14], in which a pair of loudspeakers was placed in front of the vehicle’s wheels and the resulting TDoA estimate exhibits the same pattern observed in Figure 7a.

In order to track the noise emitted by the tires of each axle separately, two distinct data sets must be provided to the curve fitting algorithm. Firstly, an “average curve” is calculated to define the points which belong to each axle. This curve is represented by the red line in Figure 7d and is obtained by averaging the indexes of the

non-zero values of each column of the binary image. Then, this curve is used as the border line to separate the data into two sets. The data to the left of the average curve, represented in blue in Figure 7e, is associated with the emission of the front-axle tires (since the left curve appears ahead in time). Likewise, the data to the right of the average curve, displayed in orange in Figure 7e, is associated with the emission of the rear-axle tires. These two data sets, corresponding to the left and right pixels with respect to the average TDoA curve, respectively, are delivered to the curve fitting stage.



**Figure 7.** Illustration of pre-processing and curve fitting results. (a) Gray-scale image; (b) binary image; (c) image after morphological opening; (d) mean TDoA curve; (e) selected data; (f) fitted curves.

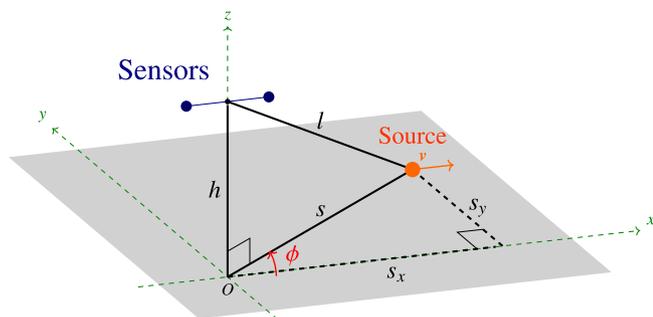
### 3.3 Curve fitting method

The theoretical model obtained for the TDoA in Equation (7) serves here as prototype to fit the data, and the least-squares algorithm is used as optimizer. In light of the non-linear theoretical TDoA graph in Figure 4, the curve fitting method is applied to the image resulting from the pre-processing stage using a trust region for non-linear optimization [29–32]. In addition, spurious data may not be eliminated in pre-processing and may appear as anomaly in the selected data. This is especially problematic, given the known sensitivity of least-squares algorithms to outliers. In this sense, a robust least squares approach is employed.

The cost function is weighted iteratively using bisquare weights [33, 34].

The curve parameters are initialized with random values from a Gaussian distribution. The unknown variables in Equation (7) are the instant  $t_0$ , distance  $s_y$  and the speed  $v_x$ . For each of them, we can define lower and upper limits. These intervals can be used as prior knowledge for initialization by setting the mean values of the random distributions as the centers of the defined intervals. The result is a fitted curve for each data vector, as shown in Figure 7f.

The fitting algorithm finds the model parameter values that minimize the mean squared error between data points



**Figure 8.** 3D representation of the measurement setup: microphone array parallel to the vehicle trajectory; height  $h$  from the floor up to the array center; distance  $l$  between the source and the reference sensor.

and the corresponding fitted curve values. The speed is a parameter of the model of Equation (7) and is estimated directly in the optimization process. The two curves are adjusted independently and the estimated parameters, including speed, may be slightly different for each axle. For simplicity, the average value is used as speed estimate. On the other hand, the wheelbase is estimated as the distance between the fitted curves. This distance is evaluated when the delay  $\tau = 0$ , that is, when each axle is symmetrically in front of the microphone array. In a controlled test scenario, the optimized parameters can be compared to the actual values as a measure of the accuracy of the algorithm.

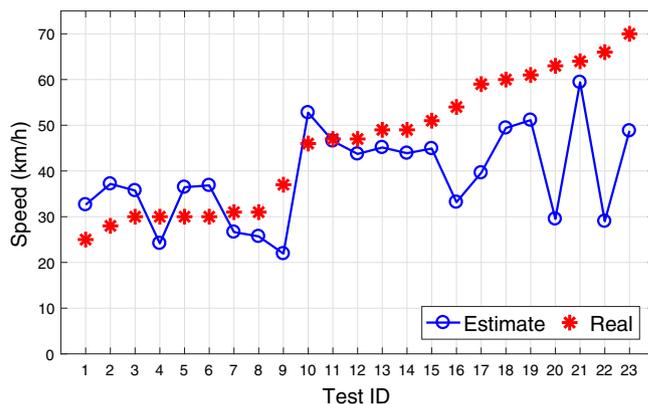
## 4 Experimental results

A set of experiments was conducted with signals acquired by an array of two microphones, aligned horizontally and spaced by  $d = 0.25$  m. It consisted of cars passing on front of the array, one at a time, in a quiet region, without traffic and with negligible background noise, located at the Brazilian National Institute of Metrology (INMETRO), Rio de Janeiro, Brazil. Four different passenger cars were used in the experiments, as detailed in Table 1. Each car passed by the microphone array at constant speeds of around 30, 50, 60 and 70 km/h and with a distance  $s_x = 2.0 \pm 0.2$  m from the array origin, as depicted in Figure 8. A GPS device placed inside the vehicle was used to estimate the speed, based on the car position and time. The vehicle wheelbases were obtained from information provided by their respective manufacturers.

The performance of the proposed system was evaluated by comparing the actual values of the speed and wheelbase of the vehicles, measured during the experiments, with the estimated ones. The speed estimates are represented by blue circles in Figure 9 for each test, while measured values are represented by red stars. Test results were sorted in ascending order of the measured speed values for visualization purposes. The average absolute error was 10.8 km/h. It can be seen in this figure that the estimation error clearly increases for higher speeds, especially for values above 50 km/h. If only the experiments with measured speed below 50 km/h are considered, the average absolute error is 6.3 km/h.

**Table 1.** Vehicles used in pass-by tests.

Vehicle ID	Car model	Transmission
1	Volkswagen Gol 1.0	Manual
2	Jeep Renegade 1.8	Automatic
3	Mitsubishi ASX 2.0	Automatic
4	Hyundai Creta 1.6	Automatic



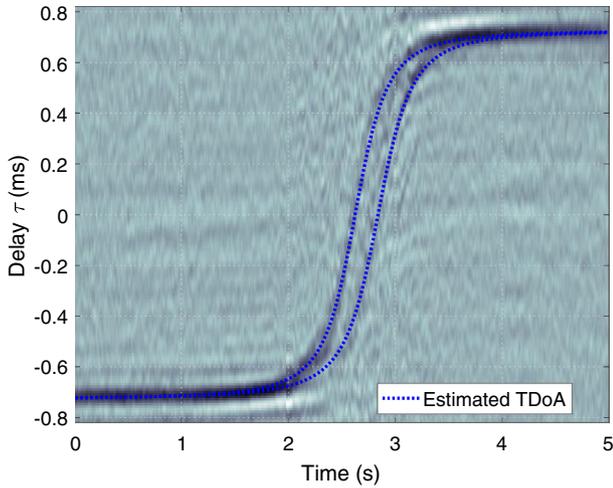
**Figure 9.** Speed estimates obtained using the proposed system with random parameter initialization.

On the other hand, if only the experiments with measured speed above 50 km/h are considered, the average absolute error increases to 18.1 km/h.

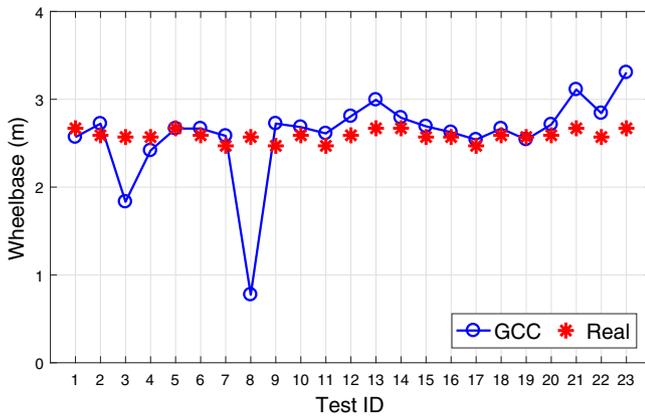
A tendency of underestimating the car velocity is observed in Figure 9, especially for speeds above 50 km/h. One possible cause could be the simplistic model used for the theoretical evolution of the TDoA over time, which assumed constant speed much smaller than the speed of sound and did not take into account the Doppler effect. The curves obtained for Test 22, which presented the largest speed error, are shown in blue in Figure 10. Despite the poor speed estimation result, the fitted curves are well adjusted to the delays obtained for the noise of the two axles of the vehicle, indicating that the error is not caused by the fitting optimization process, but by the theoretical model of the curve.

The wheelbase estimation results are presented in Figure 11 for the experiments in ascending speed order. Blue circles and red stars indicate estimated and measured values, respectively. The average estimation error was 26 cm. Unlike speed estimation, wheelbase results do not appear to be correlated to the vehicle speed. This is in agreement with the sensitivity study carried out in [13], where the increase of vehicle speed had no influence in wheelbase estimation, but caused larger mean error and standard deviation in speed estimation.

Tests 3 and 8 stand out in Figure 11 for presenting high errors. These two tests share similarities which explain the obtained result. The audio files used in Tests 3 and 8 contain sound emissions from the same vehicle, identified as Car 2 in Table 1. Car 2 emissions were also registered in Tests 4, 15, 16, 19 and 22, which resulted in estimation errors below 30 cm. However, for pass-by trials 3 and 8,



**Figure 10.** Tracking results for the highest-speed-error test.

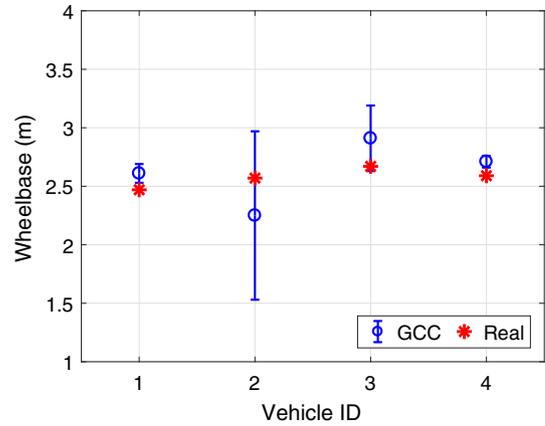


**Figure 11.** Wheelbase estimates obtained using the proposed system with random parameter initialization. Errors of Test IDs 3 and 8 are due to the loud engine noise of Car 2, which was forced to not change gears during the passage.

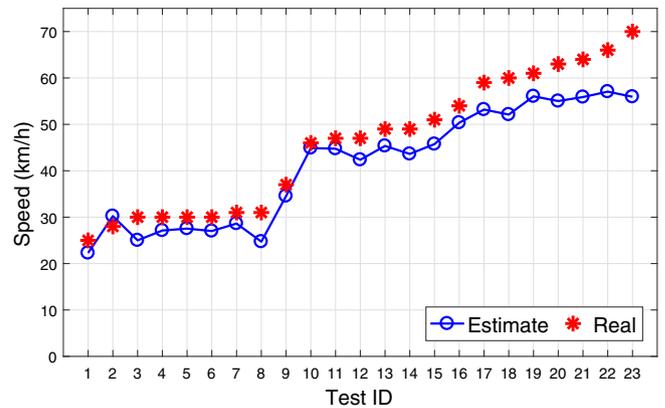
the vehicle was forced to travel in low gear and high engine speed. Engine noise is affected by engine speed [4] and could increase to a level comparable to or higher than tire noise. The proposed system assumes that the tire noise is the dominant sound source and, when this condition is not maintained, such as in Tests 3 and 8, it is expected that the estimation of delay and wheelbase will fail.

Given that the same four vehicles were used throughout the 23 experiments, an average wheelbase estimate is calculated and depicted in Figure 12. For each vehicle, blue circle indicates the wheelbase estimate averaged over all trials in which the vehicle was recorded and red star indicates the measured value. Vehicle 2 presented the highest wheelbase estimation error, equal to 32 cm, due to increased engine noise, whereas Cars 1, 3 and 4 presented errors of 14, 24 and 12 cm, respectively.

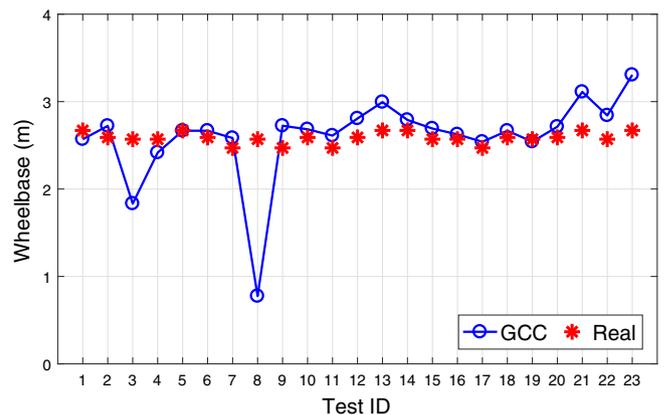
A second set of estimations is performed using measured speed values to initialize the parameters, instead of random initialization. As expected, speed estimation error decreases, as depicted in Figure 13, and the average absolute error in



**Figure 12.** Average wheelbase estimate for each vehicle.



**Figure 13.** Speed estimates obtained using the proposed system and parameter initialization with measured speed values.



**Figure 14.** Wheelbase estimates obtained using the proposed system and parameter initialization with measured speed values.

this scenario is 4.9 km/h. The underestimate tendency is even more noticeable, as it happens for 22 (or 95%) of the 23 tests. The increasing error for higher speeds is still present, although the absolute error for speeds above 50 km/h decreased to 7.4 km/h.

In contrast, wheelbase estimation is barely affected by this change in initialization. The estimates in Figure 14

are almost identical to the ones in Figure 11 and the same 26 cm average error was obtained. The wheelbase is estimated as the distance between both fitted curves when  $\tau = 0$ . Therefore, this value is highly correlated to parameter  $t_0$ , which indicates the instant when the curve crosses  $\tau = 0$ . Speed parameter, in contrast, affects the curve slope around  $t_0$  and for that reason does not present an impact on wheelbase estimates.

## 5 Conclusion

In this paper, we presented a method for separately tracking the axles of a two-axle vehicle using a pair of microphones. The approach is divided in two main steps: first, a time difference of arrival estimator calculates short-time cross-correlations between the microphone signals over an observation time interval, which are pre-processed and stored in a data matrix. Secondly, the matrix with the accumulated treated data is used to obtain two curves, corresponding to the direction of the tire noise of the two axles. From the curve fitting results, the speed and wheelbase of the vehicle are estimated. This modularized approach allows easier testing new algorithms and models and comparing their performance.

In the pre-processing stage, the concatenated cross-correlation data is processed altogether using image processing techniques. Thresholding and opening operations are applied to the image for noise reduction. The choice of the threshold value is critical for the overall system performance and should be further investigated and optimized. Image pixels are then separated into two data vectors, which are separately used in the two-curve fitting model. The proposed system relies on a theoretical model for the time difference of arrival which should be improved in further studies.

The results for vehicle noise tracking were satisfactory and can be applied, as intended, for obtaining source models to be used in acoustic virtual reality systems. The speed estimate is not accurate, especially at high speeds, with the absolute mean speed estimate calculated only over low speed experiments equal to 6.3 km/h and calculated over all experiments equal to 18.1 km/h. In contrast, the wheelbase estimation errors were not correlated to the speed estimation errors, with average absolute error equal to 26 cm. With this, we can continue to build improved models of virtual vehicles as sources in virtual acoustic environments.

It is planned to apply the technique in long-term measurements at busy roads. With additional methods to identify categories of vehicle types and their speeds by video annotation, the aim is to fill a database of parametric data of street vehicles.

## Acknowledgments

The authors would like to thank CAPES and DAAD for partially supporting this research. We are also grateful

to the National Institute of Metrology, Standardization and Industrial Quality (INMETRO), in the figures of Paulo M. Massarani and Zemar M. D. Soares for recordings assistance and to prof. Fernando A. N. C. Pinto, from Laboratory of Acoustics and Vibration from COPPE/UFRJ, for instrumentation support.

## Conflict of interest

Author declared no conflict of interests.

## References

1. M. Vorländer: Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, 2nd ed. Springer Nature, 2020.
2. B. Masiero, W.D. Fonseca, M. Müller-Trapet, P. Dietrich, Auralization of pass-by beamforming measurements, in EAA EUROREGIO, 15–18 September 2010, Ljubljana, Slovenia. 2010.
3. M. Nilsson, J. Forssén, P. Lundén, A. Peplow, B. Hellström: Listen Auralization of Urban Soundscapes. Stockholm University, Chalmers University of Technology, Sonic Studio, KTH Royal Institute of Technology, University College of Arts, Crafts and Design, 2011.
4. R. Pieren, T. Büttler, K. Heutschi: Auralization of accelerating passenger cars using spectral modeling synthesis. Applied Sciences 6, 1 (2016) 5.
5. L. Jiang, M. Masullo, L. Maffei, F. Meng, M. Vorländer: A demonstrator tool of web-based virtual reality for participatory evaluation of urban sound environment. Landscape and Urban Planning 170 (2018) 276–282.
6. P.H.T. Zannin, F.B. Diniz, W.A. Barbosa: Environmental noise pollution in the city of Curitiba, Brazil. Applied Acoustics 63, 4 (2002) 351–358.
7. P.H. Zannin, A. Calixto, F.B. Diniz, J.A. Ferreira: A survey of urban noise annoyance in a large Brazilian city: The importance of a subjective analysis in conjunction with an objective analysis. Environmental Impact Assessment Review 23, 2 (2003) 245–255.
8. B. Jakovljevic, K. Paunovic, G. Belojevic: Road-traffic noise and factors influencing noise annoyance in an urban population. Environment International 35, 3 (2009) 552–556.
9. S. Agarwal, B.L. Swami: Road traffic noise, annoyance and community health survey – a case study for an Indian city. Noise and Health 13, 53 (2011) 272–276. <https://doi.org/10.4103/1463-1741.82959>.
10. K. Heutschi, E. Bühlmann, J. Oertli: Options for reducing noise from roads and railway lines. Transportation Research Part A: Policy and Practice 94 (2016) 308–322.
11. U. Sandberg: Tyre/road Noise: Myths and Realities. Statens väg-och transportforskningsinstitut, 2001.
12. D. O’Boy, A. Dowling: Tyre/road interaction noise – numerical noise prediction of a patterned tyre on a rough road surface. Journal of Sound and Vibration 323, 1–2 (2009) 270–291.
13. P. Marmaroli, J.-M. Odobez, X. Falourd, H. Lissek: A bimodal sound source model for vehicle tracking in traffic monitoring, in 2011 19th European Signal Processing Conference, IEEE. 2011, pp. 1327–1331.
14. P. Marmaroli, M. Carmona, J.-M. Odobez, X. Falourd, H. Lissek: Observation of vehicle axles through pass-by noise: A strategy of microphone array design. IEEE Transactions on Intelligent Transportation Systems 14, 4 (2013) 1654–1664.

15. V. Cevher, R. Chellappa, J.H. McClellan: Vehicle speed estimation using acoustic wave patterns. *IEEE Transactions on Signal Processing* 57, 1 (2008) 30–47.
16. R. López-Valcarce, C. Mosquera, F. Pérez-González: Estimation of road vehicle speed using two omnidirectional microphones: A maximum likelihood approach. *EURASIP Journal on Advances in Signal Processing* 2004, 8 (2004) 929146.
17. P. Borkar, L.G. Malik: Review on vehicular speed, density estimation and classification using acoustic signal. *International Journal for Traffic & Transport Engineering* 3, 3 (2013) 331–343.
18. F. Peréz-González, R. López-Valcarce, C. Mosquera: Road vehicle speed estimation from a two-microphone array, in 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, IEEE. 2002, p. II–1321.
19. G.D. Rocha, F.R. Petraglia, J.C.B. Torres, M.R. Petraglia: Direction of arrival estimation of acoustic vehicular sources, in Proceedings of the 23rd International Congress on Acoustics, 9–13 September, Aachen, Germany. 2019.
20. C. Knapp, G. Carter: The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24, 4 (1976) 320–327.
21. D. Hertz, M. Azaria: Time delay estimation between two phase shifted signals via generalized cross-correlation methods. *Signal Processing* 8, 2 (1985) 235–257.
22. G. Doblinger: Localization and tracking of acoustical sources, in Topics in Acoustic Echo and Noise Control, Springer. 2006, pp. 91–122.
23. S. Vorobyov: Principles of minimum variance robust adaptive beamforming design. *Signal Processing* 93, 1 (2013) 3264–3277.
24. J. Capon: High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE* 57, 8 (1969) 1408–1418.
25. F. Reed, P. Feintuch, N. Bershad: Time delay estimation using the lms adaptive filter—static behavior. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29, 3 (1981) 561–571.
26. E. Ferrara: Fast implementations of lms adaptive filters. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 4 (1980) 474–475.
27. R.M. Haralick, L.G. Shapiro: *Computer and Robot Vision*, Vol. 1. Addison-Wesley Reading, 1992.
28. R.M. Haralick, S.R. Sternberg, X. Zhuang: Image analysis using mathematical morphology, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 4, IEEE. 1987, pp. 532–550.
29. J.J. Moré, D.C. Sorensen: Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing* 4, 3 (1983) 553–572.
30. M.A. Branch, T.F. Coleman, Y. Li: A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing* 21, 1 (1999) 1–23.
31. R.H. Byrd, R.B. Schnabel, G.A. Shultz: Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming* 40, 1–3 (1988) 247–263.
32. T.F. Coleman, Y. Li: An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization* 6, 2 (1996) 418–445.
33. P.W. Holland, R.E. Welsch: Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods* 6, 9 (1977) 813–827.
34. J.O. Street, R.J. Carroll, D. Ruppert: A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* 42, 2 (1988) 152–154.

**Cite this article as:** Rocha GD, Torres JCB, Petraglia MR & Vorländer M. 2021. Direction of arrival estimation of partial sound sources of vehicles with a two-microphone array. *Acta Acustica*, 5, 18.