DIRECTION OF ARRIVAL ESTIMATION OF VEHICLE SOUND SOURCES IN A TWO-MICROPHONE ARRAY

Gabriela Dantas Rocha

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: Mariane Rembold Petraglia
Julio Cesar Boscher Torres

Rio de Janeiro
Junho de 2020

DIRECTION OF ARRIVAL ESTIMATION OF VEHICLE SOUND SOURCES IN A TWO-MICROPHONE ARRAY

Gabriela Dantas Rocha

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Orientadores: Mariane Rembold Petraglia
Julio Cesar Boscher Torres

Aprovada por: Prof. Mariane Rembold Petraglia
Prof. Julio Cesar Boscher Torres
Prof. Fernando Augusto de Noronha Castro Pinto
Dr. Paulo Medeiros Massarani

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2020

Dantas Rocha, Gabriela

Direction of Arrival Estimation of Vehicle Sound Sources in a Two-Microphone Array/Gabriela Dantas Rocha. – Rio de Janeiro: UFRJ/COPPE, 2020.

XII, 46 p.: il.; 29, 7cm.

Orientadores: Mariane Rembold Petraglia

Julio Cesar Boscher Torres

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2020.

Referências Bibliográficas: p. 42 – 46.

1. direction of arrival. 2. traffic noise. 3. wheelbase estimation. I. Rembold Petraglia, Mariane *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Dedico este trabalho a todos que considero minha família.

Agradecimentos

Primeiro gostaria de agradecer a todos que possibilitaram a realização deste trabalho. Aos meus orientadores, Professora Mariane e Professor Julio, pela orientação e confiança ao longo desses anos. À Bruna Croce pela ajuda e pela companhia na jornada que foi desvendar o Unity. Ao Professor Michael Vorländer, Jonas e outros membros do grupo de Realidade Virtual Acústica da Universidade RWTH Aachen, pela importante contribuição dada a este trabalho. Ao Professor Fernando Pinto e ao aluno Frederico, do Lavi, por cederem os equipamentos e auxiliarem na gravação dos áudios aqui utilizados. Aos Engenheiros Paulo Massarani e Zemar, por abrirem as portas do INMETRO e auxiliarem nas medições do ruído. E por fim, ao CNPQ, CAPES e PROBRAL pelo auxílio financeiro sem o qual eu não teria a possibilidade de me dedicar exclusivamente a este projeto.

Agora agradeço a todos que, se não contribuíram diretamente com o trabalho, contribuíram por criar um ambiente de amor, carinho, compreensão e acolhimento dentro e fora da universidade, permitindo que eu me mantivesse saudável, mental e fisicamente, em meio às dificuldades. Agradeço assim a todos os amigos e familiares, responsáveis por formar a pessoa que escreveu esse trabalho. Em especial, agradeço aos amigos do PADS, Fernanda, Gustavo, Tiago e Odair, pelos almoços descontraídos e pós-almoços ainda mais descontraídos, pelos mates oferecidos e recusados, e por todos os planos e ideias incríveis que construímos juntos. Agradeço ainda, em especial, ao meu pai, que me apoiou durante todo esse processo, durante toda a minha vida, e que sempre acreditou em mim.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

DIRECTION OF ARRIVAL ESTIMATION OF VEHICLE SOUND SOURCES IN A TWO-MICROPHONE ARRAY

Gabriela Dantas Rocha

Junho/2020

Orientadores: Mariane Rembold Petraglia
Julio Cesar Boscher Torres

Programa: Engenharia Elétrica

Métodos convencionais de estimação da direção de chegada de fontes sonoras são capazes de gerar estimativas precisas em aplicações com apenas uma fonte. No contexto de ruído de trânsito, cada veículo possui pelo menos duas importantes fontes sonoras, associadas aos pneus dos eixos dianteiro e traseiro. Uma extensão de métodos de estimação da direção de chegada de fontes sonoras a partir de um conjunto de dois microfones é proposta neste trabalho, de modo a lidar com múltiplas fontes. O método proposto é divido em duas etapas: na primeira, um dos algoritmos de estimação da direção de chegada convencionais é usado para gerar uma matriz de atrasos; na segunda, a matriz de atrasos é tratada, utilizando-se técnicas de processamento de imagens, e em seguida usada como entrada para um algoritmo de ajuste de curvas. Quatro métodos são testados para a geração da matriz de atrasos: *Generalized Cross-Correlation* (GCC), *Interaural Time Difference* (ITD), *Least Mean Square* (LMS) e *Eigenvalue Decomposition* (EVD). Duas curvas são obtidas como resultado da segunda etapa, representando as diferenças de tempo de chegada estimadas dos sons emitidos pelos pneus do carro nos dois microfones ao longo do tempo. Além das direções de chegada das duas fontes sonoras principais, estimativas da velocidade e da distância entre os eixos dos carros podem ser obtidas a partir das duas curvas. Medições do som emitido durante a passagem individual de carros de quatro modelos em diferentes velocidades são utilizadas para testar o desempenho do sistema. Os resultados experimentais indicam que, com as modificações introduzidas na segunda etapa, o sistema foi capaz de estimar a distância entre os eixos dos carros com erro médio de 26 cm usando o algoritmo GCC, e de estimar a velocidade com erro médio de 10.01 km/h usando o algoritmo LMS.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

DIRECTION OF ARRIVAL ESTIMATION OF VEHICLE SOUND SOURCES IN A TWO-MICROPHONE ARRAY

Gabriela Dantas Rocha

June/2020

Advisors: Mariane Rembold Petraglia
Julio Cesar Boscher Torres

Department: Electrical Engineering

Conventional methods for the direction of arrival estimation of sound sources are capable of generating accurate estimates in applications with only one source. In the context of traffic noise, each vehicle has at least two important sound sources, associated with the tires on the front and rear axles. An extension of methods for the direction of arrival estimation of sound sources from a set of two microphones is proposed in this work, in order to deal with multiple sources. The proposed method is divided into two stages: in the first, one of the traditional arrival direction estimation algorithms is used to generate the delay matrix; in the second, the delay matrix is treated, using image processing techniques, and then used as an input to a curve fitting algorithm. Four methods are tested for generating the delay matrix: Generalized Cross-Correlation (GCC), Interaural Time Difference (ITD), Least Mean Square (LMS) and Eigenvalue Decomposition (EVD). Two curves are obtained as a result of the second stage, representing the time difference of arrival estimated of the sounds emitted by the car tires on the two microphones over time. In addition to the directions of arrival of the two main sound sources, estimates of the speed and distance between the axles of the cars can be obtained from the two curves. Measurements of the sound emitted during the individual passage of cars of four models at different speeds are used to test the performance of the system. The experimental results indicate that, with the modifications introduced in the second stage, the system was able to estimate the distance between the axles of the cars with an average error of 26 cm using the GCC algorithm, and to estimate the speed with an average error of 10.01 km/h using the LMS algorithm.

Contents

List of Figures

List of Tables

List of Abbreviations

DFT Discrete Fourier Transform

EVD Eigenvalue Decomposition

END Environmental Noise Directive

FIR Finite Impulse Response

GCC Generalized Cross-Correlation

IDFT Inverse Discrete Fourier Transform

ITD Interaural Time Difference

LMS Least Mean Square

WHO World Health Organization

PHAT Phase transform

PSD Power Spectral Density

RPM Revolutions per Minute

Chapter 1

Introduction

Serious annoyances, sleep disorders and an increased risk of cardiovascular disease are related to the harmful effects of noise exposure on people's quality of life and health [1]. Road traffic is identified by many studies as the main source of environmental noise [2–8], which the World Health Organization (WHO) has declared as a public health problem. A large part of the world population lives in cities, where traffic noise is more intense, being continuously exposed to noise and its serious consequences. Strategies for assessing traffic noise are therefore of great interest to the scientific community and the general public.

Noise mapping was declared as a common noise assessment method in the European Directive 2002/49/EC, known as the Environmental Noise Directive (END) [9]. Noise maps provide a graphical representation of a noise indicator, defined in the END as a physical scale to describe environmental noise and its harmful effects. However, the relationship between noise indicators and human perception is not intuitive. This problem is aggravated because, in addition to people with technical knowledge, such as architects, urban planners and engineers, noise maps generally need to be evaluated by non-specialists, such as decision makers, politicians and the general public. In this context, auralization stands out as a more intuitive noise assessment strategy, in which a virtual acoustic environment is generated, providing the listener with an immersive experience.

The auralization systems are based on three main modules responsible for the generation, transmission and reproduction of sound, respectively [10]. The first module provides the "dry" input signal obtained by anechoic recordings [11] or synthesis [12]. Sound transmission introduces effects accumulated by the sound signal in the path between the source and the receiver. Finally, sound reproduction adds other effects to simulate the desired listening experience. In the context of the auralization of traffic noise, road vehicles stand as important sources of disturbance and must be well represented in the sound generation module. Therefore, an accurate simulation of traffic noise depends on knowledge of the characteristics of the vehicle's sound

sources.

It would be impractical or impossible to accommodate a moving car inside an anechoic chamber to measure its emission characteristics or record the sound emitted by the vehicle in several directions. For this reason, synthesis strategies have been employed to model urban noise sources in previous implementations of auralization systems [13–15]. Another strategy is to acquire vehicle sound emissions, among other sound sources, and select only vehicle emissions by applying signal processing techniques to filter, both in time and space, the recorded audio data. From the filtered data, it is possible to infer spectral characteristics of the signal, which can be used to synthesize the vehicle noise emissions. Therefore, it is necessary to know the vehicle's position, or at least its direction, to perform spatial filtering and obtain the desired sound.

A simple configuration for the location of acoustic sources is defined by a pair of microphones that record emissions from a sound source. Each microphone receives the signal from the source delayed in a slightly different amount, and this time difference indicates which direction the original signal came from. Direction of arrival (DOA) estimators are often used in speech applications [16]. Researches on speaker tracking usually deal with highly reverberated environments and non-overlapping dominant sources. On the other hand, the present work addresses the problem of vehicle sound tracking, where reverberation is not an important issue and the sound source of interest is composed of several contributions that overlap in time and frequency. In addition, vehicle tracking differs from speaker tracking as the first cannot always be approached as a point source, depending on the distance between the car and the microphone array. Car noise emission is actually a combination of several noise sources associated with different elements of the vehicle. The main sources are the tires, the engine and the exhaust system, where the former are dominant for vehicle speeds above 30 km/h [17]. Therefore, instead of tracking the vehicle as a whole, we propose to track tire noise, which has been shown to be dominant in most pass-by recordings [18–20]. In addition, the contributions of the front and rear tires are tracked separately, by modifying the conventional direction of arrival estimation algorithms, designed for single sound sources.

Four methods for estimating the direction of arrival are investigated in this work [16] for application in the context of the location of vehicle sound sources in urban environments. The first method uses a traditional correlation approach to deal with the problem, calculating a modified (or "generalized") cross-correlation function between two recordings. The second method explores the physical constraints of the microphone array to test all possible directions of arrival and select the most suitable one. The two final methods use adaptive filters as an effort to model the impulse response of the channel between sources and microphones. All four algorithms are

based on the estimation of the delay difference between the microphone signals, known as *time difference of arrival* (TDOA). These techniques are robust for speaker tracking in reverberating environments and in a scenario where there is only one active voice at a time [16].

An extension, common to all four methods, is proposed to deal with the two dominant sources observed in emissions from two-axle vehicles. This extension is necessary due to the significant difference between tracking vehicles instead of speakers. In this work, we use a sequence of time difference of arrival estimates, calculated by one of the original methods, and ignore the final step, in which a dominant source is assumed and a single direction of arrival (DOA) estimate per frame is calculated. Instead, image processing and curve fitting optimization techniques are applied to obtain the TDOA estimates. A sequence of delay estimations, calculated by one the conventional DOA method, for the entire duration of the car passage is taken into account simultaneously, being interpreted as an image rather than examined frame by frame. A dynamic model of a single vehicle pass-by is combined with the geometric characteristics of the experimental setup to obtain a model for the TDOA curves. Two curves are fitted to the data points, corresponding to the TDOA estimates over time for the sounds coming from the two axles' directions.

Moreover, given the TDOA model used to adjust the two curves, the optimization of their parameters translates into computing estimates for the vehicle speed and the wheelbase. Several studies have been published on acoustic velocity estimates [18, 21–23] and wheelbase estimates [19, 21] in the last two decades. In [22], for example, a direct maximum likelihood approach is proposed for speed estimation. By calling it a "direct" approach, the authors emphasize the difference with respect to the conventional approach of analyzing short-term cross-correlations. The direct approach resulted in more accurate and robust estimates. Similarly, we propose to replace the short-time maximum value assessments by an optimization which accounts for all time frames simultaneously.

Publications on acoustic vehicle tracking are hardly found in the literature. In [18], the vehicle is modeled as a bi-modal sound source, in order to account for both axles, and the generalized cross-correlation method associated with particle filtering is used. This approach resulted in more accurate estimates of the speed and wheelbase, compared to an approach that uses uni-modal source modeling. In a later work [19], a methodology for choosing an appropriate distance between sensors was investigated in order to optimize the estimates obtained from [18]. In these works, the tracking problem of road vehicle was addressed, but, as far as we know, tracking vehicle axles separately has not been addressed before.

The purpose of this work is to present a method to estimate the time difference of arrival of vehicular two-source sounds in a two-microphone array. In addition,

we intend to compare four conventional direction of arrival estimation algorithms that can be as part of the proposed system, indicating which ones are suitable for vehicle noise applications. The proposed method consists of two main stages: in the first, conventional direction of arrival estimators are implemented and one of them is chosen to compute a delay matrix; in the second, image filtering followed by curve fitting optimization are applied to the delay matrix, resulting in direction of arrival estimates for the sounds coming from the front and rear axles' tires. In other words, the first stage applies strategies found in the literature to generate a matrix with intensities related to TDOA values along the time, in which the elements corresponding to the delays of the dominant source signals are emphasized; the second stage is responsible for searching the TDOA matrix for the values associated with two sources, corresponding to an extension of the original methods that consider only one source. The delay estimation performed by the proposed system applies to sound emissions of two-axle cars in which tire noise is the main sound source. Noise measurements from four different passenger car models, passing through an array of two microphones at different speeds, are used as input data for the validation of the proposed method.

The algorithms employed in this work for the estimation of the direction of arrival of one-source signals are presented in Chapter 2, and the proposed two-source extension is presented in Chapter 3. Chapter 4 contains a description of the experimental tests carried out to validate the proposed method, as well as the results for the directions of arrival, speed and wheelbase estimates. The conclusion of this work and proposal for future work are presented in Chapter 5.

Chapter 2

Direction of Arrival Estimation

Four direction of arrival estimators, based on signal delay between two microphones, are presented next. The acoustic scenario described in Fig. 2.1 is used throughout this chapter.

Two microphones are placed d meters apart from each other and receive the signal emitted by a sound source. Each microphone receives source signal at slightly different instants and the microphone signals can be modeled as

$$\begin{cases} x_1(t) = s(t) + v_1(t) & \text{(2.1a)} \\ x_2(t) = s(t - \tau_0) + v_2(t), & \text{(2.1b)} \end{cases}$$

where $s(t)$ is the source signal and $v_1(t)$, $v_2(t)$ are the noise components. In other words, the signal at the furthest microphone is modeled as a delayed version of the signal at the closest one, plus noise. Furthest and closest refer here to the distance between source and both microphones.

Assuming free-field conditions, direction of arrival might be estimated from delay τ_0 between sensor signals, as indicated by

$$\phi = \arccos\left(\frac{c \cdot \tau_0}{d}\right), \tag{2.2}$$

where c is the speed of sound and ϕ is the azimuth angle.

2.1 Generalized Cross-Correlation Method

One strategy for direction of arrival estimation lies in using the Generalized Cross-Correlation (GCC) method [16, 24]. Cross-correlation function embodies an idea of similarity between signals, which is expressed as a function of the delay τ between them. Thus, for two nearly identical signals that differ only by a time delay τ_0, cross-correlation function reaches its maximum value when the applied lag τ equals

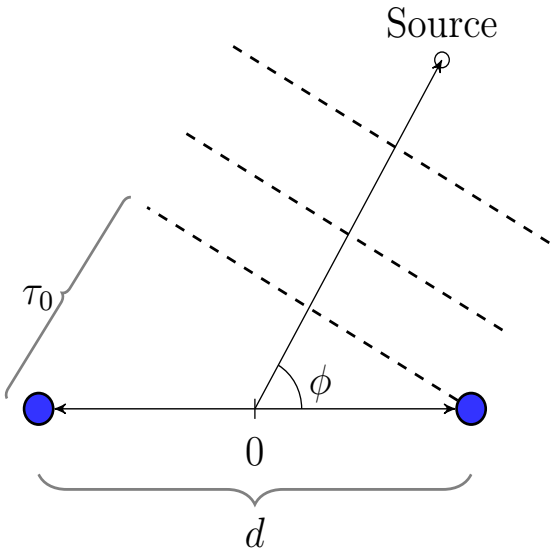

Figure 2.1: Two-microphone setup for delay τ_0 and direction of arrival ϕ estimation.

the actual delay τ_0.

As in [16], GCC function is estimated in frequency domain using cross-power spectrum. Discrete-time representation of sensor signals of (2.1) are given, respectively, by

$$\begin{cases} x_1(nT) = s(nT) + v_1(nT) & \text{(2.3a)} \\ x_2(nT) = s(T(n - \tau_0/T)) + v_2(nT), & \text{(2.3b)} \end{cases}$$

with sampling period T.

A frequency-domain representation, obtained by applying Fourier Transform in Eqs. (2.3), is

$$\begin{cases} X_1(e^{j\Omega}) = S(e^{j\Omega}) + V_1(e^{j\Omega}) & \text{(2.4a)} \\ X_2(e^{j\Omega}) = S(e^{j\Omega})e^{-j\Omega\frac{\tau_0}{T}} + V_2(e^{j\Omega}). & \text{(2.4b)} \end{cases}$$

Cross-power spectrum is defined as

$$S_{x_1x_2}(\Omega) = E\left[X_1(e^{j\Omega})X_2^*(e^{j\Omega})\right] = S_{ss}(\Omega)e^{j\Omega\frac{\tau_0}{T}}, \quad (2.5)$$

where $E[.]$ is the expected value operator and $S_{ss}(\Omega)$ is the source signal power spectrum. However, since each random process represented by a microphone signal is observed for a single realization, during a limited period of time, $S_{x_1x_2}(\Omega)$ cannot be computed from (2.5) and needs to be estimated. Assuming that $x_1(n)$ and $x_2(n)$ are ergodic processes, the expected value operation in (2.5) is replaced by a time average, which must be taken in small intervals in order to guarantee stationarity. Thus, power spectrum estimation is carried out over windowed signal frames, as it

follows:

$$\hat{S}_{x1x2}(m,k) = \alpha\hat{S}_{x1x2}(m-1,k) + (1-\alpha)X_1(m,k)X_2^*(m,k), \tag{2.6}$$

where m is the frame index, k the discrete frequency index and $X_1(m,k)$ and $X_2(m,k)$ are Discrete Fourier Transform (DFT) representations of windowed microphone signals. Exponential weighting coefficient α is typically set between 0.7 and 0.8.

Generalized cross-correlation function is defined as

$$R_{x_1x_2}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\psi_{12}(e^{j\Omega})S_{x_1x_2}(\Omega)e^{j\Omega n}d\Omega, \tag{2.7}$$

and the only difference between the expression above and the regular, well-known, cross-correlation is the weighting function $\psi_{12}(e^{j\Omega})$. If $\psi_{12}(e^{j\Omega}) = 1$, (2.7) becomes the inverse Fourier Transform of cross-power spectrum, as in classical cross-correlation definition [25]. An interpretation for the weighting function is presented in [24], where term $\psi_{12}(e^{j\Omega})$ is introduced to the correlation equation by means of a pre-filtering operation, applied to the input signals in order to enhance correlation peaks. the Phase Transform (PHAT) weighting function, used in the present implementation, is defined as

$$\psi_{12}(e^{j\Omega}) = \frac{1}{|S_{x1x2}(\Omega)|}. \tag{2.8}$$

In practice, generalized cross-correlation (2.7) is obtained using inverse DFT (IDFT) of cross-power spectrum estimate (2.6), thus

$$\hat{R}_{x1x2}(m,n) = \frac{1}{N}\sum_{k=0}^{N-1}\frac{\hat{S}_{x1x2}(m,k)}{\left|\hat{S}_{x1x2}(m,k)\right|}e^{j\frac{2\pi}{N}nk}, \quad n = 0,1,...,N-1, \tag{2.9}$$

for discrete frequency index k, frame index m and frame length N.

Finally, delay difference between microphones 1 and 2 is estimated from $\hat{R}_{x1x2}(m,n)$ for each frame m, as it follows

$$\frac{\tau_{0m}}{T} \approx n_{0m} = \arg\max_n \hat{R}_{x1x2}(m,n). \tag{2.10}$$

2.2 Interaural Time Difference Method

Interaural difference, or the delay difference of a sound between the ears, is used by humans and other animals as a tool for identifying sound sources location. Similarly,

time difference of arrival between the two microphone signals is used in ITD (Interaural Time Difference) method for the same purpose. Main idea here is to generate a set of all possible delay differences between microphones and then search within this set to identify the actual delay. Assuming that both signals differ only by a time delay, the search is conducted by introducing artificial delays to the original signals and checking the similarity between these delayed versions of original signals. Time delay of arrival is estimated as the artificial delay responsible for making both signals to coincide.

Azimuth range $\phi \in [0, \pi]$ is divided into I equally separated sectors, where I must be an odd number. Each angle of arrival corresponds to a time delay, as indicated in (2.2), and a TDOA set is generated by

$$\tau_i = \frac{d}{2c} \sin\left(\frac{i-1}{I-1}\pi - \frac{\pi}{2}\right), \quad i = 1, 2, ..., I, \tag{2.11}$$

where d is the distance between microphones and c is the sound propagation speed. The algorithm is implemented in frequency domain, where delays τ_i defined in (2.11) correspond to phase factors

$$p_k(i) = e^{-j\frac{2\pi}{N}kf_s\tau_i}, \quad k = 0, 1, ..., \frac{N}{2}, \quad i = 1, 2, ..., I, \tag{2.12}$$

with sampling frequency $f_s = 1/T$.

The Discrete Fourier Transform of signals, $X_1(m, k)$ and $X_2(m, k)$, are computed for each frame m and multiplied by the generated phase factors. Phase-shifted DFTs are then compared for each frequency bin k, as shown in Figure 2.2. Only vertically aligned pairs are compared and, as a result of array geometry, all possible phase differences are covered in this way.

Comparison is carried out in the coincidence detection box, illustrated in Figure 2.2, by means of the squared difference of delayed signals

$$\Delta_i(m, k) = |p_k(i)X_1(m, k) - p_k(I - i - 1)X_2(m, k)|^2, \qquad i = 1, 2, ..., I \tag{2.13}$$

which is minimized with respect to the index i, resulting in

$$i_{otm}(m, k) = \arg\min_i \Delta_i(m, k), \qquad k = 0, 1, ..., \frac{N}{2}. \tag{2.14}$$

Optimum index i_{otm} indicates which among the artificial delays $p_k(i)$ generated the most similar delayed signals. Time delay estimate is given by the delay corresponding to the optimum index, or $\tau(m, k)|_{i=i_{otm}(m,k)}$.

A histogram averaging strategy is introduced in order to handle phase ambiguity and fluctuations in delay estimates over frequency. Optimum indices $i_{otm}(m, k)$ are

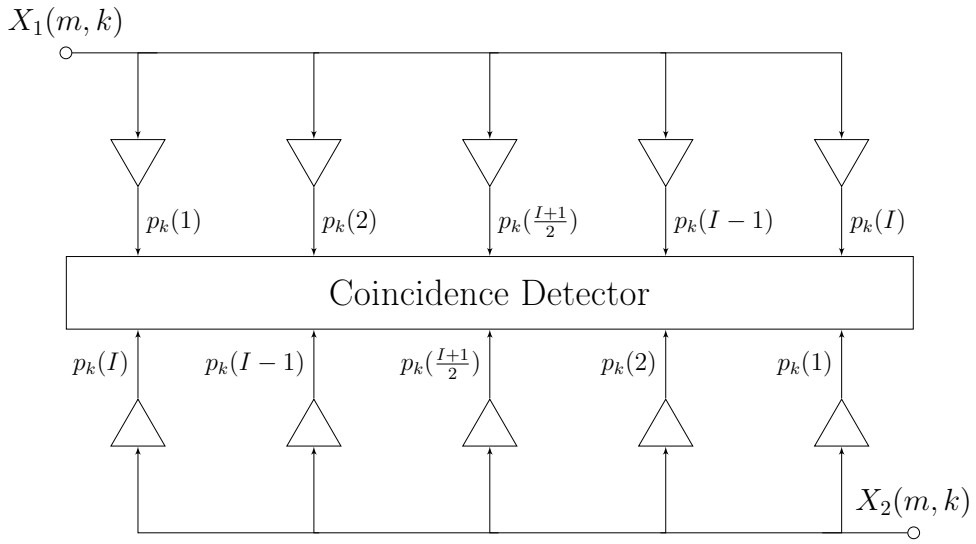

Figure 2.2: ITD system schematic.

accumulated in a histogram map $P_k(\tau_i, m)$, which is updated for each new evaluated frame. Map $P_k(\tau_i, m)$ contains information from current frame m and exponentially weighted past frames, as indicated

$$P_k(\tau_i, m) = \alpha P_k(\tau_i, m-1) + \delta(i - i_{otm}(m,k)), \quad \begin{cases} i = 1, 2, ..., I \\ k = 0, 1, ..., \dfrac{N}{2}, \\ m = 0, 1, 2, ... \end{cases} \tag{2.15}$$

where $\delta(.)$ is the Kronecker delta function[1] and τ_i is defined in (2.11). Forgetting factor α is chosen between 0.85 and 0.95. Average delay estimate is obtained by summing $P_k(\tau_i, m)$ over frequency bins

$$P_{sum}(\tau_i, m) = \sum_{k=0}^{\frac{N}{2}} P_k(\tau_i, m), \tag{2.16}$$

and checking which delay corresponds to maximum histogram sum

$$\tau(m) = \arg\max_{\tau_i} P_{sum}(\tau_i, m). \tag{2.17}$$

2.3 Least Mean Square Method

An adaptive method is now presented as an alternative, assuming that microphone signals are simply delayed versions of the source signal, i.e., plane wave propagation in free field. Delay estimation in this approach is obtained by modeling the impulse

[1] $\delta(x)$ equals 1 for $x = 0$ and 0 otherwise.

responses between source and microphones.

In Least Mean Square (LMS) algorithm [26], one of the two sensor signals is filtered by a finite impulse response (FIR) system and the result is compared to the remaining signal, as depicted in Figure 2.3. This comparison results in error signal $e(n)$, which is fed back to the adaptive filter. Filter coefficients are updated in the sense of minimizing the error according to a normalized version of LMS algorithm [27, 28], given by

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \frac{\mu}{\|\boldsymbol{x}_2(n)\|^2} e(n)\boldsymbol{x}_2(n), \tag{2.18}$$

where μ is the step size. After adaption, filter coefficients $w(n)$ would ideally all be zero, except for the sample corresponding to delay τ_0.

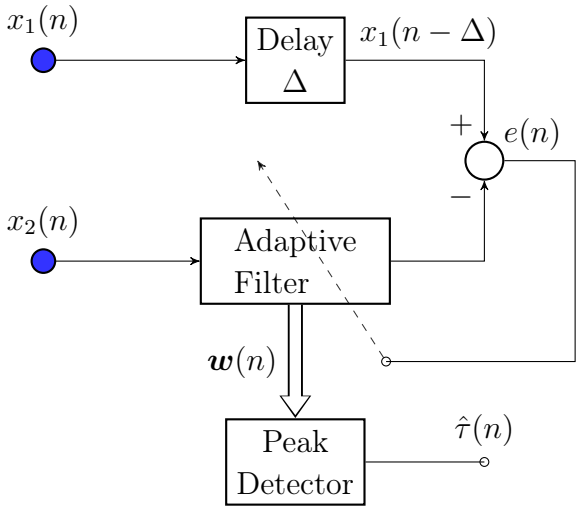

Figure 2.3: Adaptive LMS algorithm schematic.

The FIR system of length L is implemented in frequency domain. Signal samples are grouped into frames of length $N = 2L$ and filter coefficients are held constant while each frame is being processed. Overlap-save method is used towards implementing linear convolution operation, which is required for filtering signal $\boldsymbol{x}_2(n)$.

Filter coefficients are updated at each frame m according to the following equations:

$$X_2(m,k) = \sum_{n=0}^{N-1} x_2(mL+n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0, 1, ..., N-1 \tag{2.19}$$

$$y(m,n) = \frac{1}{N}\sum_{k=0}^{N-1} W(m,k)X_2(m,k)e^{j\frac{2\pi}{N}nk}, \quad n = 0, 1, ..., N-1 \tag{2.20}$$

$$\tilde{e}(m,n) = \begin{cases} 0, & n = 0, 1, ..., L-1 \\ x_1(mL+n-\Delta) - y(m,n), & n = L, L+1, ..., N-1 \end{cases} \tag{2.21}$$

$$E(m,k) = \sum_{n=0}^{N-1} \tilde{e}(m,n) e^{-j\frac{2\pi}{N}nk}, \quad k = 0, 1, ..., N-1 \tag{2.22}$$

$$S_{x_2x_2}(m,k) = \alpha S_{x_2x_2}(m-1,k) + (1-\alpha)|X_2(m,k)|^2, \quad k = 0, 1, ..., N-1 \tag{2.23}$$

$$W(m+1,k) = W(m,k) + \frac{\mu}{S_{x_2x_2}(m,k)+\epsilon} X_2^*(m,k) E(m,k), \quad k = 0, 1, ..., N-1. \tag{2.24}$$

The ideal scenario of one single non-zero filter coefficient is not possible, given the finite model used for channel impulse response nad, also, the wave propagation effects which are not considered in the model. However, one coefficient should still dominate over the others. Delay estimates are thus obtained by maximizing filter coefficients for each frame.

2.4 Eigenvalue Decomposition Method

The second adaptive algorithm performs delay estimation using an adaptive eigenvalue decomposition of a covariance matrix. First, the channel between the source and each microphone is modeled as a linear and time invariant system, with impulse responses $h_1(n)$ and $h_2(n)$, as depicted using dashed lines in Figure 2.4.

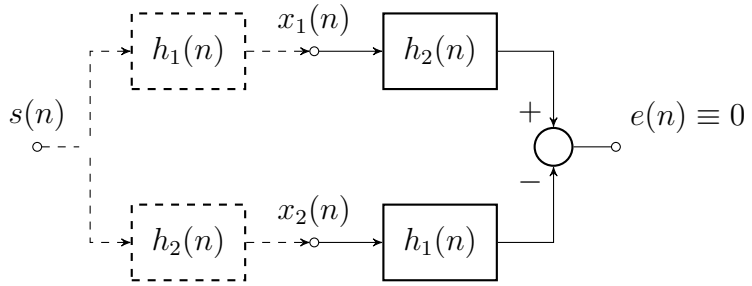

Figure 2.4: Left: Signal model, in dashed lines. Right: impulse response estimator, in solid lines.

Filtering each microphone signal by the opposite impulse response, as indicated in Figure 2.4, results in the following relation:

$$x_1(n) * h_2(n) = h_1(n) * s(n) * h_2(n) = h_1(n) * x_2(n), \tag{2.25}$$

where associative and commutative properties of linear convolution were employed. Equation 2.25 enables formulating an eigenvalue problem, as shown next.

First, (2.25) is written in vector notation

$$\boldsymbol{x}_1^T(n)\boldsymbol{h}_2(n) - \boldsymbol{x}_2^T(n)\boldsymbol{h}_1(n) = 0, \tag{2.26}$$

which is achieved by modeling impulse responses $h_1(n)$ and $h_2(n)$ as length L FIR filters and grouping all data into the following $L \times 1$ vectors

$$\boldsymbol{x}_i = \begin{bmatrix} x_i(n) \\ x_i(n-1) \\ \vdots \\ x_i(n-L+1) \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{h}_i = \begin{bmatrix} h_i(0) \\ h_i(1) \\ \vdots \\ h_i(L-1) \end{bmatrix}, \tag{2.27}$$

$i = 1, 2$. Furthermore, extended $2L \times 1$ vectors are defined by stacking up the $L \times 1$ vectors defined in (2.27)

$$\boldsymbol{x}(n) = \begin{bmatrix} \boldsymbol{x}_1^T(n) \\ \boldsymbol{x}_2^T(n) \end{bmatrix}, \qquad \boldsymbol{u} = \begin{bmatrix} \boldsymbol{h}_2^T \\ -\boldsymbol{h}_1^T \end{bmatrix}. \tag{2.28}$$

Now, replacing (2.28) into (2.26) results in

$$\boldsymbol{x}^T(n)\boldsymbol{u} = 0. \tag{2.29}$$

Covariances matrix $\boldsymbol{R}_{\boldsymbol{xx}}(n) = E\left\{\boldsymbol{x}(n)\boldsymbol{x}^T(n)\right\}$ is introduced by first multiplying (2.29) on the left by $\boldsymbol{x}(n)$ and then applying the expected value operator, which leads to

$$\boldsymbol{R}_{\boldsymbol{xx}}(n)\boldsymbol{u} = \boldsymbol{0}. \tag{2.30}$$

Vector $\boldsymbol{u}$ in (2.30) might be interpreted as the eigenvector of covariances matrix $\boldsymbol{R}_{\boldsymbol{xx}}(n)$ associated with eigenvalue 0. Therefore, computing vector $\boldsymbol{u}$, which is the same as computing vectors $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, depends on finding one eigenvector of the covariances matrix. As demonstrated in [29], the eigenvector of $\boldsymbol{R}_{\boldsymbol{xx}}(n)$ corresponding to eigenvalue 0 can be obtained by minimizing function $\boldsymbol{u}^T\boldsymbol{R}_{\boldsymbol{xx}}\boldsymbol{u}$ subject to the constraint $\boldsymbol{u}^T\boldsymbol{u} = 1$.

An adaptive algorithm is used for implementing the required minimization. The error signal indicated in Fig. 2.4, obtained using (2.26), is given by

$$e(n) = \boldsymbol{u}^T(n)\boldsymbol{x}(n) \tag{2.31}$$

and should ideally be zero. Cost function

$$J(n) = \frac{1}{2}E\left\{e^2(n)\right\} = \frac{1}{2}\boldsymbol{u}^T(n)\boldsymbol{R}_{\boldsymbol{xx}}(n)\boldsymbol{u}(n) \tag{2.32}$$

can be minimized by an adaptive gradient descent approach. Gradient descent update rule is given by

$$\boldsymbol{u}(n+1) = \boldsymbol{u}(n) - \mu\boldsymbol{R}_{\boldsymbol{xx}}(n)\boldsymbol{u}(n), \tag{2.33}$$

where $\boldsymbol{R_{xx}}(n)\boldsymbol{u}(n)$ is the cost function gradient with respect to vector $\boldsymbol{u}$. The constraint $\boldsymbol{u}^T\boldsymbol{u} = 1$ is taken into account by enforcing updated vector $\boldsymbol{u}(n+1)$ to be normalized. Thus, defining $\boldsymbol{v}(n) = \boldsymbol{u}(n) - \mu\boldsymbol{R_{xx}}(n)\boldsymbol{u}(n)$ and applying normalization, (2.33) becomes

$$\boldsymbol{u}(n+1) = \frac{\boldsymbol{v}(n)}{\sqrt{\boldsymbol{v}^T(n)\boldsymbol{v}(n)}}. \tag{2.34}$$

The algorithm is implemented in frequency domain, following a similar structure as the fast LMS algorithm (see equations (2.19) - (2.24)). Filter coefficients are updated according to the following equations:

$$X_i(m,k) = \sum_{n=0}^{N-1} x_i(mL+n)e^{-j\frac{2\pi}{N}nk}, \qquad i = 1,2, \qquad k = 0,1,...,N-1 \tag{2.35}$$

$$e(m,n) = \frac{1}{N}\sum_{k=0}^{N-1}\left[U_1(m,k)X_1(m,k) + U_2(m,k)X_2(m,k)\right]e^{j\frac{2\pi}{N}nk}, \qquad n = 0,1,...,N-1 \tag{2.36}$$

$$\tilde{e}(m,n) = \begin{cases} 0, & n = 0,1,...,L-1 \\ e(m,n), & n = L, L+1, ..., N-1 \end{cases} \tag{2.37}$$

$$E(m,k) = \sum_{n=0}^{N-1}\tilde{e}(m,n)e^{-j\frac{2\pi}{N}nk}, \quad k = 0,1,...,N-1 \tag{2.38}$$

$$S_{x_ix_i}(m,k) = \alpha S_{x_ix_i}(m-1,k) + (1-\alpha)|X_i(m,k)|^2, \qquad i = 1,2, \qquad k = 0,1,...,N-1 \tag{2.39}$$

$$U_i(m+1,k) = U_i(m,k) - \frac{\mu}{S_{x_ix_i}(m,k)+\epsilon}X_i^*(m,k)E(m,k), \qquad i = 1,2, \qquad k = 0,1,...,N-1. \tag{2.40}$$

Here, a full description of vector $\boldsymbol{u}$ and impulse responses $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ is not required, since the objective is simply to estimate the delay between signals. This is achieved by calculating the difference between negative and positive peaks in $\boldsymbol{u}$, associated with the peaks in $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ respectively (see (2.28)). The strategy to calculate the difference between opposite peaks is based on conveniently initializing

vector $\boldsymbol{u}$, as follows:

$$\boldsymbol{u}_i(0) = \begin{cases} 0, & 0 \leq i \leq \left\lfloor \frac{L}{2} \right\rfloor - 1 \\ 1, & i = \left\lfloor \frac{L}{2} \right\rfloor \\ 0, & \left\lfloor \frac{L}{2} \right\rfloor + 1 \leq i \leq 2L - 1 \end{cases} . \tag{2.41}$$

Therefore, $\boldsymbol{u}$ is initialized with all zeros except for the central element of its first half. This is equivalent to initializing the impulse responses as $-\boldsymbol{h}_1(0) = [0 \;\; 0 \ldots 0 \ldots 0]^T$ and $\boldsymbol{h}_2(0) = [0 \;\; 0 \ldots 1 \ldots 0 \;\; 0]^T$. One positive peak stays fixed in the middle of vector $\boldsymbol{u}$ first half, associated with the direct path in $\boldsymbol{h}_2$, while a negative peak evolves in the second half, associated to the direct path in $-\boldsymbol{h}_1$. Delay estimate is finally obtained by calculating the difference in the position of positive and negative peaks.

Chapter 3

Extension: Two-Axle Vehicle Tracking

A strategy for taking into consideration both dominant sources observed in car pass-by recordings is presented in this chapter. The proposed system is based on extending single-source DOA estimators by appending further processing stages and thus adapting them for traffic noise applications.

Before getting into the extension details, let us illustrate the consequences of using original DOA estimators for car noise emissions. Grayscale images on the background of Fig. 3.1 represent, on the vertical axis, the generalized cross-correlation function for GCC, the histogram map for ITD, and the evolution of filter coefficients for LMS and EVD methods. The horizontal axis represents sequential time frames in all four images. These different functions generate quite similar images, in which two distinguishable peak regions are visible. Tire noise dominates car emissions for most typical car speeds and the two visible regions are associated with the noise emitted by front and rear tires. This assumption regarding the dominant sources is supported by the results in [19]. A pair of loudspeakers was attached to vehicle wheels during pass-by experiments carried out in the referred work and the calculated cross-correlation function presented a similar pattern as observed in Fig. 3.1. When the original algorithms are used, which assume a single sound source and compute only one TDOA estimate per frame, the delay estimates results, represented by a blue line in Fig. 3.2, are obtained. The dominant sound source alternates between the front and rear axles during car pass-by, and so does the estimate.

An extended system is proposed as a solution to the noisy and inaccurate estimates delivered by traditional DOA estimators in car noise applications. The extra processing stage is indicated in Fig. 3.3 named as “Two-source Extension" and, as depicted, it is appended to one of the single-source estimators represented in “Single-source Estimation" block. Despite exhibiting distinct approaches and implementations, all four single-source estimators share a few essential similarities which

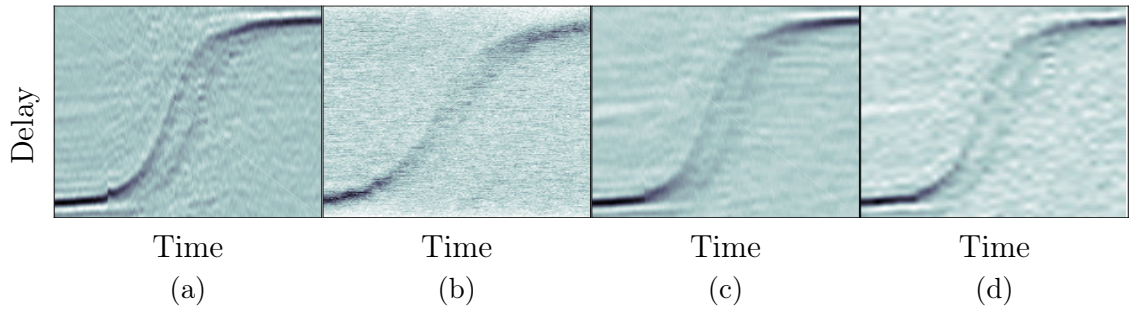

Figure 3.1: Images generated using (a) GCC, (b) ITD, (c) LMS and (d) EVD algorithms.

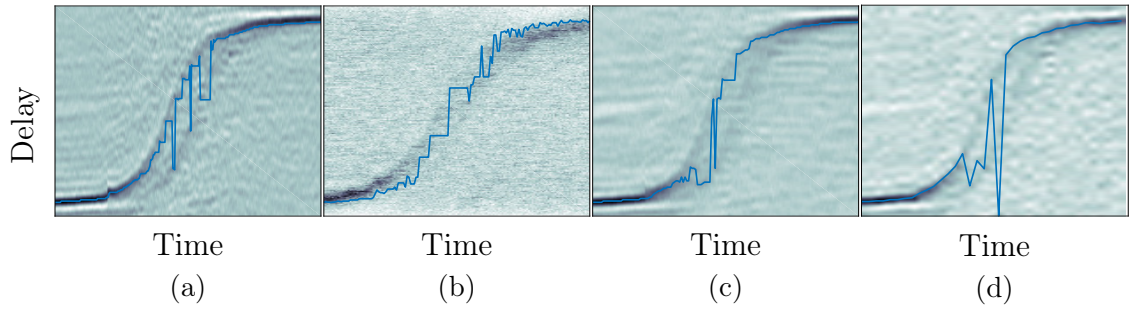

Figure 3.2: Time delay estimated by original algorithms: (a) GCC, (b) ITD, (c) LMS and (d) EVD.

are exploited here, in order to make them interchangeable in the proposed system. All of them use two sensor signals as input and, besides, all four keep track of sequential TDOA estimations by accumulating short-time calculations into a matrix of dimension $D \times M$, where D is the number of discrete values used to represent the delay and M is the number of time frames. Thus, "Single-source Estimation" stage might be interpreted as a "black box" in which a 2-channel audio recording enters and a $D \times M$ matrix comes out. Whether this matrix used to represent a cross-correlation function, a histogram map or the evolution of filter coefficients inside the black box, it is nothing more than a generic matrix $\boldsymbol{A} \in \mathbb{R}^{D \times M}$ for the remaining system.

Delay matrix $\boldsymbol{A}$ is a critical piece of our system, as it provides information about how likely each delay is to represent the actual time delay of arrival. The amount of noisy data in $\boldsymbol{A}$ and the sharpness of its peaks depend on the chosen DOA estimator and on the parameters used to set it, such as window length, forgetting factor and adaption step size. Algorithms sensitivity to these parameters was investigated in a previous work [30]. If DOA estimators fail in calculating matrix $\boldsymbol{A}$, the post-processing performed in the extension is also fated to fail and to produce inaccurate TDOA estimates.

Original estimators deliver a single DOA estimated curve as output, obtained by accumulating maximum value evaluations for each time frame, as illustrated in Fig.

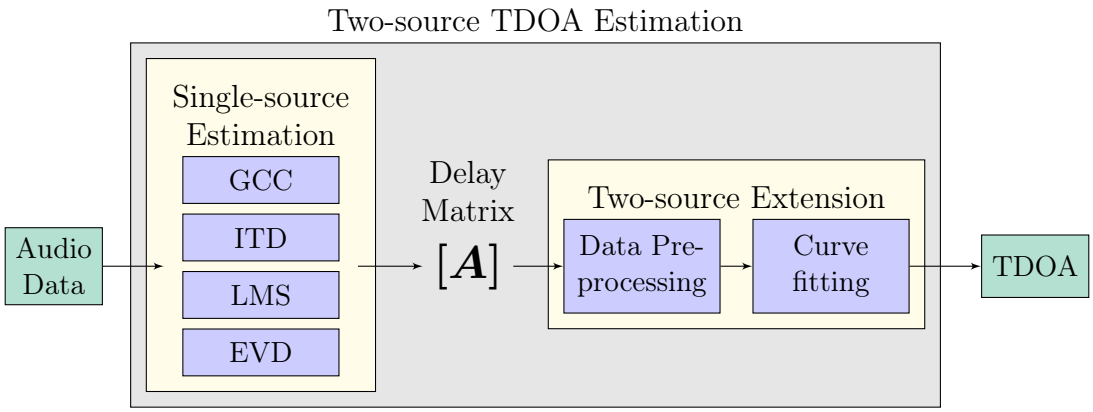

Figure 3.3: Schematic diagram of the proposed two-source TDOA estimation system.

3.2. Single-source estimates are ignored by the extended system and replaced by the two stages indicated in the diagram: pre-processing and curve fitting. Matrix $\boldsymbol{A}$ is first treated as an image and processed for noise removal. Next, pixels from this filtered image are used as input data for curve fitting. Besides data points, curve fitting algorithm must be fed with a curve model. A rough TDOA model is derived in agreement to pass-by dynamic and geometric characteristics, as shown in Section 3.1. Moreover, stages comprised in “Two-source Extension" block are explained in Sections 3.2 and 3.3.

3.1 Curve Model

Time difference of arrival evolution is predicted by calculating sound path difference between the source and two microphones separated by a distance d.

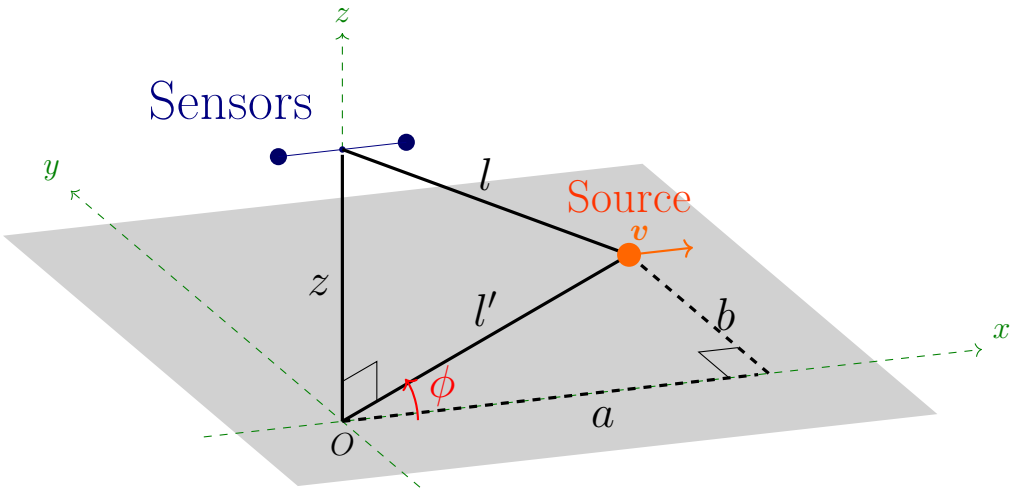

Figure 3.4: Pass-by 3D representation. Height z measured between the floor and the center of the array. Distance l between the source and one of the microphones.

A three-dimensional pass-by model is pictured in Fig. 3.4, from which is possible

to extract distance l between the source and one of the sensors. Source is assumed to lie in $z = 0$ plane, thus it is assumed to be at pavement level, while the microphone array belongs to $y = 0$ plane. The following calculations are performed considering two microphones at the same x-coordinate, although a similar derivation might be carried for other configurations. Distance b, between $y = 0$ plane and the line containing the vehicle path, height z, measured from the floor to the array center, and vehicle speed $\boldsymbol{v}$, assumed to be constant, represent the known, measurable and time-invariant data. A pass-by snapshot for an arbitrary time instant is represented in Fig. 3.4, but the reader should keep in mind that a and l are changing over time, as indicated in the following time dependent equations.

Vehicle velocity, besides a constant magnitude, is assumed to have its direction parallel to x-axis, or, in other words $\boldsymbol{v} = \boldsymbol{v_x}$. Moreover, velocity orientation is assumed to be the same as x-axis orientation and thus magnitude $v_x = |\boldsymbol{v_x}|$ is used throughout the text. This simplification results in no loss in generality since a speed in the opposite direction might be represented by reversing v_x sign. Doppler effect is neglected in this derivation as the sound speed is assumed to be much larger than the car speed. Horizontal component of the source position, $a(t)$, is calculated by

$$a(t) = (t - t_0)\, v_x \tag{3.1}$$

where t_0 represents the instant in which the vehicle crosses y-axis. Distances in Fig. 3.4 are equated by straightforward trigonometric relations

$$l'^2(t) = b^2 + a^2(t) \tag{3.2}$$

$$l^2(t) = z^2 + l'^2(t) \tag{3.3}$$

and, replacing (3.2) in (3.3):

$$l^2(t) = z^2 + b^2 + a^2(t). \tag{3.4}$$

Equation (3.4) computes the distance between the source and a sensor located in the array center, where $x = 0$. Thus, the distance to a sensor located at $x = -d/2$ and $x = +d/2$ is, respectively

$$l_1^2(t) = z^2 + b^2 + (a(t) - d/2)^2 \tag{3.5}$$

$$l_2^2(t) = z^2 + b^2 + (a(t) + d/2)^2. \tag{3.6}$$

Time delay of arrival is finally calculated by

$$\tau = \frac{l_2(t) - l_1(t)}{v_x}$$
$$\tau = \frac{\sqrt{z^2 + b^2 + (a(t) + d/2)^2}}{v_x} - \frac{\sqrt{z^2 + b^2 + (a(t) - d/2)^2}}{v_x} \tag{3.7}$$

When $v_x > 0$ and the vehicle moves along x-axis positive direction, $a(t)$ is negative for $t < t_0$ (see (3.1)), resulting in $l_1(t) > l_2(t)$ (Eqs. (3.5) and (3.6)). In other words, TDOA is negative ($\tau < 0$) while the source approaches the array, and it is positive ($\tau > 0$) while the source moves away from the array, passing through zero delay when the vehicle is exactly in front of the array. This behavior can be observed in Fig. 3.5, where the theoretical curve (3.7) is plotted for $v_x = 60\,\text{km/h}$ and $t_0 \approx 1.5\,\text{s}$.

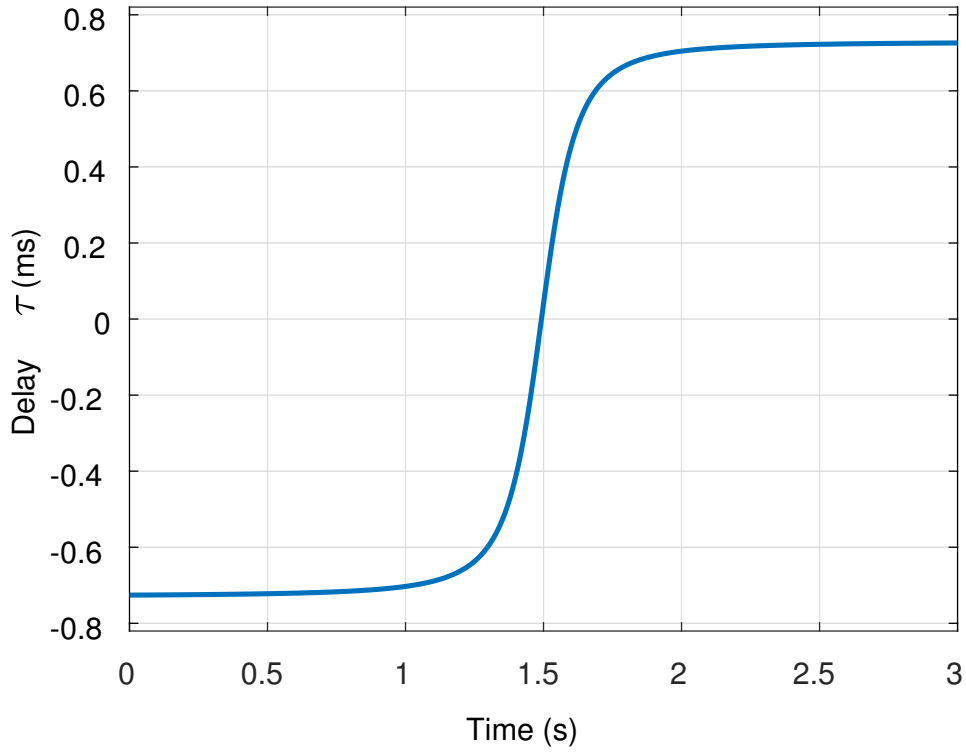

Figure 3.5: TDOA model for a vehicle moving at 60 km/h.

3.2 Data pre-processing

Data scaling and noise reduction are implemented in order to prepare data for curve fitting. The system is required to work regardless of which TDOA estimator is used. All single-source estimators deliver a frame by frame estimate obtained by maximizing matrix $\boldsymbol{A}$ columns. However, distinct strategies are used by each method in order to generate matrix $\boldsymbol{A}$ and, as a result, data values might be at different

scales depending on the chosen algorithm. Data must then be scaled as soon as it enters the extension block. Besides adjusting the range of values to be handled by the algorithm, data normalization allows the two-dimensional data structure to be treated as a digital image representation, with values within grayscale range. This image will later be processed by image filters.

Due to sound attenuation with increasing distance, relevant audio data concentrates around $t = t_0$, when source-receiver distance is minimum and the signal-to-noise ratio is maximum. Data scaling is thus followed by data cropping, which is intended to discard irrelevant portions of audio data, acquired when the vehicle was far away from the microphones. A 3 s data window is selected around the sample corresponding to $t = t_0$, and the remainder is ignored.

In a real scenario, interfering noise generated by other sources, rather than the vehicle, might corrupt the recordings. Depending on the interfering noise intensity, spectral content and position, such interference might severely degrade curve fitting performance. For this reason, input data is filtered as an attempt to reduce noise.

First, grayscale image is converted to binary by use of a threshold value, selected based on empirical tests. Distinct threshold values are obtained for each DOA estimator, which is consequence of pixel distribution differences in generated images. Histogram plots in Fig. 3.6, calculated for Fig. 3.1 images, illustrate pixels distribution over grayscale level for each estimator. Low standard deviation is observed in GCC- and LMS-generated images, corroborating the distinguishable peaks in Figures 3.1a and 3.1c. In addition, the high standard deviation observed in ITD histogram plot matches the smooth transitions in 3.1b. A suitable threshold value choice is therefore highly relevant. Binary images generated using the same algorithm (GCC) and different threshold values are depicted in Fig. 3.7. A trade-off between removing noise and keeping enough relevant data is observed. Similar results would be observed on images generated by the other algorithms.

Next, morphological opening (dilatation followed by erosion) [31] is performed on the binary image using a squared structuring element. Opening is applied in order to remove undesired foreground pixels, represented by dark pixels in Fig. 3.1. This helps filling “silent" regions on the image with zeros and therefore removing noise.

Typical results from the mentioned image processing steps might be observed in Fig. 3.8, in which a GCC-generated image is used as example. First, a grayscale image obtained after data normalization and cropping is illustrated in Fig. 3.8a. This image is then converted to the binary image in Fig. 3.8b using a threshold value of 0.1. Next, morphological opening is applied, resulting in Fig.3.8c image, from which most noisy pixels were eliminated.

If the image processing steps described above are successful, at this point the

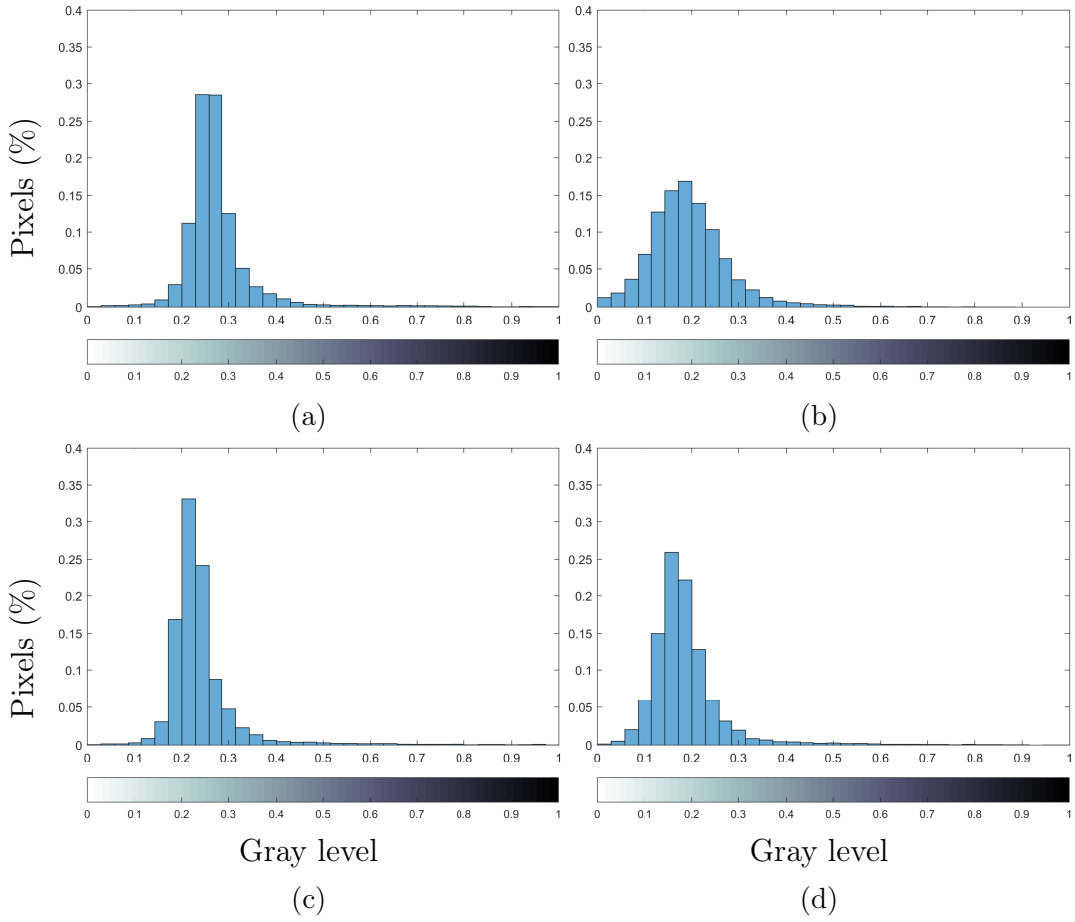

Figure 3.6: Histogram plots of images generated using (a) GCC, (b) ITD, (c) LMS and (d) EVD.

image should display two distinguishable curves which indicate the presence of two dominant noise sources. Time-shift between curves is equivalent to the average wheelbase in passenger cars and is assumed to be result of the noise emitted by vehicle front and rear axles.

Curve fitting optimization requires data points and a model to fit the data. Assuming identical front and rear tires, the noise emitted by each axle should, ideally, differ only by a delay. The same model is thus used to model sound emissions from both axles. Data points, on the other hand, must be selected from the filtered image and assigned to one of the car axles. Thus, before curve fitting, data selection is performed by separating data points from image $\boldsymbol{A}$ into two vectors, one for each car axle. For each column in the binary image, line indices holding non-zero data values are stored in a vector, from which the mean value is calculated. The mean value of all columns gathered in sequence results in a mean curve of non-zero-data indices, indicated by the red line in Fig. 3.8d. This mean curve is used as border line for separating data into two vectors. All data lying on the left (or above) of the mean curve is assigned to the left curve, as represented in blue in Fig. 3.8e. The left

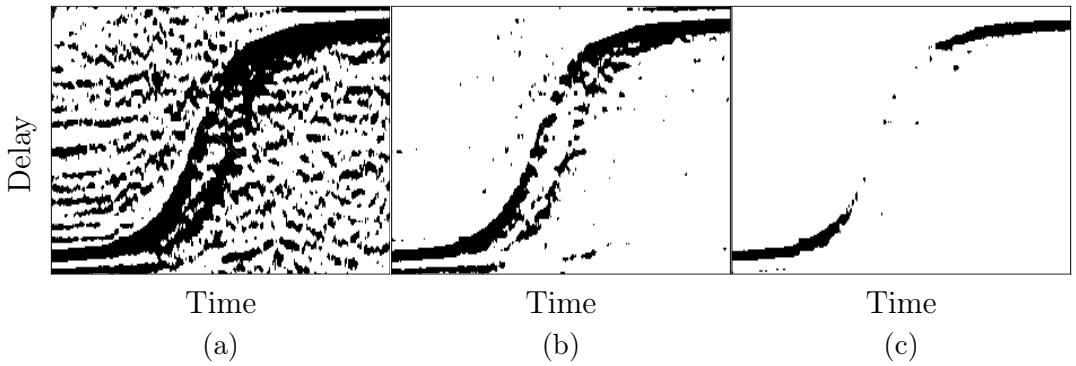

Figure 3.7: Binary images generated using GCC algorithm and different threshold values: (a) 0.025, (b) 0.1 and (c) 0.25.

curve appears ahead in time and is therefore associated with the front-axle emission. Similarly, data lying on the right (or below) of the mean curve is associated with the rear-axle emission and displayed in orange in 3.8e. These two data vectors, corresponding to the left and right curves, are delivered to the curve fitting stage.

Tracing one mean curve to separate data carries the assumption of two dominant sources being present in the audio data, and thus of two axles being present in the vehicle. The idea could be extended to more axles by iteratively computing mean curves for the subdivided spaces, until the number of axles in the vehicle is reached. This would allow thw system to be applied for heavier vehicles, which also play a relevant role in traffic noise.

3.3 Curve fitting

Curve fitting algorithm must be provided with (a) a TDOA curve model, (b) two selected data vectors, and (c) a initial value for curve parameters. Items (a) and (b) are obtained as indicated in Sections 3.1 and 3.2, respectively. Unknown variables in (3.7) are instant t_0, distance b and speed v_x . These curve parameters can be initialized using measured data, if there is any available, or random values. Instant t_0 corresponds to the time when the vehicle is in front of the microphone array, and therefore the signal energy at the receivers is maximum. Thus, an estimate $\hat{t}_0$ is calculated based on the maximum energy level and used for initialization. Signal Power Spectral Density (PSD) is computed and summed over frequency bins. The instant corresponding to the higher summed PSD is defined as $\hat{t}_0$ and used to initialize curve parameter t_0.

Given the nonlinear TDOA curve model (see 3.7), curve fitting is performed using trust-region algorithm for nonlinear optimization [32–35]. Moreover, spurious data might withstand pre-processing and appear as outliers in selected data. This is specially harmful given the known sensitivity of least-squares algorithms to outliers.

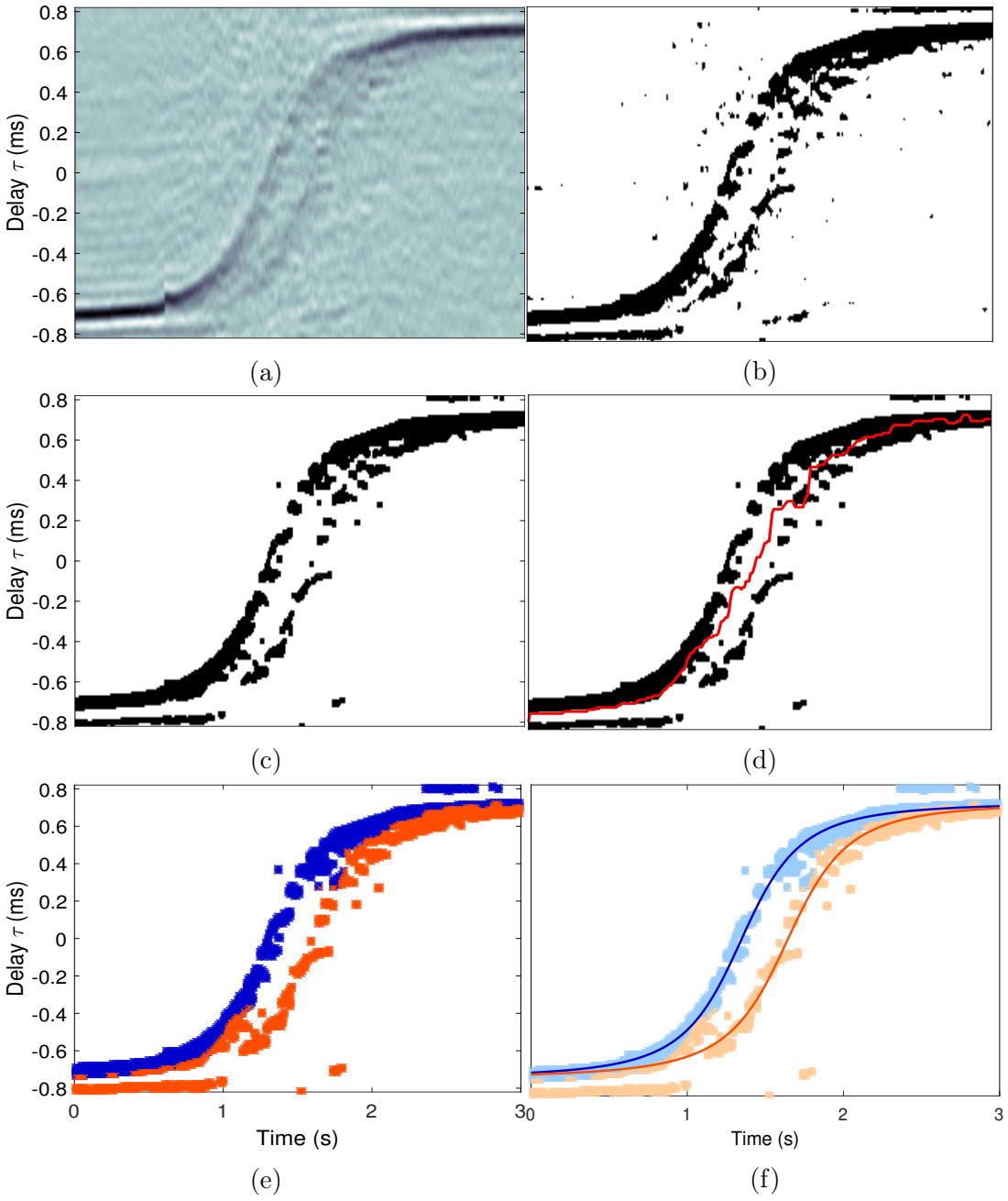

Figure 3.8: Pre-processing and curve fitting steps. (a) Grayscale image. (b) Binary image. (c) morphological opening output. (d) Mean curve in red. (e) Separated data and (f) fitted curves.

In this regard, a robust least-squares approach is preferred over the traditional one. Cost function is iteratively re-weighted using bisquare weights [36, 37].

Curve fitting results in optimized values for model parameters, which minimize the squared error between fitted curve and data points. Speed, as one of the model variables of (3.7), is directly estimated by this optimization. Curves corresponding to front and rear axles are fitted independently and the estimated parameters, including speed, might be slightly different for each axle. For simplicity, speed estimate is defined as the average between both optimized values. Wheelbase, conversely, is

estimated as the distance between the fitted curves. This distance is evaluated for delay $\tau = 0$, when the midpoint between axles is exactly in front of the microphone array. A typical result for the adjusted curves is observed in Fig. 3.8f. Data points are maintained in this figure, in lighter colors, and adjusted curved are pictured in blue and orange dark solid lines, for front and rear axles respectively.

Chapter 4

Experimental Results

Experimental tests were carried out for validating the proposed system and evaluating its performance. A comparison between DOA estimators is performed in order to indicate which could be used in traffic noise applications. Details regarding data acquisition and test results are also presented next.

Pass-by noise measurements are registered by two microphones 0.25 m apart. The linear array formed by these two sensors is placed with its axis parallel to both the ground and the vehicle path. The top view of this setup is illustrated in Fig. 4.1.

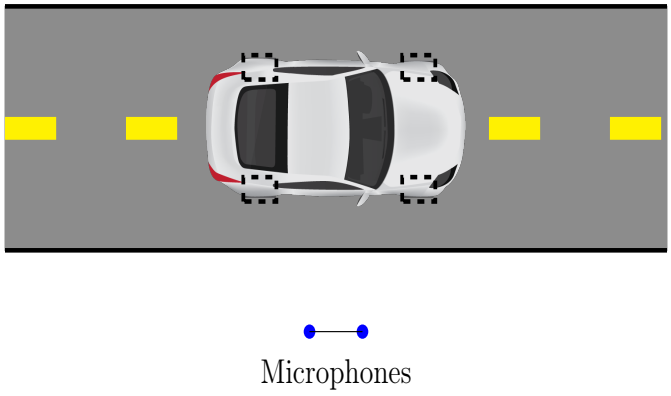

Figure 4.1: Relative positioning of array and vehicle path.

Four passenger cars were used in the experiment trials, as detailed in Table 4.1, as well as three different drivers. Each trial consisted in one car passing by the microphone array at a constant speed of around 30, 50, 60 or 70 km/h, or accelerating. Vehicle speed was tracked by a GPS equipped device placed inside the vehicle. Previous experiments had indicated that tire noise prevails over other noise sources in constant speed pass-by tests, hence speeding up tests were carried out as an attempt of observing other noise sources. Pass-by trials took place in a quiet location at the Brazilian National Institute of Metrology (INMETRO), resulting in low background noise level composed mainly by birds and wind. Moreover, video recordings were

captured by a cell phone camera fixed at the array center. Later, these recordings were analyzed in order to extract estimates for distance b, between car and array, which might be used for initializing parameters in curve fitting optimization.

Table 4.1: Vehicles used in pass-by tests

Car ID	Car Model	Transmission	Wheelbase (m)
1	Volkswagen Gol 1.0	Manual	2.47
2	Jeep Renegade 1.8	Automatic	2.57
3	Mitsubishi ASX 2.0	Automatic	2.67
4	Hyundai Creta 1.6	Automatic	2.59

Multi-source localization is often approached using array processing techniques [38–41], with arrays that might scale up to hundreds of elements [42, 43]. These methods might deliver good estimations, but they require a complex and costly implementation, not to mention the higher computational cost resulted from using many input signals. A two-microphone setup is proposed here as a simpler and cheaper alternative to the large-array approach. The inner distance of 0.25 m between sensors has shown to result in images with the sharpest peaks among sensor distance options in [44].

A subset of 23 out of the total 28 pass-by recordings is selected to be used in the following tests. Selection criteria is based on vehicle speed, which is assumed constant in the proposed system, and therefore only constant speed pass-by trials are selected. Information regarding each of the 23 selected audio tests is presented in Table 4.2. This table contains the car identification and speed in each test, and whether the gear was shifted up as the car speed increased or a low gear was imposed. Three parameters are chosen for evaluating the proposed method: a visual analysis of how delay estimates fit matrix $\boldsymbol{A}$ peaks, and error analysis of both wheelbase and speed estimates.

As mentioned previously, curve fitting optimization requires parameter initialization. Instant t_0 is initialized using maximum energy based estimate $\hat{t}_0$, as discussed in Section 3.3. For the tests carried out here, there is measured data available that could be used to initialize remaining parameters b and v_x. However, the proposed system is intended to work even when there is no prior knowledge regarding vehicle speed and distance, and thus initialization uses random values from a normal distribution.

Implementing the proposed method also requires setting the parameters used by each of the four direction of arrival estimators. The values used in the present implementation, indicated in Table 4.3, were empirically chosen based on the delay

Table 4.2: Pass-by tests characteristics: identification number, car speed for $t = t_0$, car model identification and whether the gear was shifted according to the car speed or fixed at a low gear.

Test ID	Speed (km/h)	Car ID	Gear Shifts
1	25	3	Free
2	28	4	Free
3	30	2	Fixed at 2^{nd} gear
4	30	2	Fixed at 2^{nd} gear
5	30	3	Free
6	30	4	Free
7	31	1	Free
8	31	2	Fixed at 1^{st} gear
9	37	1	Free
10	46	4	Free
11	47	1	Free
12	47	4	Free
13	49	3	Free
14	49	3	Free
15	51	2	Free
16	54	2	Free
17	59	1	Free
18	60	4	Free
19	61	2	Free
20	63	4	Free
21	64	3	Free
22	66	2	Free
23	70	3	Free

matrices obtained for the pass-by audio recordings. There is no guarantee, however, that these values are optimum for this application. Window size N' is scaled according to the expected car speed by setting a maximum displacement of the vehicle between two consecutive blocks (without overlapping). More details on the selection of these parameters and their influence on the delay estimation can be found in [16, 30].

4.1 Speed Estimation

Speed estimates obtained using the four single-source estimators are displayed in Fig. 4.2, along with GPS-measured values. Test results were sorted in ascending order of the measured speed values for visualization purposes. This visualization choice highlights how speed estimation performance degrades when vehicle speed increases.

Table 4.3: Delay estimation parameters

Method	Parameters	
GCC	*Hamming*	# Window function
	$N = N'$	# Window length
	$N_{OV} = 75\%$	# Overlap between consecutive windows
	$\alpha = 0.7$	# Forgetting factor
	$OV = 4$	# Oversampling factor
	$th = 0.10$	# Threshold
ITD	*Hamming*	# Window function
	$N = 4N'$	# Window length
	$N_{OV} = 87.5\%$	# Overlap between consecutive windows
	$I = 180$	# Number of azimuth angle sectors
	$th = 0.25$	# Threshold
LMS	*Rectangular*	# Window function
	$N = N'/2$	# Window length
	$N_{OV} = 50\%$	# Overlap between consecutive windows
	$\mu = 0.25$	# Adaptive step size
	$th = 0.15$	# Threshold
EVD	*Rectangular*	# Window function
	$N = N'/2$	# Window length
	$N_{OV} = 50\%$	# Overlap between consecutive windows
	$\mu = 0.25$	# Adaptive step size
	$th = 0.25$	# Threshold

This behavior might be observed for all methods but it is especially evident for ITD. The absolute speed estimation error for each method is displayed in Table 4.4.

Table 4.4: Speed average absolute error, in km/h, for all estimators and different speed ranges.

Speed range	GCC	ITD	LMS	EVD
All	10.82	17.08	10.01	11.36
≤ 50 km/h	6.13	10.26	6.81	8.30
> 50 km/h	18.11	27.68	11.23	16.04

If all 23 pass-by tests are considered, LMS presents the smallest estimation error, followed closely by GCC. Instead, if only low speed tests are considered (below 50 km/h), all estimators deliver lower absolute errors. In this scenario, LMS and GCC still provide the smallest errors, but GCC error is slightly smaller. Finally, taking into account speeds above 50 km/h, absolute error increases for all methods, in especial for ITD. However, the lowest error is now observed for LMS, followed by EVD. This analysis thus suggests a superior performance of adaptive methods for higher vehicle speeds. Adaptive methods require less assumptions regarding sound

propagation, as they are based on modeling the propagation channel, and this might cause the performance difference. On the other hand, ITD poor speed estimation performance for all speed ranges is a result of ITD-generated images showing no sharp peaks, as already observed on histogram plots (see Fig. 3.6).

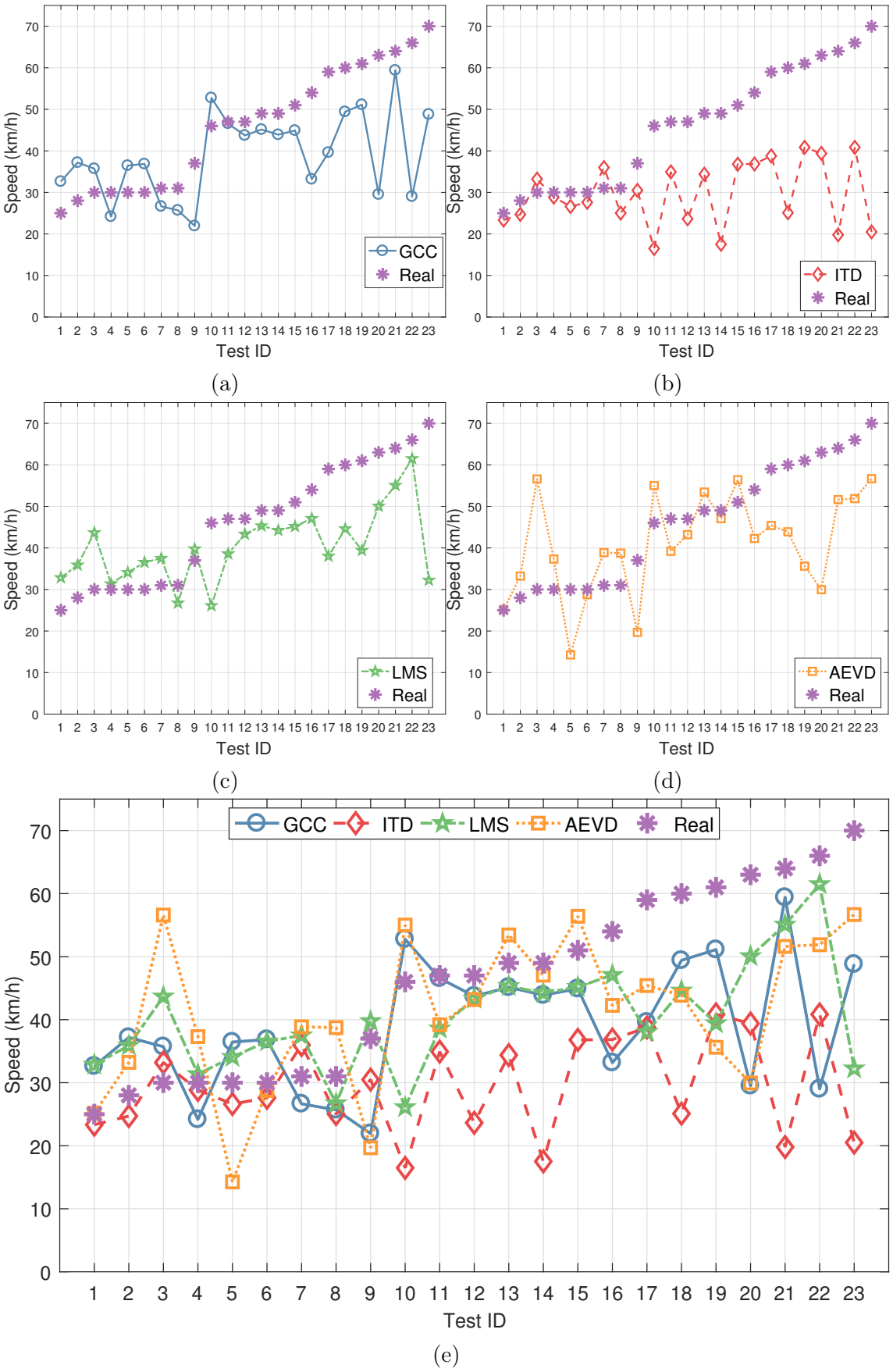

Figure 4.2: Speed estimates obtained by curve fitting using (a) GCC, (b) ITD, (c) LMS and (d) EVD generated data. All estimators compared in (e).

Another similarity observed for all estimators is a tendency to underestimate the vehicle speed. One possible reason is the simplistic model used for TDOA, which apparently is not able to accurately describe the expected delay behavior. Time delay estimate for Test 23, which resulted in the higher speed estimation error for LMS, is presented in Fig. 4.3. Despite the large estimation error, curves are well adjusted to image peaks corresponding to the delays of the vehicle axles. This result indicates the error is not caused by the fitting optimization, but by the chosen theoretical model.

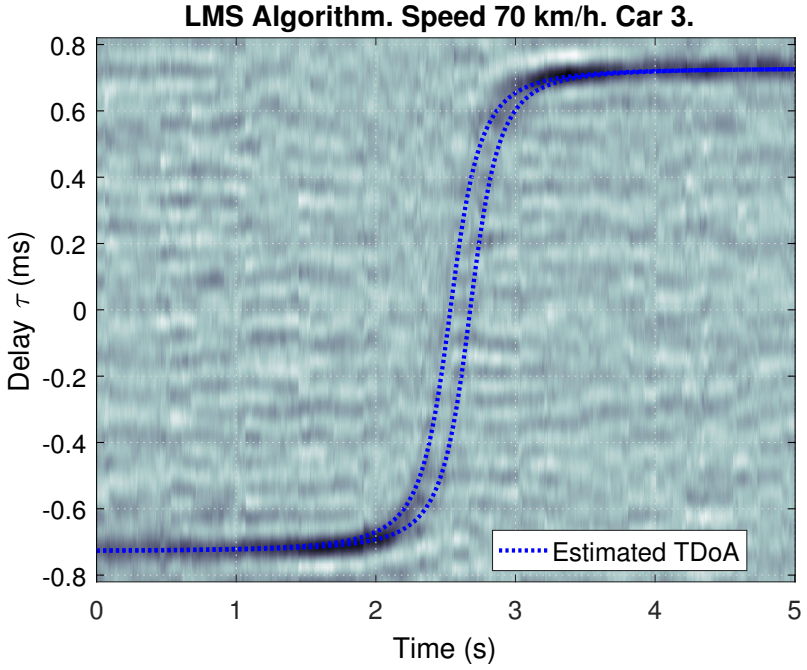

Figure 4.3: Tracking results for the highest-speed-error test obtained with LMS estimator.

4.2 Wheelbase Estimation

Wheelbase estimation results are presented in Fig. 4.4 along with real values, for the experiments in ascending speed order. The reference wheelbase values provided by the manufactures can be found in Table 4.1. Regardless of speed, average wheelbase estimation error is higher when ITD is used as single-source delay estimator, as indicated in Table 4.5. The lowest average error is obtained using GCC, when all speed values or only speed values below 50 km/h are considered, and LMS, for speed values above 50 km/h. Wheelbase estimation results obtained using GCC are comparable to the ones in [19]. Our system delivered an estimation error lower than 30 cm in 78% of the cases, while Marmaroli *et al.* [19] approach, in a slightly better performance, did the same in 86% of the cases.

Table 4.5: Wheelbase average absolute error, in meters, for all estimators at different speed ranges.

Speed range	GCC	ITD	LMS	EVD
All	0.26	1.04	0.38	0.81
$\leq 50\,\mathrm{km/h}$	0.30	1.10	0.55	0.98
$> 50\,\mathrm{km/h}$	0.20	0.96	0.12	0.55

Wheelbase estimation error is, on average, higher for lower speeds than for higher speeds, as it can be observed for all estimators in Table 4.5. Therefore, wheelbase estimation assumes an opposite behavior with respect to increasing vehicle speed to the one observed for speed estimation. There is no agreement in the literature concerning the speed above which tire noise prevails over other vehicular sound sources, ranging from values between 30 km/h and 50 km/h[17, 21, 45]. However, other noise sources clearly become less relevant for higher speed values. In this light, and given that the wheelbase estimation performed here depends on localizing vehicle wheels, the larger estimation error obtained for lower speeds is consistent with the higher difficulty level in distinguishing tire noise emissions from the noise emitted by other vehicle components.

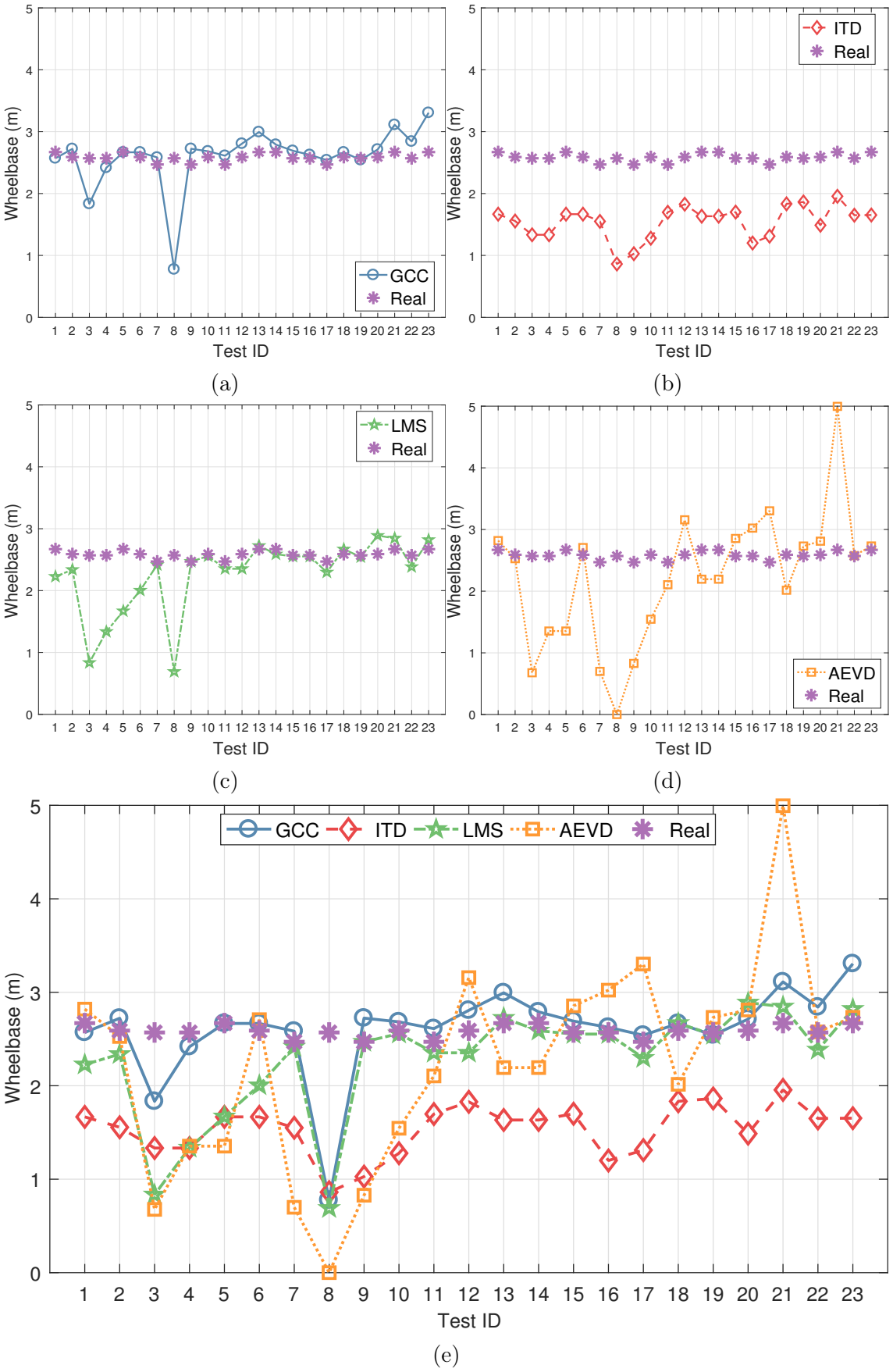

Figure 4.4: Wheelbase estimates obtained by curve fitting using (a) GCC, (b) ITD, (c) LMS and (d) EVD generated data. All methods compared in (e).

Let us now focus on GCC and LMS estimation results, which delivered the lowest estimation errors for wheelbase and speed. Wheelbase estimates for Tests 3 and 8 stand out in Figures 4.4a and 4.4c, whose absolute errors are far from the average. Besides the low vehicle speeds, of around 30 km/h, these two tests share further similarities which helps explaining the high estimation error. A same vehicle, identified as Car 2 in Table 4.1, was recorded in both audio files used as input to Test 3 and 8. Although Car 2 was recorded in several other trials, many of which resulted in low estimation errors, in these two specific pass-by trials the vehicle was conducted in a different manner. Vehicle 2 was forced to travel in low gear during pass-by trials 3 and 8, which forced the vehicle engine into a high revolutions per minute (RPM) operation. Engine speed affects engine noise emissions [15] and forcing higher RPMs might cause engine noise to increase its relevance in comparison to tire noise. As mentioned before, the proposed delay estimation system assumes tire noise as dominant sound sources and, if this condition fails to be true, the estimation will also fail.

4.3 Performance Comparison

Estimation error analysis provides a basis of comparison between the direction of arrival estimators tested in this work. The estimators were evaluated concerning the estimated speed and wheelbase values and for different vehicle speed ranges: all speeds, low speeds (below 50 km/h) and high speeds (above 50 km/h). The methods which delivered the lowest estimation errors in each of these scenarios are indicated in Fig. 4.5.

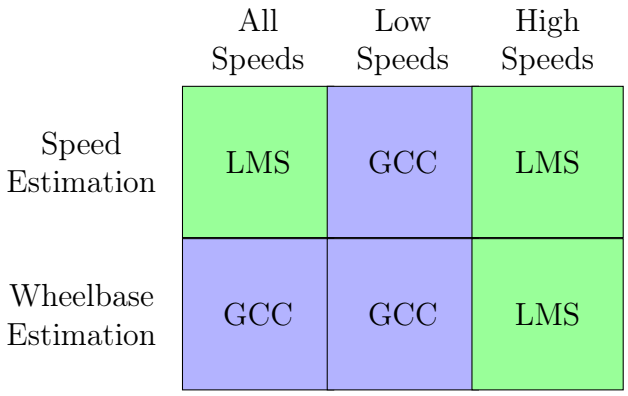

Figure 4.5: Best performing algorithms for each evaluated scenario.

The estimators GCC and LMS display a similar performance as each of them presents the lowest estimation error for half of the evaluated scenarios. However, these six possible scenarios should not be given the same relevance in evaluating the proposed system. The main objective of the present work is to estimate the time

delay of arrival of the sound emitted by vehicle tires and this should be taken into account for deciding the best algorithm.

Speed estimate is a model parameter optimized during curve fitting and it influences the slope of the delay curves. However, speed estimation has presented high errors even when delay estimates were well adjusted to data (see Fig. 4.3), indicating the delay model does not accurately describe the actual delay behavior. Estimated speed values are thus not reliable for representing the system performance.

Wheelbase estimate, on the other hand, is computed based on the distance between the adjusted curves for $\tau = 0$, when the curves cross the horizontal axis. Thus, accurate delay estimation implies in accurate wheelbase estimation. The inverse statement is not true, since a pair of delay curves might cross $\tau = 0$ axis with the same time distance with infinite different slopes. This would be the equivalent, for example, to the delay estimates of a same vehicle traveling at distinct speeds. Nevertheless, wheelbase estimation has shown to be better than speed estimation as indicator for the system performance with respect to time delay of arrival estimation.

Wheelbase estimation error is defined as decision criteria for choosing the best direction of arrival estimator. The results summarized in Fig. 4.5 indicate GCC algorithm is the most suitable method to be used in car noise applications.

Time delay of arrival estimation results for all tests and estimators can be found in Appendix A. Adjusted curve results corroborate the assumed correlation between wheelbase and time delay estimation. Similarly to wheelbase estimation, GCC and LMS display the best results, ITD sub estimates the distance between axles and AEVD delivers unstable estimates.

Chapter 5

Conclusion

An extended direction of arrival estimation system was proposed in this work, by introducing post-processing stages to the traditional one-source direction of arrival methods. This modification aimed to adapt the estimators to be used in applications with multiple sources. In the traffic noise scenario, which is the focus of this dissertation, the original methods fail to provide accurate estimates of the directions of arrival of the main sound sources. The system developed in this work uses the signals of a two-microphone array as inputs and provides time-difference of arrival estimates for the two dominant sources, associated with the vehicle's front and rear tires.

The system consisted in two main blocks. First, a conventional estimator is used to produce a delay matrix, which indicates, for each time frame, delay values with the greatest potential to represent the actual delay between the signals; second, the delay matrix is processed by image filters and the filtered data is used to feed a curve fitting algorithm. The optimization of the curve fit takes into account two dominant sources, separating the data received into two vectors, one for each source. Only vehicles with two axles were considered, and an extension to more axles can be achieved by modifying the approach used to select the data that fits each curve.

Four direction of arrival estimators were tested to compose the system, namely: the Generalized Cross-Correlation (GCC), Interaural Time Difference (ITD), Least Mean Square (LMS) and Eigenvalue Decomposition (EVD), where the last two are adaptive methods based on channel estimation. The vehicle speed and wheelbase estimates were compared to the actual measured values, revealing that the GCC and LMS were the most appropriate choices among the tested algorithms. The vehicle's wheelbase was estimated with an average error as low as 26 cm, obtained by the GCC, and the speed was estimated with a minimum average error of 10.01 km/h, using the LMS. However, the wheelbase estimate proved to be more suitable to be used as an evaluation criterion than the speed estimate, and thus it was concluded that the Generalized Cross-Correlation was the best algorithm to be used in the

vehicle noise TDOA estimates.

Results were promising, given that a simple array of only 2 microphones was required to implement the system. This method is an alternative to the complex array strategies, often found in the literature as a solution for handling multiple sources. Regarding a similar approach found in the literature [19], our system presented a comparable, but slightly worse in estimating the wheelbase.

Regardless of the chosen algorithm, speed estimation performance was better when the vehicle speed was low (below 50 km/h), as opposed to the wheelbase estimation, which presented more accurate results for higher speeds (above 50 km/h). In addition to the vehicle speed, the engine's operating regime also influenced the performance of the estimator. Engine noise becomes more relevant when the engine speed is higher, and this can prevent tire noise from dominating car noise emissions, as has been assumed. If the engine noise exceeds or corresponds to the tire noise, the proposed system cannot track the position of the tires or estimate the wheelbase of the vehicle. Given that tracking the axles is the main objective of this work, instead of estimating the speed of the vehicle, the system performance is optimum when the vehicle speed is at least 50 km/h and the engine speed is low.

Improvements to the system can be achieved by modifying either the first or second main processing block. In the first block, extra direction of arrival estimators can be tested to improve the calculation of the delay matrix. The modular structure of the system makes it possible to easily replace the estimator without major changes in the rest of the system.

In the second block, a great improvement to the system would be achieved by automating the selection of the threshold value. An adequate choice of this parameter is crucial to provide sufficient relevant data for the curve fitting algorithm and, therefore, greatly influences its result. Empirical optimization, as performed here, is subject to imprecision and is impractical if new estimators need to be tested.

Appendix A

Adjusted Curves

Adjusted curves corresponding to the time delay of arrival estimates for all tests and all four estimators are displayed in Figures A.1-A.3. For spacing limits, results are divided into three figures and each figure is divided into four columns and at most 8 rows. Images generated using GCC, ITD, LMS and EVD are placed in the first to fourth columns, respectively. Numbers on the left margin inform tests identification in the same order as in Figures 4.2 and 4.4. Each individual result is represented by a square in which matrix $\boldsymbol{A}$ and both fitted curves are displayed. Matrix $\boldsymbol{A}$ is represented by grayscale pixels and TDOA adjusted curves are depicted as blue solid lines. Axles are omitted from individual results for spacing issues, but similarly to Fig. 4.3, time is represented in the horizontal axis and delay in the vertical axis.

Wheelbase estimate is calculated based on the distance between each pair of fitted curves. These two estimations, of wheelbase and time delay of arrival, are thus expected to be correlated. In agreement to the wheelbase estimates in Fig. 4.4, GCC and LMS display similar results, with well adjusted curves for the majority of tests. One exception is observed for Test 8, displayed on the 8th row of Fig. A.1, in which fitted curves are estimated too close together and a separation between them is hardly visible. As highlighted during wheelbase estimation analysis, Test 8 generated exceptionally high estimation error given the special manner in which the vehicle was conducted, and observing the adjusted curves for Test 8 helps elucidating such high error. This behavior, of little or even no distance between adjusted curves, is observed in nearly all ITD results, which explains the low average wheelbase estimate (around 1.5 m) obtained using this estimator. Lastly, EVD results are unstable, as already suggested by the bumpy estimates in Figures 4.2 and 4.4. Adjusted curves obtained using EVD might be equivalent to the ones obtained using GCC and LMS for some tests (*e.g.*: Test 1, 2, 6, 11-20, 22, 23), or similar to ITD results for other (*e.g.*: Test 3, 4, 5, 8, 9).

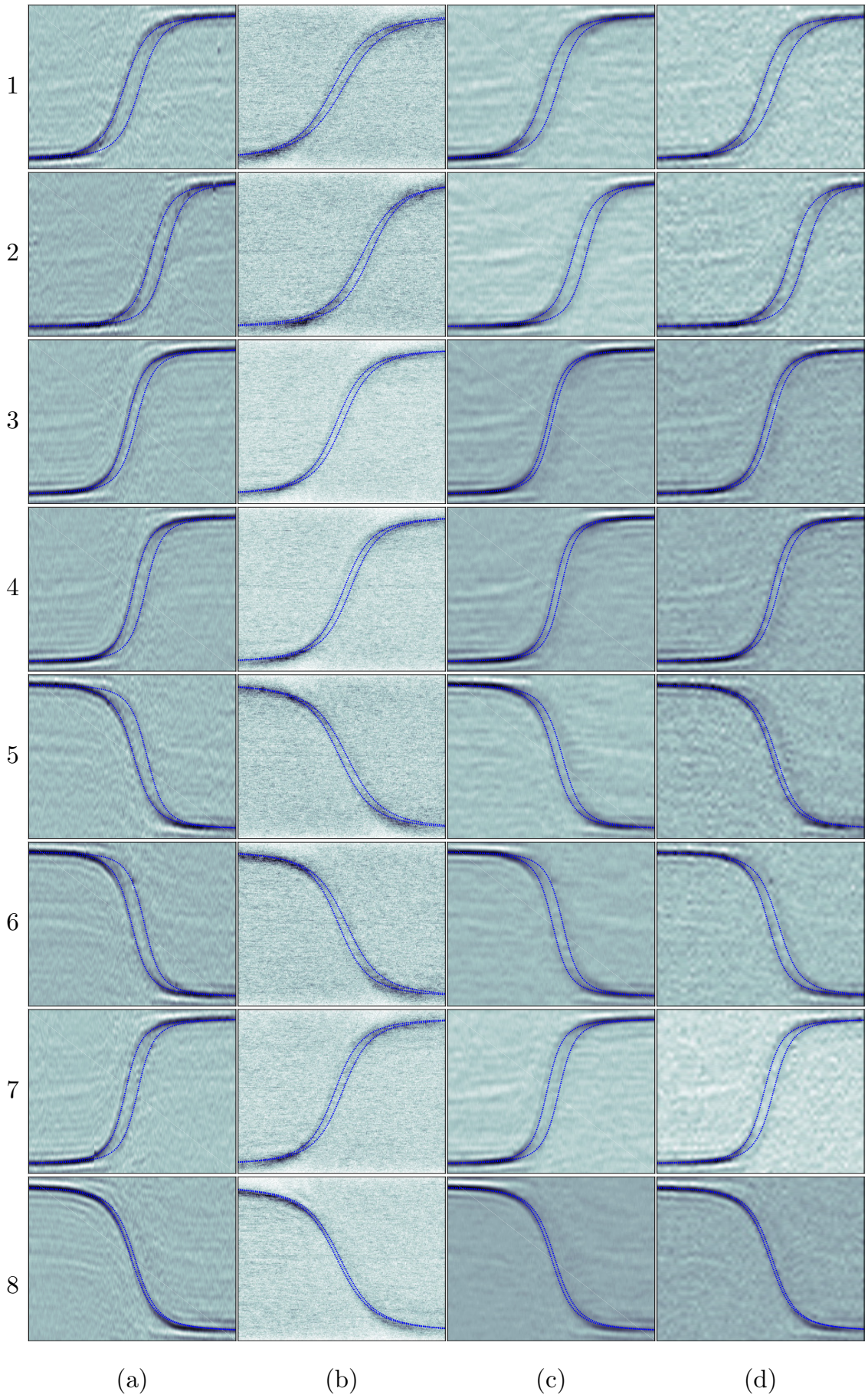

Figure A.1: Tests 1-8 in rows. Estimators in columns: (a) GCC, (b) ITD, (c) LMS and (d) EVD. Adjusted curves as solid blue lines and matrix $\boldsymbol{A}$ as grayscale pixels.

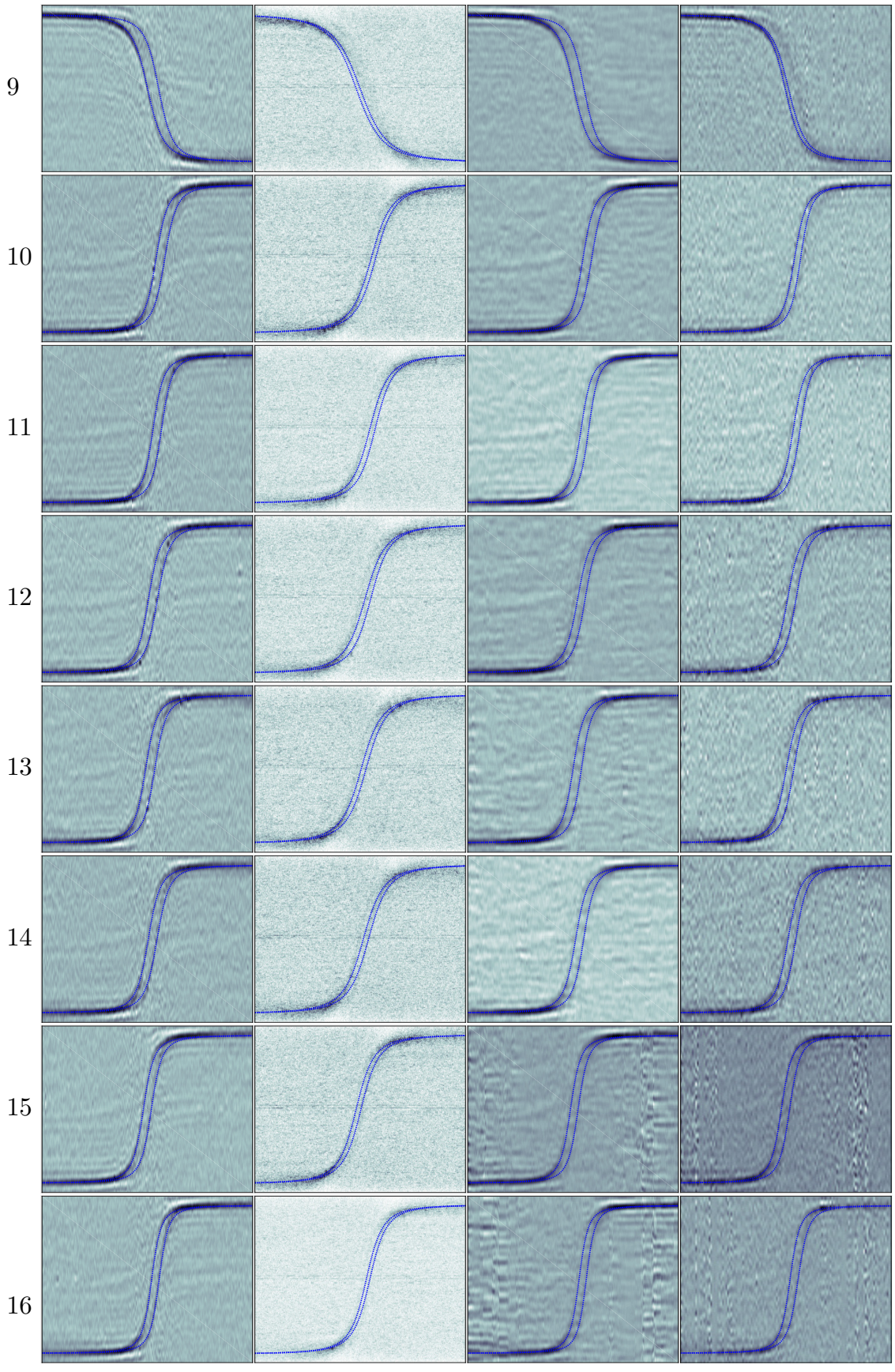

Figure A.2: Tests 9-16 in rows. Estimators in columns: (a) GCC, (b) ITD, (c) LMS and (d) EVD. Adjusted curves as solid blue lines and matrix $\boldsymbol{A}$ as grayscale pixels.

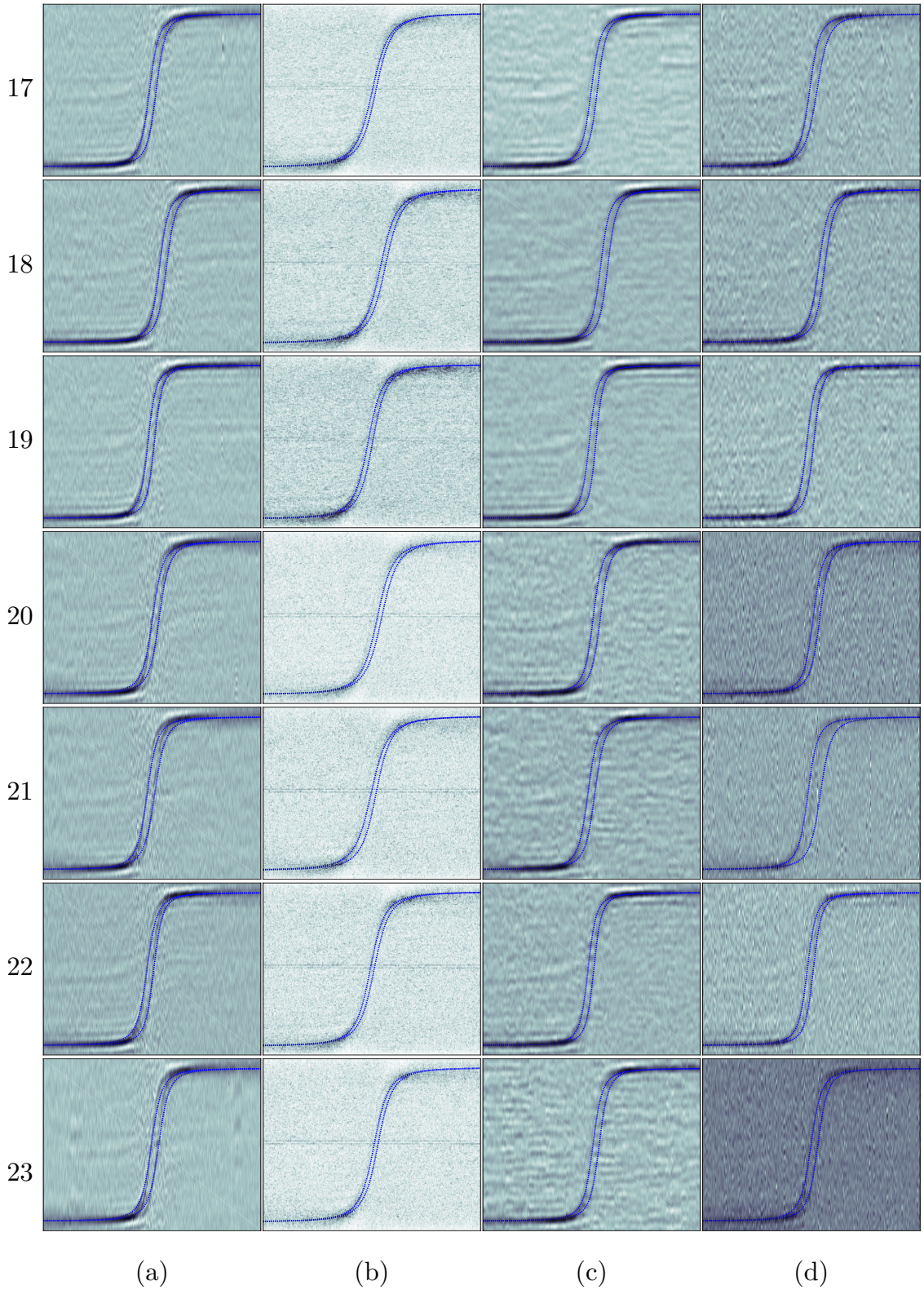

Figure A.3: Tests 17-23 in rows. Estimators in columns: (a) GCC, (b) ITD, (c) LMS and (d) EVD. Adjusted curves as solid blue lines and matrix $\boldsymbol{A}$ as grayscale pixels.

References

[1] ORGANIZATION, W. H., OTHERS. "Burden of disease from environmental noise: Quantification of healthy life years lost in Europe". In: *Burden of disease from environmental noise: quantification of healthy life years lost in Europe*, p. 126, 2011.

[2] AGARWAL, S., SWAMI, B. L., OTHERS. "Road traffic noise, annoyance and community health survey-A case study for an Indian city", *Noise and Health*, v. 13, n. 53, pp. 272, 2011.

[3] ZUO, F., LI, Y., JOHNSON, S., et al. "Temporal and spatial variability of traffic-related noise in the City of Toronto, Canada", *Science of the Total Environment*, v. 472, pp. 1100–1107, 2014.

[4] ZANNIN, P. H. T., DINIZ, F. B., BARBOSA, W. A. "Environmental noise pollution in the city of Curitiba, Brazil", *Applied Acoustics*, v. 63, n. 4, pp. 351–358, 2002.

[5] HAMMER, M. S., SWINBURN, T. K., NEITZEL, R. L. "Environmental noise pollution in the United States: developing an effective public health response", *Environmental health perspectives*, v. 122, n. 2, pp. 115–119, 2013.

[6] NUGENT, C., BLANES, N., FONS, J., et al. "Noise in europe 2014", *European Enviroment Agency*, v. 10, pp. 2014, 2014.

[7] SEONG, J. C., PARK, T. H., KO, J. H., et al. "Modeling of road traffic noise and estimated human exposure in Fulton County, Georgia, USA", *Environment international*, v. 37, n. 8, pp. 1336–1341, 2011.

[8] PINTO, F. A. D. N. C., MARDONES, M. D. M. "Noise mapping of densely populated neighborhoods—example of Copacabana, Rio de Janeiro—Brazil", *Environmental monitoring and assessment*, v. 155, n. 1-4, pp. 309–318, 2009.

[9] DIRECTIVE, E. "Directive 2002/49/EC of the European parliament and the Council of 25 June 2002 relating to the assessment and management of

environmental noise", *Official Journal of the European Communities, L*, v. 189, n. 18.07, pp. 2002, 2002.

[10] VORLÄNDER, M. *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer Science & Business Media, 2007.

[11] RIZZI, S. A., SULLIVAN, B. M., SANDRIDGE, C. A. "A three-dimensional virtual simulator for aircraft flyover presentation", 2003.

[12] ARNTZEN, M., SIMONS, D. "Modeling and synthesis of aircraft flyover noise", *Applied Acoustics*, v. 84, pp. 99–106, 2014.

[13] NILSSON, M., FORSSÉN, J., LUNDÉN, P., et al. "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.

[14] JAGLA, J., MAILLARD, J., MARTIN, N. "Sample-based engine noise synthesis using an enhanced pitch-synchronous overlap-and-add method", *The Journal of the Acoustical Society of America*, v. 132, n. 5, pp. 3098–3108, 2012.

[15] PIEREN, R., BÜTLER, T., HEUTSCHI, K. "Auralization of accelerating passenger cars using spectral modeling synthesis", *Applied Sciences*, v. 6, n. 1, pp. 5, 2016.

[16] DOBLINGER, G. "Localization and tracking of acoustical sources", *Topics in acoustic echo and noise control*, pp. 91–122, 2006.

[17] SANDBERG, U. *Tyre/road noise: myths and realities*. Statens väg-och transportforskningsinstitut, 2001.

[18] MARMAROLI, P., ODOBEZ, J.-M., FALOURD, X., et al. "A bimodal sound source model for vehicle tracking in traffic monitoring". In: *2011 19th European Signal Processing Conference*, pp. 1327–1331. IEEE, 2011.

[19] MARMAROLI, P., CARMONA, M., ODOBEZ, J.-M., et al. "Observation of vehicle axles through pass-by noise: A strategy of microphone array design", *IEEE Transactions on Intelligent Transportation Systems*, v. 14, n. 4, pp. 1654–1664, 2013.

[20] ROCHA, G. D., PETRAGLIA, F. R., TORRES, J. C. B., et al. "Direction of arrival estimation of acoustic vehicular sources". In: *Proceedings of the 23rd International Congress on Acoustics*, 2019.

[21] CEVHER, V., CHELLAPPA, R., MCCLELLAN, J. H. "Vehicle speed estimation using acoustic wave patterns", *IEEE Transactions on signal processing*, v. 57, n. 1, pp. 30–47, 2008.

[22] LÓPEZ-VALCARCE, R., MOSQUERA, C., PÉREZ-GONZÁLEZ, F. "Estimation of road vehicle speed using two omnidirectional microphones: A maximum likelihood approach", *EURASIP Journal on Advances in Signal Processing*, v. 2004, n. 8, pp. 929146, 2004.

[23] BORKAR, P., MALIK, L. G. "REVIEW ON VEHICULAR SPEED, DENSITY ESTIMATION AND CLASSIFICATION USING ACOUSTIC SIGNAL." *International Journal for Traffic & Transport Engineering*, v. 3, n. 3, 2013.

[24] KNAPP, C., CARTER, G. "The generalized correlation method for estimation of time delay", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 24, n. 4, pp. 320–327, 1976.

[25] HAYES, M. H. *Statistical digital signal processing and modeling*. John Wiley & Sons, 2009.

[26] REED, F., FEINTUCH, P., BERSHAD, N. "Time delay estimation using the LMS adaptive filter–static behavior", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 29, n. 3, pp. 561–571, 1981.

[27] FERRARA, E. "Fast implementations of LMS adaptive filters", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 28, n. 4, pp. 474–475, 1980.

[28] MANSOUR, D., GRAY, A. "Unconstrained frequency-domain adaptive filter", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 30, n. 5, pp. 726–734, 1982.

[29] BENESTY, J. "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization", *The Journal of the Acoustical Society of America*, v. 107, n. 1, pp. 384–391, 2000.

[30] ROCHA, G. D. "Estimação Das Direções De Chegada De Fontes Sonoras Veiculares Usando Arranjo De Microfones". 2018.

[31] HARALICK, R. M., SHAPIRO, L. G. *Computer and robot vision*, v. 1. Addison-wesley Reading, 1992.

[32] MORÉ, J. J., SORENSEN, D. C. "Computing a trust region step", *SIAM Journal on Scientific and Statistical Computing*, v. 4, n. 3, pp. 553–572, 1983.

[33] BRANCH, M. A., COLEMAN, T. F., LI, Y. "A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems", *SIAM Journal on Scientific Computing*, v. 21, n. 1, pp. 1–23, 1999.

[34] BYRD, R. H., SCHNABEL, R. B., SHULTZ, G. A. "Approximate solution of the trust region problem by minimization over two-dimensional subspaces", *Mathematical programming*, v. 40, n. 1-3, pp. 247–263, 1988.

[35] COLEMAN, T. F., LI, Y. "An interior trust region approach for nonlinear minimization subject to bounds", *SIAM Journal on optimization*, v. 6, n. 2, pp. 418–445, 1996.

[36] HOLLAND, P. W., WELSCH, R. E. "Robust regression using iteratively reweighted least-squares", *Communications in Statistics-theory and Methods*, v. 6, n. 9, pp. 813–827, 1977.

[37] STREET, J. O., CARROLL, R. J., RUPPERT, D. "A note on computing robust regression estimates via iteratively reweighted least squares", *The American Statistician*, v. 42, n. 2, pp. 152–154, 1988.

[38] BRANDSTEIN, M. S., SILVERMAN, H. F., OTHERS. "A practical methodology for speech source localization with microphone arrays", *Computer Speech & Language*, v. 11, n. 2, pp. 91–126, 1997.

[39] DI CLAUDIO, E. D., PARISI, R. "Multi-source localization strategies". In: *Microphone Arrays*, Springer, pp. 181–201, 2001.

[40] CHEN, J. C., YAO, K., HUDSON, R. E. "Acoustic source localization and beamforming: theory and practice", *EURASIP Journal on Advances in Signal Processing*, v. 2003, n. 4, pp. 926837, 2003.

[41] TIANA-ROIG, E., JACOBSEN, F., GRANDE, E. F. "Beamforming with a circular microphone array for localization of environmental noise sources", *The Journal of the Acoustical Society of America*, v. 128, n. 6, pp. 3535–3542, 2010.

[42] SILVERMAN, H. F., YU, Y., SACHAR, J. M., et al. "Performance of real-time source-location estimators for a large-aperture microphone array", *IEEE Transactions on Speech and Audio Processing*, v. 13, n. 4, pp. 593–606, 2005.

[43] LEIBA, R., OLLIVIER, F., MARCHAL, J., et al. "Large array of microphones for the automatic recognition of acoustic sources in urban environment".

In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, v. 255, pp. 2662–2670. Institute of Noise Control Engineering, 2017.

[44] DE AMORIM AUGUSTO, G. H. "Análise De Desempenho De Um Algoritmo De Estimação De Direções De Chegada De Fontes Sonoras Veiculares". 2019.

[45] O'BOY, D., DOWLING, A. "Tyre/road interaction noise—Numerical noise prediction of a patterned tyre on a rough road surface", *Journal of Sound and Vibration*, v. 323, n. 1-2, pp. 270–291, 2009.