Large array of microphones for the automatic recognition of acoustic sources in urban environment

Raphaël Leiba, François Ollivier, Jacques Marchal, Nicolas Misdariis, Régis Marchiano

inter.noise
27-30 AUGUST HONG KONG 2017

Large array of microphones for the automatic recognition of acoustic sources in urban environment

Raphaël LEIBA*[1,2]; François OLLIVIER[1]; Jacques MARCHAL[1]; Nicolas MISDARIIS[2]; Régis MARCHIANO [1]

[1]Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7190, Institut Jean Le Rond ∂'Alembert, France
[2]STMS Ircam-CNRS-UPMC, France

ABSTRACT

Characterising the urban sonic environment is usually achieved by measuring energetic indicators such as the average A-weighted level (LAeq) or the day-evening-night Level(Lden). The European 2002/49/EC directive compels large cities to set up noise maps and implement action plans aimed at reducing the citizens' exposition to excessive noise. Yet the diagnosis step of the process requires the sonic environment to be quantified and the soundscape concept suggests that the nature of sound sources is central in this description. Furthermore, the nature of sources on noise annoyance have been considered homogeneous for city dwellers in some previous studies.
Therefore, our work focuses on the nature of sound sources in urban areas and their automatic assignment to given perceptual categories. This is done using very large microphone array processing and video tracking to extract individual passing-by vehicles audio signal from a scene involving multiple vehicles. These moving source signals feed a categorisation process consisting in a supervised machine learning algorithm (Support Vector Machine). The paper presents a benchmark experiment that allowed to gather reference signals out of traffic for different passing-by vehicle types (heavy vehicle, utility truck, personal car, motorcycle, ...) and driving condition (constant speed, acceleration, deceleration). These signals are used as training samples for the machine learning algorithm so that the categorisation process allows to characterize the vehicle on perceptual categories (combining both vehicle types and driving conditions). Another experiment led *in situ* is also analysed. The automatic recognition robustness is discussed and improved by adding the *in situ* extracted signals to the training samples.

Keywords: Urban noises, Beamforming, Machine learning
I-INCE Classification of Subjects Numbers: 72, 50

1. INTRODUCTION

From 2002, the European Union has set up a standardized legislation in order to quantify, inform on noise exposition of city dwellers and, if need be, act to reduce it. This 2002/49/EC directive[1] (known as Environmental Noise Directive – END) imposes the construction of maps characterizing the noise induced by different sources such as road, air or rail traffic for cities with more than 100,000 inhabitants. The maps have to be established in L_{DEN} (day-evening-night level) a noise level implementing the increase of annoyance in the evening (5 dB added to the measurements) and at night (10 dB added).

In addition to national initiatives, this directive has been a big step forward in reducing noise exposure. But Raymond Murray Schafer [2] tells us that a good regulation has to focus on the classification of sources in order to act on the most annoying ones. Yet this directive does not provide a segregated information. For instance, it does not analyze the psychoacoustic composition of the road traffic and discriminate the specific annoyance of each type of vehicle in its various driving conditions.

In a lot of recent surveys, in France, road traffic noise is pointed out as the first annoying noise source. Following this analysis some studies have been driven on understanding the annoyance induced by road traffic focusing on the audio signal of the sources [3, 4].

Such as R.M. Schafer, Morel *et al.*[5] pointed out the perceived difference between the different types and driving conditions of road vehicles. They studied how the subjects were grouping different vehicle passing-by sounds and proposed a perceptual typology. This typology is split into seven categories,

*raphael.leiba@upmc.fr

sometimes mixing different types of vehicles in the same driving condition, sometimes with only one type of vehicle and one driving condition. These categories are listed below:

Cat. 1 : Two-wheel vehicles passing-by at constant speed,
Cat. 2 : Two-wheel vehicles in acceleration,
Cat. 3 : Buses, light and heavy vehicles passing-by at constant speed,
Cat. 4 : Two-wheel vehicles in deceleration,
Cat. 5 : Buses, light vehicles and heavy vehicles in deceleration,
Cat. 6 : Light vehicles in acceleration,
Cat. 7 : Buses and heavy vehicles in acceleration.

Classifying road traffic according to this typology would considerably enrich noise maps and could help to build annoyance models. Therefore we propose to investigate the feasibility of an automatic classification coupled with this perceptual taxonomy of road traffic noise. We choose to make the classification rely on the specific source audio signals. This requires to identify each potential source (the vehicles) and extract its specific signal out of the global acoustic scene. Various sound extraction methods have been proposed in the last decades, particularly in the field of speech enhancement. They often consist in denoising mono-channel signals, sometimes using a Non-negative Matrix Factorization (see for instance [6]) or Deep Neural Networks (such as [7, 8]). We choose to adopt an original approach consisting in using large microphone arrays and a standard inverse method adapted to moving sound source extraction : conventional beamforming (see [9, 10]).

In this paper, we investigate a road traffic classification method based on the combination of a video tracking process, the use of a large microphone array for spatial filtering of source signals together with machine learning algorithms for sources classification. The methodology is presented in the first section. In the following, the classification process is depicted, describing in particular the tuning and training steps working on a number of dedicated scenarios involving single passing-by vehicles. Finally, the method is applied to the case of a real urban environment for which we present a quantified analysis of the road traffic flow for a period of one day.

2. METHODOLOGY

In this section the classification method for road traffic flow is exposed. We are interested only in the noise generated by vehicles taken individually, meaning that we are only looking for mobile sources. Of course this assumption is restrictive and could be overcome in the near future. The method proceeds in three distinct steps. First, the mobile source trajectories are obtained from video processing. Then, the audio signals from each moving source is extracted along its identified trajectory by means of an adaptive tracking beamforming algorithm. The features retrieved from the source signals are finally fed to a supervised machine learning classifier in order to assign the sources to categories. .

2.1 SOURCE TRAJECTORY BY VIDEO TRACKING

A real road traffic scene is usually composed of different vehicles with distinct trajectories. In this step, the objective is to isolate each vehicle automatically by processing a video stream provided by a video camera located at the center of the microphone array (see section below). We use the algorithm described below for each frame of the video. It is based on background subtraction and makes use of the `OpenCV`[1] library:

1. the background (the frame without moving targets) and the frame to be processed are converted to gray levels and blurred,
2. the difference between the two frames is computed and the resulting "delta" frame is blurred and thresholded,
3. a contour detection algorithm that provides rectangles isolating the moving vehicles.

The trajectory of the target is derived from the successive positions of the centroids and provides at each time sample t the position $\mathbf{x}_i = (x_i, y_i, z_i)$ of the i^{th} moving target.

2.2 SOURCE AUDIO SIGNAL EXTRACTION

When the trajectory is known, it can be used to extract an audio signal of the source. This is performed using a well-known inverse method: conventional beamforming. It aims at estimating the audio signal $\mathrm{s}_i(t)$ at the i^{th} moving vehicle position by phase shifting the recorded signals of a microphone array. The beamformed signal is computed in the frequency domain by estimating the spectral components $\hat{\mathrm{s}}_i(\omega)$ according to:

[1] opencv.org/

$$\hat{\mathrm{s}}_i(\omega) = \frac{1}{N_m} \sum_m r_{mi} e^{j\omega r_{mi}/c_0} . \hat{p}_m(\omega), \tag{1}$$

with $r_{mi} = \|\mathbf{x}_m - \mathbf{x}_i\|$, the varying distance between the m^{th} microphone and the tracked source, N_m the number of microphone and $\hat{p}_m(\omega)$ the spectral component of the m^{th} microphone. $\hat{p}_m(\omega)$ results from the short time Fourier transform (STFT) of the pressure signal $p_m(t)$.

The process can be written in matrix form such that:

$$\hat{\mathbf{s}} = \frac{1}{N_m} \mathbf{A}^H \hat{\mathbf{p}}_m, \tag{2}$$

with ^{H} denoting the hermitian transpose, $\hat{\mathbf{p}}_m = [\hat{p}_1(\omega) \cdots \hat{p}_m(\omega) \cdots \hat{p}_{N_m}(\omega)]^T$ and $\mathbf{A}$ a $N_m \times N_i$ phase shifting matrix

The spectral process is parallelized on a GPU[11] in order to reach real-time capabilities. Finally, the target time history signal is obtained by applying an inverse Fourier transform. The overall process involves 50 ms overlapping frames with a 12.5 ms step (75% of overlapping). Prior to the STFT the frames are tapered using a squared sine which ensures the conservation of the energy of the overlapped signals. Finally, the target source signal is obtained by summing the overlapping successive short time histories.

2.3 SOURCE CLASSIFICATION

This section investigates the ability of a machine learning algorithm to estimate the perceptual category associated to a vehicle given its audio signal. Support Vector Machines (SVM) belong to the supervised machine learning algorithms. SVM has been widely used to perform audio signals classification (see for example [12, 13]). It aims at separating at least two categories based on several feature samples for each category. Figure 1 illustrates the limit (called hyperplane) between two categories represented by two features. The purpose of the SVM algorithm is to minimize $\|\vec{w}\|$, the normal vector of the hyperplane, in order to find the hyperplane that maximizes the distance between categories, limited by the two hyperplanes $\vec{w} \cdot \vec{x} - b = 1$ and $\vec{w} \cdot \vec{x} - b = -1$. The three $\vec{x_i}$ vectors are called the support vectors.

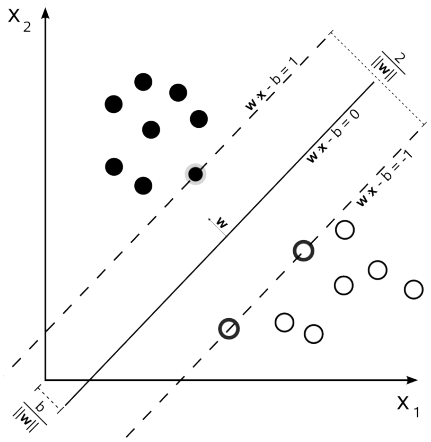


Figure 1 – Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors. From Wikipedia[2]

Valero *et al.* [14] compare different signal extracted features and show the relevance of the Mel Frequency Cepstral Coefficients (MFCC), a feature widely used in automatic speech and speaker recognition. Indeed when using SVM, they provide a good classification rate, confirming previous studies (see [13, 12]). It has to be noted that several other standard physical and psychoacoustical features have been tested providing poorer results : L_{eq}, Spectral centroid, Roughness and Loudness.

The MFCCs are computed by filtering the power spectral density (PSD) of a signal over a mel-filterbank (triangular overlapping windows mimicking the cochlear filterbank[15]). The filtered spectra are summed and logarithmically scaled. Finally, the resulting coefficients undergo a cosine transform, providing the MFCCs. The `python_speech_features`[3] library has been used to compute the MFCCs, using the default number of 26 filters. This provides finally 13 MFCCs.

In order for the SVM algorithm to perform a relevant classification according to the seven chosen categories, each MFCC is normalized with respect to the highest. The classification process makes use of the C-Support Vector Classification class of the SVM module of the `Scikit-Learn`[4] Python's library.

[2] wikipedia.org/wiki/Support_vector_machine [3] github.com/jameslyons/python_speech_features [4] scikit-learn.org

3. VALIDATION ON SINGLE VEHICLES PASS-BY

In order to validate the proposed methodology, a measurement campaign was conducted to collect a data basis of various single vehicles passing-by with various driving conditions. The experiment implemented a dedicated 256 microphone array over the test track of La Ferté-Vidame (France). The microphone array is a 20×2.25 m rectangular frame of MEMS microphones. According to the by-passing measurements standard, it is set-up 7.5m away from the vehicles trajectory (see figure 2). Nine road vehicles, including heavy commercial vehicles, light vehicle and two-wheel vehicles, have been used during this campaign (more details in [16]). Note that all Morel's categories are present in this data basis.

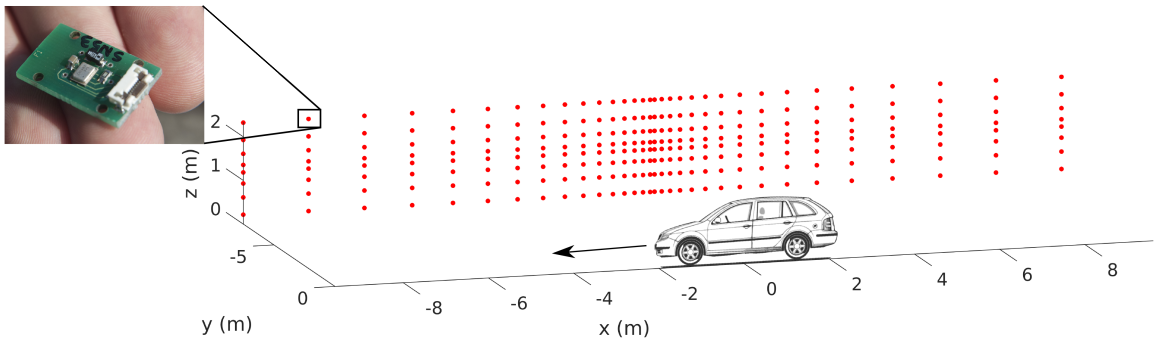


Figure 2 – Experimental set-up for pass-by measurements - Array of 256 microphones (red dots) and picture of an elementary MEMS microphone and circuit. From [16]

The target signal extraction task is first tested by finding the audio signal of a controlled loudspeaker installed on the side of a by-passing car and emitting a 2 kHz pure tone. Figure 3 presents the power spectral density for a MEMS microphone signal (blue) and the beamformed signal (orange) when the vehicle passes by the array center. The tonal components (fundamental and harmonic) are perfectly retrieved and the overall background noise (MEMS numerical noises and vehicle noise) is lowered by about 10dB above 1400 Hz, suggesting that the beamforming process efficiency is limited around the sources spectral signature.

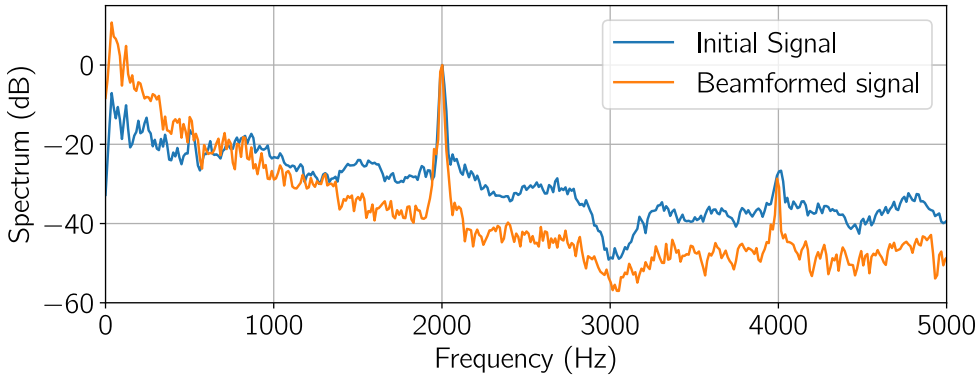


Figure 3 – Power spectral density comparison between a Mems microphone signal (blue) and the beamformed signal (orange)

Some preliminary tests show the importance of weighting the MFCC vector to improve the classification process. In order to estimate the classification error, the model is trained on 88% of the 68 signals and tested on the remaining 8 (12%). The driving condition, obtained by video tracking, is used in the training data set. It is added to the data set using three binary inputs (constant speed = 0 or 1, acceleration = 0 or 1 and deceleration = 0 or 1).

A parametric study has been led to identify the MFCC weighting vector. The classification process is tested for each combination of the MFFC weighting vector taking integer values between 0 and 5. Among the 1,679,616 possible combinations, 26,301 result in a 100% success. One of these perfect configuration weighting vector is chosen arbitrarily :

$$\text{DataWeighting} = [\underbrace{2,1,1,4,2,1,1,0,1,1,1,1}_{\text{MFCC}}, \quad \underbrace{1,1,1}_{\text{Driving conditions}} \quad].$$

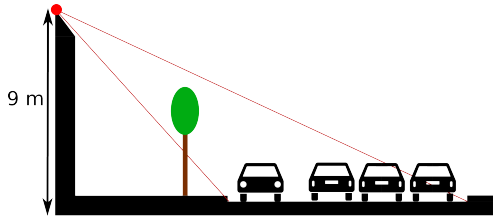
According to these results, the validation test is conclusive. The different steps follow each other properly and the large number of weighting vector providing good results builds confidence for the use of this method in the urban environment.

4. *IN SITU* RESULTS

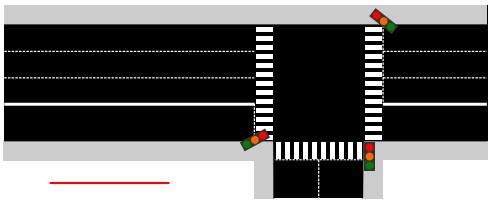
For the *in situ* implementation of the method, a very versatile 128 microphone array has been designed to facilitate the use in the city. The road traffic flow classification is tested on a 3×1 way street in Paris (France). This section presents the experimental set-up and discusses its performances. The traffic flow statistics obtained in the process are presented over a day period.

4.1 EXPERIMENTAL SETUP

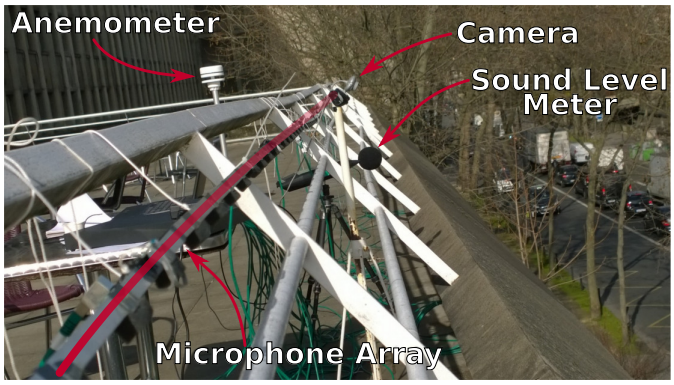
The microphone array used for this measurement campaign is presented in figure 4. A linear 128 MEMS microphone array is set-up on a balcony overlooking the road at a 9 m height (see the sectional view of the street on figure 4a). Note that some trees are located in the field between the road and the antenna. Figure 4b presents, in top view, the overall setup: microphone position (red line) circulation lines and traffic lights. The campaign was conducted in winter, so that there were no leaves on the trees. Figure 4c shows a picture of the linear array, the camera located at its center. A sound level meter provides an overall sound level reference and an anemometer is used to estimate the influence of wind perturbations.



(a) Scheme of the antenna disposition - Section



(b) Scheme of the antenna disposition - Top

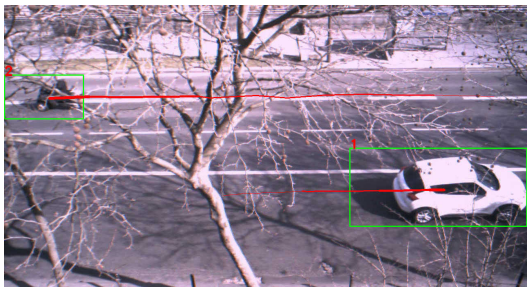


(c) Picture of the antenna and other instrumentations

Figure 4 – 128 microphones antenna for urban acoustic imaging - From [16]

For this experiment, the data processing requires some adaptations, in particular for the video tracking to be successful dealing with a large number of moving objects and the presence of trees. Figure 5 shows a tracking result for a white car accelerating and a scooter at constant speed, passing each other. The figure 5a shows the threshold image on which the contour detection is performed. The blurring step effect is visible on this image resulting in the detection of three objects: two road vehicles and one pedestrian. As can be seen see on figure 5b, the trajectories (in red line) are well estimated in this example. Sometimes the number of vehicle increases, the blurring step returns a degraded tracking result. Consequently multiple vehicles can be grouped into a single moving object.

(a) Threshold image

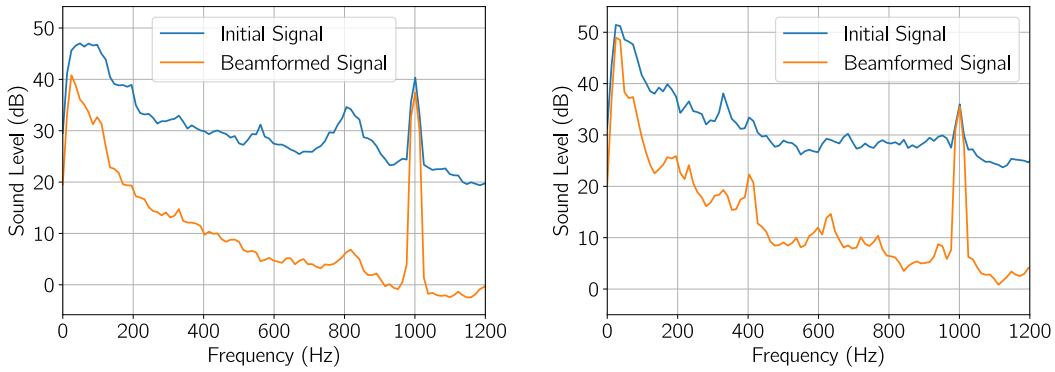


(b) Trajectory detection

Figure 5 – Video tracking examples with trajectory extraction based on background subtraction

In a prior calibration step, the audio signal extraction is tested on a controlled non moving source set at different locations in the scene. The source is a loudspeaker emitting a 1 kHz pure tone. Figure 6 shows the power spectral density of the central MEMS microphone signal and the beamformed signal, the road vehicles are ignored. Figure 6a presents the results for the source located at 12.8 meters away from the array center. As we can see, the peak due to the source is visible in the spectrum (blue curve) but with a strong background noise. The beamformed signal (orange curve) shows a 25 dB dynamic enhancement around the source peak and reducing at low frequencies to about 6 dB. Doubling the distance (Figure 6b),

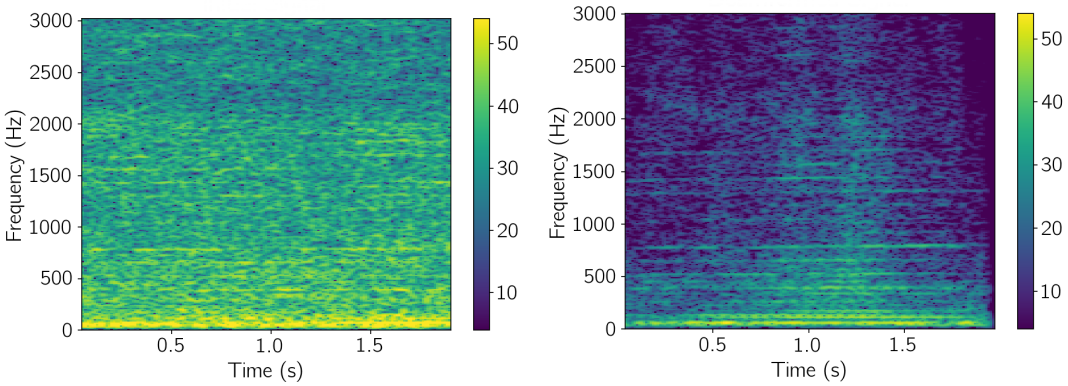
the source signal is barely audible by the central microphone, while the beamforming efficiency drops by a few dBs still showing a raise of dynamics of 20 dB.



(a) 12.8 meters between the source and the array center (b) 26.7 meters between the source and the array center

Figure 6 – Power Spectral Density of array central microphone signal and beamformed signal for two positions of the acoustic source - Signal emitted: 1kHz pure tone.

The efficiency of the chosen array being proved for static pure tone sources, the specific source signal extraction of moving targets in the urban environment is investigated. Figure 7 presents the spectrograms of initial and beamformed signals of a passing-by two-wheel vehicle. In figure 7a no frequency components emerge from the large bandwidth noise. Whereas, when the signal of the two-wheel vehicle is extracted from the mixture (figure 7b), the tonal components of the engine and the broadband noise due to the tire-road contact are clearly visible. This exhibits the filtering capability of the system.



(a) Central microphone array signal

(b) Beamformed signal of the passing-by two-wheel vehicle

Figure 7 – Passing-by two-wheel vehicle signals spectrograms – Dynamic: 40 dB

4.2 CLASSIFICATION RESULTS

In the following the classification process is applied to the beamformed signals of a number of vehicles embedded in the traffic flow along the day. In order to quantify the classifier accuracy, ten minutes of road traffic have been tagged in terms of vehicle types and driving conditions, and finally reduced in Morel's categories.

With the 186 *in situ* tagged signals added to the training data set for the SVM model fitting, the global classification error is minimized to 13.94% for only six combinations of the 26,301 ones found previously. One of the possible weighting vector is:

$$\text{DataWeighting} = [\underbrace{5,1,0,4,2,0,2,0,2,1,0,0}_{\text{MFCC}}, \quad \underbrace{1,1,1}_{\text{Driving conditions}} \quad].$$

The confusion matrix is given table 1. It exhibits the classifier performance, comparing the expected categories (manually tagged categories, i.e. the "ground truth") to the ones estimated by the SVM

algorithm. When the matrix is diagonal, the classifier is 100% functional. Though promising, the results still bare substantial errors for the first category (two-wheel vehicles at constant speed) and for the seventh one (heavy vehicles in acceleration). The number of vehicle of the 3rd and 6th categories appears to be overestimated while in the others, the vehicle number is underestimated. This error can be use to adjust the classification results.

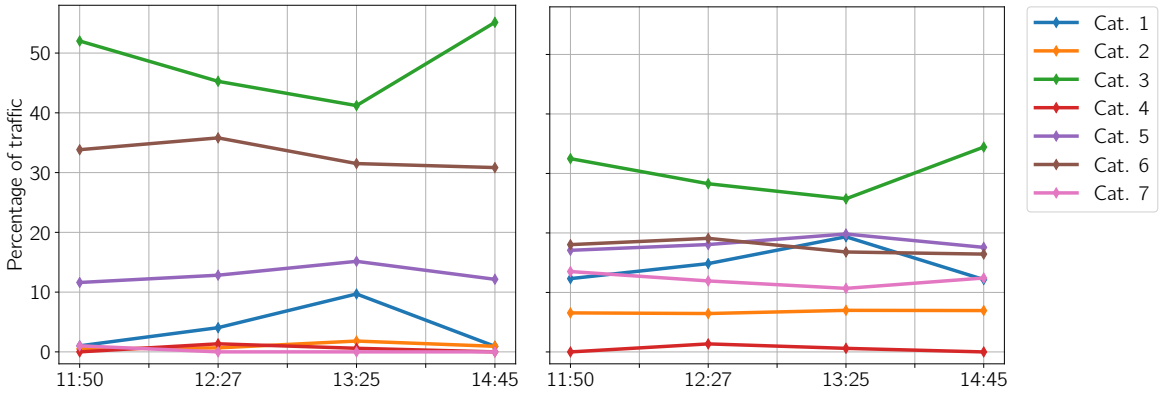
		Estimated						
		Cat.1	Cat.2	Cat.3	Cat.4	Cat.5	Cat.6	Cat.7
Expected	Cat.1	69.23	7.69	15.38	0	0	0	7.69
	Cat.2	0	75	16.67	0	0	8.33	0
	Cat.3	0.68	0	90.41	0	0	8.9	0
	Cat.4	0	0	0	100	0	0	0
	Cat.5	3.12	0	9.38	0	81.25	6.25	0
	Cat.6	0	2.86	5.71	0	5.71	85.71	0
	Cat.7	0	0	0	0	0	37.5	62.5
	Sum	73.03	85.55	137.55	100	86.96	146.69	70.19

Table 1 – Confusion matrix in percentage of vehicle signal per perceptive category - Only test track data available for the 4th category

This classification process is applied to other data recorded at the same place and various moments in the day to account for the variations of the road traffic. Figure 8a presents the results. According to the previous results the global classification error is estimated to about 15%. The proportions of the different categories in the traffic show to be quite stable, except at 1:25 pm where the number of two-wheel vehicles passing at constant speed (cat. 1) increases to 9.69%, while the light and heavy vehicles passing at constant speed (cat. 3) decreases from 52% (at 11:50 am) to 41%. The fourth category is underestimated because of the difficulty to extract the two-wheel vehicles trajectories when passing by other idling vehicles.

We can also notice that the location of the experiment participates to the classification statistics. Indeed, as presented in figure 4b, the system was set-up close to a traffic light so that the one-way part of the road bares mostly accelerating vehicles while on the three-way part, the vehicles are mostly at constant speed when passing in front of the array.

The global error of estimation presented in the table 1 is used to adjust the classification results. The adjusted results are represented in figure 8b. The adjustments induce the 5th category (light and heavy vehicles in deceleration) to be as important as the 6th one. The categories 1 and 7 have also been significantly raised up.



(a) Raw results (b) Adjusted results

Figure 8 – Percentage of the traffic flow composition over time in terms of Morel's categories

The proposed classification process appears to be working in the urban environment with an overall error around 15%. The relative stability of the results obtained at different moments around midday guarantees the robustness of the process.

5. CONCLUSIONS

This work investigates the usability of microphone arrays for the purpose of urban road traffic classification according to perceptual categories. First, a three steps method has been proposed to classify the road traffic into perceptual categories (grouping types and driving conditions of the vehicles). It is based on a Support Vector Machine algorithm using the vehicle audio signal features (MFCC) which are

extracted from the global acoustic scene thank to the beamforming technique and the video tracking.

This classification process has first been tested on vehicles isolated on a test track and the importance of the driving condition knowledge has been pointed out in order to reach a fully successful process.

The proposed method has been applied *in situ*. Ten minutes of a real traffic flow have been tagged in terms of perceptual categories and this data set has been added to a machine learning training set in order to provide a more robust classifier. The resulting global error is 13.94%, which is quite good, regarding the classical machine learning based classifiers capabilities [14, 13]. Finally, the classifier has been applied to other data recorded at the same date and appears to be quite robust, given the stability of the consistent analysis.

The paper demonstrates the capability of the beamforming technique associated with machine learning algorithms to analyze the urban road traffic in terms of typical noise sources categories. The further studies will consist of translating these classified data in terms of annoyance by implementing a psychoacoustic model yet to be defined.

ACKNOWLEDGEMENTS

The authors wish to thank: Dominique BUSQUET (UPMC), Pascal CHALLANDE (UPMC), Jean-Christophe CHAMARD (PSA), Hélène MOINGEON (UPMC), Christian OLLIVON (UPMC) et Vincent ROUSSARIE (PSA).

This research benefited from the support of the Chair "Mobilty and quality of life in urban surroundings" led by The UPMC Foundation and sponsored by the Donors (PSA Peugeot-Citroën and RENAULT)

REFERENCES

[1] Directive européenne 2002/49/ce relative à l'évaluation et à la gestion du bruit dans l'environnement, 2002.

[2] Raymond Murray Schafer. *Le Paysage Sonore*. Wildproject, 4ème edition, 2010.

[3] A. Klein, C. Marquis-Favre, R. Weber, and A. Trollé. Spectral and modulation indices for annoyance-relevant features of urban road single-vehicle pass-by noises. *J. Acoust. Soc. Am.*, 137(3):1238–1250, Mar 2015.

[4] J. Morel, C. Marquis-Favre, and L.-A. Gille. Noise annoyance assessment of various urban road vehicle pass-by noises in isolation and combined with industrial noise: A laboratory study. *Applied Acoustics*, 101:47–57, Jan 2016.

[5] J. Morel, C. Marquis-Favre, D. Dubois, and M. Pierrette. Road traffic in urban areas: A perceptual and cognitive typology of pass-by noises. *Acta Acustica united with Acustica*, 98(1):166–178, Jan 2012.

[6] Nasser Mohammadiha, Paris Smaragdis, and Arne Leijon. Supervised and unsupervised speech enhancement using nonnegative matrix factorization. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(10):2140–2151, oct 2013.

[7] Emmanuel Vincent, Shinji Watanabe, Aditya Arie Nugraha, Jon Barker, and Ricard Marxer. An analysis of environment, microphone and data simulation mismatches in robust speech recognition. *Computer Speech & Language*, dec 2016.

[8] Yong Xu, Jun Du, Li-Rong Dai, and Chin-Hui Lee. A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(1):7–19, jan 2015.

[9] John E. Adcock. *Optimal filtering and speech recognition with microphone arrays*. PhD thesis, Brown University, 2001.

[10] Ines Hafizovic, Carl-Inge Colombo Nilsen, Morgan Kjølerbakken, and Vibeke Jahr. Design and implementation of a MEMS microphone array system for real-time speech acquisition. *Applied Acoustics*, 73(2):132–143, Feb 2012.

[11] C. Vanwynsberghe, R. Marchiano, F. Ollivier, P. Challande, H. Moingeon, and J. Marchal. Design and implementation of a multi-octave-band audio camera for realtime diagnosis. *Applied Acoustics*, 89:281–287, Mar 2015.

[12] Chien-Chang Lin, Shi-Huang Chen, Trieu-Kien Truong, and Yukon Chang. Audio classification and categorization based on wavelets and support vector machine. *IEEE Transactions on Speech and Audio Processing*, 13(5):644–651, sep 2005.

[13] Asma Rabaoui, Manuel Davy, Stéphane Rossignol, and Noureddine Ellouze. Using one-class SVMs and wavelets for audio surveillance. *IEEE Transactions on Information Forensics and Security*, 3(4):763–775, dec 2008.

[14] Xavier Valero and Francesc Alías. Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys. *Archives of Acoustics*, 37(4), jan 2012.

[15] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics - Facts and Models*. Berlin : Springer Verlag, 3rd edition, 2007.

[16] Raphaël Leiba, François Ollivier, Régis Marchiano, Nicolas Misdariis, and Jacques Marchal. Urban acoustic imaging : from measurement to the soundscape perception evaluation. In *INTER-NOISE*, 2016.