ACOUSTIC IMAGING AND MACHINE LEARNING FOR SOURCES LOCALIZATION AND IDENTIFICATION : APPLICATION TO IN SITU VEHICLE PASS-BY NOISE

Lucille Pinel[1*], **Simon Bouley**[1], **Antoine Purier**[1],
Enzo Tiffeneau[1], **Fabien Lepercque**[1]
[1] MicrodB, 28 Chemin du Petit Bois, 69130, Écully, France

ABSTRACT

Acoustic monitoring aims at detecting, identifying and classifying sound sources. Recently, applications such as street sound events detection or vulnerable areas protection have taken advantage of its techniques. Sound sources of interest can be as varied as dog barks, engine or drone noises and even gunshots sounds. Noises measurement and data processing are two key aspects of such applications. Noise measurement can be carried out with a single microphone but in most environment, the presence of both multiples sources and high background noise (wind, traffic) requires a microphone array to extract the signal of interest and localize the sound sources. This paper presents a methodology to control loud vehicles pass-by noise in streets. First, a specific noise camera, including acoustic and video measurements, allows to localise the sound sources. Limited to one line of microphones for cost and compactness reasons, this array is not able to separate the source of interest (a vehicle to control) from the background noise in its main directivity lobe. To do so, statistical learning is leveraged, using multimodal measurements to eliminate false alarms. This paper details the proposed joined methodology to localize and identify vehicles pass-by noise, and presents applications on experimental data.

Keywords: *source localization, source identification, pass-by noise, acoustic monitoring*

**Corresponding author: lucille.lamotte@microdb.fr.*

1. INTRODUCTION

In recent years, noise events detection has become a priority to tackle urban noise pollution by monitoring vehicle noise in field conditions to obtain realistic noise database and later define penalties [1].
The objective is to quantify the noise sources generated by passing vehicles and to harmonize their levels to the standard ISO 362-1:2022 [2]. Source of noise are numerous in busy cities, with human activities (construction or road works, conversations, etc.), natural life (singing birds, dog barks, etc.) and environmental noise (rain, storm, etc.). In these complex environments with a multitude of noise sources, the use of a single microphone/sound level meter for standardized vehicle pass-by noise measurements implies to control that no other noise sources have contributed to the ambient noise. Coupling the audio measurement with video analysis can help [3, 4] but is not sufficient for accurate sound level estimation. Contrariwise, acoustic imaging methods coupled with video enables the identification of noise sources in space.
Two main approaches can be used to verify that the measured noise is emitted by a single passing vehicle: a first one is to make use of triangulation to locate dominating sources and check if one of them is positioned onto the vehicle. However, as stated by Elias [5], as triangulation lacks of accurate spatial resolution, the whole acoustic scene must be overwhelmed by one dominant source, which may not be the case in real-life traffic situations. An alternative way is to take advantage of microphone array processing to precisely estimate the sound level of potential sources in the area surrounding the vehicle position [6]. This paper addresses the latter as this specific process allows to monitor vehicle sound emission

even with low signal-to-noise ratio and gives accurate sound quantification. Moreover, array processing can spatially filter the source of interest from the others (close vehicle, high background noise).

This type of measurement and processing with a microphone array and video monitoring provides interesting information for sound events detection, such as their location, trajectory, frequency spectrum or temporal behavior. In a nutshell, acoustic imaging localizes acoustic hot spots in a predetermined area around the microphone array. However, the nature of the radiating object is still to be identified. Also, array processing techniques are limited by the spatial resolution of computed acoustic maps. Hence, especially at low frequency, multiple sources can be hidden inside the most prominent hot spots. To answer these questions, acoustic imaging and video processing outputs can feed statistical learning algorithms to identify the sound source and prevent false alarms.

This paper presents the development and experiments of the foreseen methodology detailed in [7], dedicated to the acoustic monitoring of in situ vehicle pass-by noise. Section 2 briefly describes the measurement system and processing. Section 3 explains the objectives of statistical learning using multimodal measurement. Finally, Sec. 4 illustrates the proposed methodology with results of passing vehicles from experiments in real-life traffic situations.

2. ACOUSTIC IMAGING FOR SOURCE LOCALIZATION

2.1 Acoustic imaging measurement tool and set-up for vehicle pass-by noise measurement

As stated in [7], the acoustic monitoring of vehicles along traffic lanes leaded to design a linear microphone array, ensuring optimal performances with respect to the geometrical configuration, and completed with video monitoring. Such a system named “dBFlash” has been developed to withstand long measurement campaigns over several months in any weather conditions as illustrated on Fig. 1. The antenna is 1,6 m long with 52 MEMS microphones randomly distributed and protected from dust, rain and wind. As illustrated on Fig. 2, it is installed on the roadside, parallel to the lanes, at around 5 m height. This antenna shape implies an optimal directivity in a view angle of $\pm 30^\circ$ on each side of the system. In front of the antenna, a “control area” is set where the ability of array processing to distinguish close sources is optimal. A sound level meter is also synchronized with the microphone array to measure the ambient noise according to standard and certified measurement.

Figure 1. dBFlash system mounted on a mast with a microphone array and video camera on top for in-situ pass-by measurements.

The measurement is carried out over few seconds to cover the vehicle passage through the measurement area (at least +/- 10 m for speeds above 30 km/h). Data acquisition is triggered from the top camera by the video detection of passing vehicles below the microphone array. It permits to easily synchronize video and audio recordings and ensures the vehicle position regarding the audio processing. The system can be completed by an automatic plate number recognition system (ANPR) to identify the vehicle’s characteristics and its owner.

The video camera view angle allows to monitor two lanes and trigger data acquisition for any driving directions below the system (upstream or downstream). Triggering from video event rather than noise event permits to process all vehicle passages independently of their acoustic level in a noisy environment.
Audio and video data are finally processed by a computer and results are stored on a hard disk drive or transferred to a secured server if necessary to comply with General Data Protection Regulation.

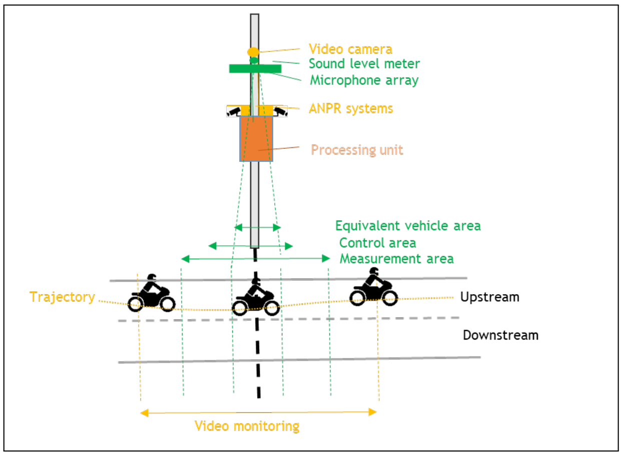

Figure 2. Pass-by noise measurement system diagram.

2.2 Array processing for noise source quantification

As reference, the sound level meter quantifies the ambient sound pressure level L_{AF} with a frequency-weighting A and a time-weighting F (fast) to comply with IEC 61672 standard [8]. The maximum time-weighted sound level measured during the vehicle passage is denoted as L_{AFmax}. To be able to quantify the contribution of the noise emitted by the sources inside the equivalent vehicle area represented on Fig. 2, acoustic imaging algorithms such as conventional beamforming [5] associated with a deconvolution step (namely CLEAN-SC [9]) are applied on microphones signals. The process is carried out in frequency domain on short time sequences for which the vehicle displacement is considered as negligible. The frequency range is limited to [200-6 000] Hz, where the array performance and directivity are validated without spatial aliasing or too poor spatial resolution which would not permit to separate two vehicles distant from 5 m in the control area.

A linear calculation grid represents the potential noise sources in the measurement area (between $\pm$ 10 m and $\pm$ 20 m) and is positioned close to the vehicle trajectory. Therefore, the linear array has no resolution in the transverse direction. Its length largely covers the equivalent sources in the control area of interest ($\pm$ 5 m) as required by the deconvolution processing. The vehicle trajectory is provided by video analysis with a good accuracy. Hence, this processing provides a fruitful acoustic trace representing sound events in a space/time reference frame.

From this acoustic trace, the potential sources in a smaller zone denoted as *equivalent vehicle area* (whose size equals one vehicle length ($\pm$ 2.5 m), Fig. 2) are spatially integrated to estimate their contribution during the measurement to a reference point located 7.5 m away, like in pass-by noise measurement standard. Therefore, the integrated dB(A) sound level of isolated passing vehicles can be estimated, as well as its maximum value. A fast time-weighting is applied to smooth the tracking with the same parameters as the sound level meter to obtain the L_{AFmax}. This methodology provides the superimposed evolution of the ambient noise level and the acoustic contribution of passing vehicle, respectively measured and processed with the sound level meter and the microphone array. A measurement campaign at Transpolis test facilities (France) with low ambient noise validated the good agreement between the maxima of vehicle pass-by sound pressure level in dB(A) with Fast weighting (L_{AFmax}) computed with the sound level meter and the array processing, respectively. Both measurement systems were located close to the lane, while the vehicle speed ranged from 30 to 90 km/h, in both traffic directions. Discrepancies between both measurements do not exceed 1 dB as illustrated on Fig. 3.

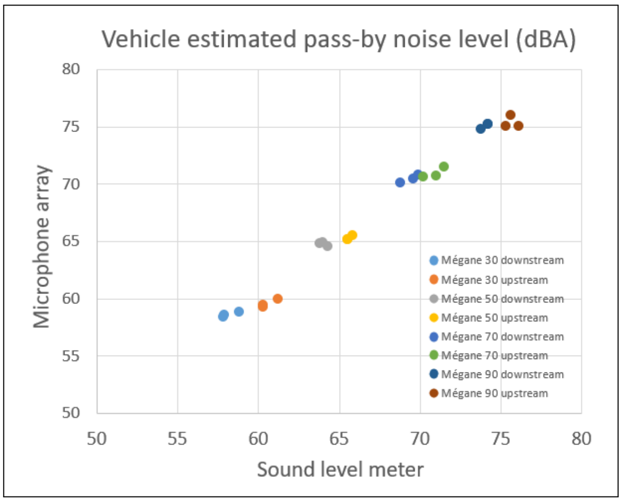

Figure 3. Sound pressure level L_{AFmax} comparison between sound level meter and microphone array processing for vehicle passage at different speeds.

In noisy environments, the measured sound level in the equivalent vehicle area may not only be due to the passing vehicle acoustic radiation as other sources in this area could be present at the same instant (other passing or parked vehicles, dog barks, shouts, car horn ...). To accurately identify the multiple elements in the acoustic scene, statistical learning algorithms can be leveraged.

3. STATISTICAL LEARNING FOR SOURCE IDENTIFICATION

3.1 Single vehicle from video analysis

The vehicle detection is carried out by statistical learning algorithms such as Hidden Markov Models and its derivatives [10]. It identifies all vehicles, their trajectory and size. When the vehicle enters the device field of view, video frames are extracted (Fig. 4, displaying both the trajectory and the vehicle size. First, video monitoring controls the presence of a single vehicle of interest in the control area over $\pm$ 5 m as well as potential masking vehicles on adjacent lanes.
All trajectories are automatically checked to ensure that no other vehicles were simultaneously present in the control area. The vehicle size is also compared to the surface covered by the second lane in the camera view angle in case one vehicle could not be detected due to masking effect (e.g. a motorbike behind a van).

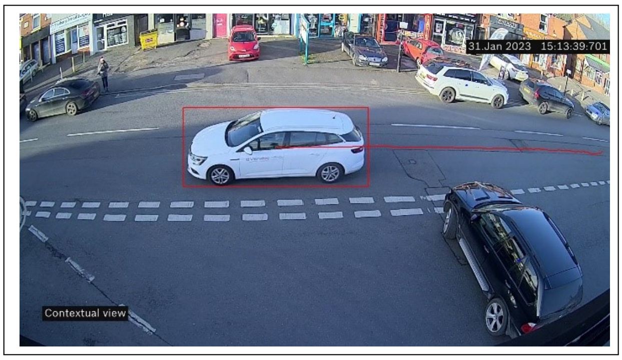

Figure 4. Contextual view from the video camera with a red bounding box around the identified vehicle and its trajectory.

3.2 Sound recognition

To identify the measured noise as a passing vehicle, its signature is processed with statistical learning algorithms initially trained from a large audio dataset of usual false alarms (human conversation, dog barks, etc.) or undesired types of noises (siren, horn).

Once vehicles are detected through video analysis, acoustic measurements are triggered. Pressure signals are then processed to compute a set of features that describe shape, statistics and entropy of the signals in temporal, spectral and cepstral domains [11]. Then, a linear support vector machine (SVM) model [12] is learned from these features. This method has been previously assessed and showed good performances in other acoustic applications [13].

4. APPLICATION TO IN-SITU VEHICLE PASS-BY NOISE

4.1 Acoustic trace

Figure 5 displays the vehicle pass-by noise trace from array processing for the scenario illustrated in Fig. 4. It describes the acoustic scene in time-space domain. In this case, it points out a dominant source emission linearly evolving in distance with time, which corresponds to the vehicle displacement at constant speed. The two close lines distant from 2.5 m correspond to the sources emitted from front and rear vehicle areas. The engine compartment emits higher levels (in second gear position) than rear wheels/road contacts. The trace is centered as the audio data acquisition is triggered when the vehicle is passing by below the antenna with $\pm$ 1,5 seconds data recording. Another acoustic trace is visible between -5 and -10 m but outside of the control area. It corresponds to another vehicle slowly moving close to the antenna and visible on the contextual view on Fig. 4.

4.2 Noise level estimation and equivalent vehicle area noise level

From the array processing and sound level meter measurements, sound pressure levels (dB(A)) are calculated with and without Fast time-weighting during the vehicle passage. The acoustic trace maximum noise level displayed in Fig. 5 is also extracted (Fig. 6). While there is a high ambient noise around 70 dB(A) (green curve), the equivalent noise source emission in the vehicle area has a maximum at 68,2 dB(A) (red curve). It corresponds to the vehicle passage below the antenna. In this case, the vehicle is 2,7 dB lower than the ambient noise during its passage. Therefore, Fig. 6 proves the system's ability to accurately

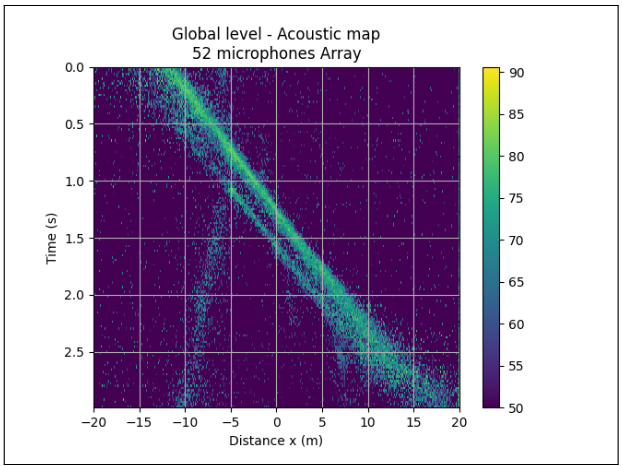

Figure 5. Acoustic trace from deconvolution processing for a passing vehicle.

quantify pass-by noise level, even if the sound pressure level of the vehicle is below the ambient level.

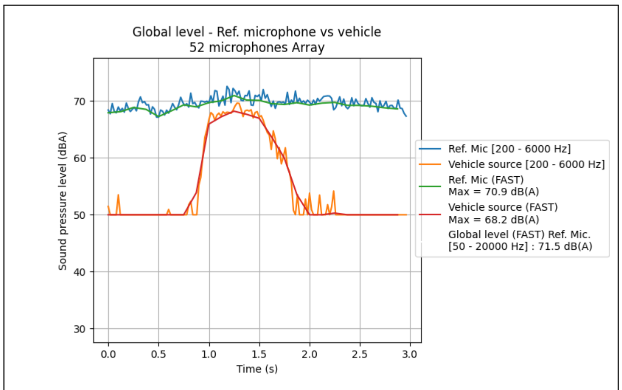

Figure 6. Ambient (blue/green curves) and equivalent vehicle area (orange/red curves) sound pressure tracking for vehicle pass-by.

4.3 Specific case with sirens

Emergency vehicles usually emit high noise levels, especially when they make use of their siren. For that reason, they make an interesting class to isolate from other vehicles in usual traffic.

Hence, a linear SVM model has been trained with an audio dataset composed of sounds of sirens from several countries, as they often largely differ. 2620 samples of 0.2 seconds have been considered, 2077 for the training set (4-fold cross-validation) and 543 for the test set. The results are display in Tab. 1. the output confusion matrix is displayed in Tab. 1, and shows accurate predictions :

Table 1. Confusion matrix of the SVM model.

		Predicted	
		Siren	Noise
Actual	Siren	376	1
	Noise	0	166

Machine learning algorithms such as SVM can also provide fruitful information about the most significant features that drive identification. Thus, the typical two-tone siren seems to be mostly identified by a set of ten features involving both temporal, spectral and cepstral domains (Fig. 7), with a significant influence of skewness and kurtosis properties as well as the Rényi entropy of the signals.

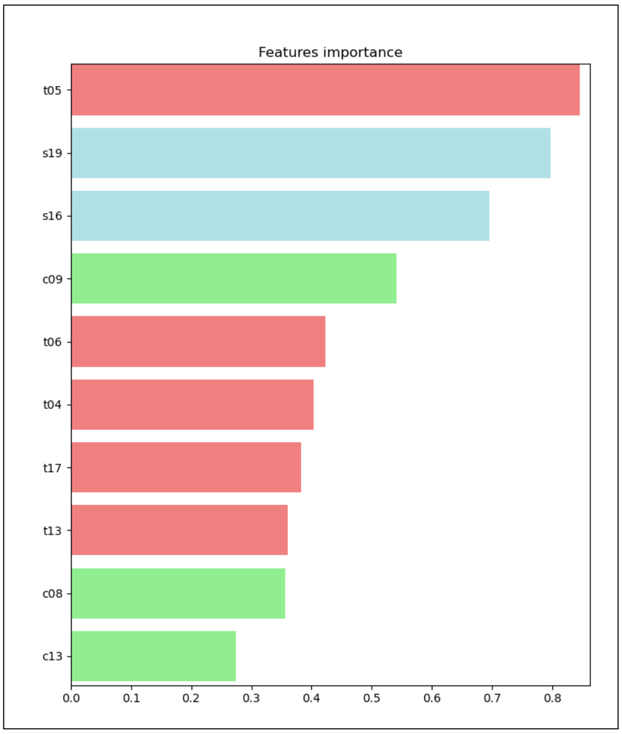

Figure 7. Most prominent features for two-tone siren identification. (temporal in red, spectral in blue and cepstral in green.)

5. CONCLUSION

In situ vehicle pass-by noise measurement with high background noise and disturbing sources requires to use acoustic imaging in combination with statistical learning to obtain accurate vehicle noise levels with trust. Synchronized with video monitoring, the proposed system can be used in automatic mode to build a large database of vehicle noises in urban areas or control vehicles for noise threshold infringment. These results point out that the acoustic trace provides valuable information which could help to describe the acoustic scene with more precision. For example, impulsive, stationary or moving sources can be directly distinguished from these maps. Also, this work opens many other possible applications aimed at detecting, identifying and classifying noise sources in various sound landscapes. As mentioned in the paper, the linear geometry of the proposed array implies its resolution in the transverse direction is limited. Therefore, further studies are required to apply this methodology to multi-lane roads, possibly using a multi-antenna configuration, able to grasp the transverse direction.

6. ACKNOWLEDGMENTS

MicrodB wishes to acknowledge the support of French ADEME organization on dBFlash project.

7. REFERENCES

[1] L. Morris, A. Lawrence, and R. Stait, "Roadside vehicle noise measurement – Phase 1 study report and technology recommendations," tech. rep., Atkins and Jacobs, mar. 2019.

[2] "ISO 362-1:2022 – Acoustics – Engineering method for measurement of noise emitted by accelerating road vehicles – Part 1: M and N categories," standard, International Organization for Standardization, Geneva, CH, Mar. 2022.

[3] A. Klos, "Noise camera: automated detection technology to identify noisy vehicles," in *Proceedings of ACOUSTICS 2006*, pp. 343–348, Australian Acoustical Society, 2006.

[4] A. Agha, W.-S. Gan, Y.-K. Chong, and B.-W. Ang, "A noisy vehicle surveillance camera (noivelcam) system," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 249, pp. 1046–1054, Institute of Noise Control Engineering, 2014.

[5] G. Elias, "Experimental techniques for source location," *Lectures Series 1997–07 Aeroacoustics and Active Noise Control*, 1997.

[6] D. Püschel, M. Auerbach, W. Bartolomaeus, and C. Marín, "Pass-by measurements using array techniques," in *2016 12th IEEE International Conference on Industry Applications (INDUSCON)*, pp. 1–7, IEEE, 2016.

[7] L. Pinel Lamotte, F. Lepercque, and V. Baron, "Accurate and controled vehicle pass-by noise emission quantification in real life traffic," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 263, pp. 5233–5245, Institute of Noise Control Engineering, 2021.

[8] "IEC 61672-1 – Electroacoustics — Sound level meters — Part 1: Specifications," standard, International Electrotechnical Commission, Geneva, CH, sep. 2013.

[9] P. Sijtsma, "Clean based on spatial source coherence," *International journal of aeroacoustics*, vol. 6, no. 4, pp. 357–374, 2007.

[10] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[11] M. Malfante, M. Dalla Mura, J.-P. Métaxian, J. I. Mars, O. Macedo, and A. Inza, "Machine learning for volcano-seismic signals: Challenges and perspectives," *IEEE Signal Processing Magazine*, vol. 35, no. 2, pp. 20–30, 2018.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[13] V. Baron, S. Bouley, M. Muschinowski, J. Mars, and B. Nicolas, "Drone localization and identification using an acoustic array and supervised learning," in *Artificial Intelligence and Machine Learning in Defense Applications*, vol. 11169, pp. 129–137, SPIE, 2019.