

# Estimation of the road traffic sound levels based on Non-Negative Matrix Factorization

Jean-Rémy GLOAGUEN  
Arnaud Can  
LAE  
Ifsttar  
jean-remy.gloaguen@ifsttar.fr

Mathieu Lagrange  
Jean-François Petiot  
LS2N  
Central School of Nantes

## Abstract

## 1 Introduction

### 1.1 Related work

With the introduction of the European Directive 2002/EC/49, cities over 100 000 inhabitants have to produce road traffic noise maps. These maps allow to estimate the number of city dwellers exposed to high noise levels and to draw up action plans to reduce it as too long exposures to these noises can generate health problems [?]. These maps are the result of a simulation process based on the estimation of the traffic density on the main roads and the use of sound propagation techniques. They express  $L_{DEN}$  and  $L_N$ , which are *Day-Evening-Night* and *Night* equivalent A-weighted sound levels respectively. However, these maps introduce lot of uncertainty generated by the numerical tools [?], by the different calculation methodologies used [?][?] or even by the calculation procedure of the number of inhabitants exposed to noise [?]. In addition, the usual road traffic noise maps are static, aggregating the exposure on the two indicators  $L_{DEN}$  and  $L_N$ , thus ignoring the sound levels evolution throughout the day. Since the creation of road traffic noise maps entails long data collection and calculation times, the use of acoustic measurements could facilitate their updating or even the generation of dynamic maps [?]. These measurements can be performed at fixed stations spread all over the cities [?] [?], which would lead to the availability of the long-term evolution of the traffic noise levels. It can also be performed with mobile stations [?] [?] covering a larger area with fewer sensors but also sparse time periods.

Currently, sensor networks in cities are spread for multiple applications (air quality assessment, measurement of meteorological parameters, ...), including the assessment of urban noise levels. DYNAMAP project [?] studied the establishment and feasibility of such installations. It focuses on sensor installations on specific roads at the

city scale in Milan and Rome [?]. In a similar way, but reduced to few neighborhoods, the CENSE project<sup>1</sup> [?] aims to combine *in situ* observations, from a sensor network, and numerical data, from noise modeling, through data assimilation techniques.

If sensors networks could improve road traffic noise estimation compared with simulated maps, the issue of the correct estimation from measurements of the traffic sound level is still unsolved [?]. Indeed, the urban sound environment is a complex environment gathering lots of different sounds (car passages, voices, bird's whistles, car horn ...) that can overlap. Consequently, the traffic sound level estimation based on measurements is not trivial task. Many recent works have focused on the detection or recognition tasks of environmental sounds without distinction between them[?], [?], [?], [?]. A two step process is generally followed : describe the audio files with a set of features (Spectrum Gravity Spectrum, harmonicity, Mel-Frequency Cepstral Coefficient ...) and classify them with the help of classifiers (Support Vector Machines, Gaussian Mixture Models, Hidden Markov Model, Artificial Neural Networks). A description of there features and classifiers can be found in [?] and their application can be found in [?], [?], [?]. Recently, an Anomalous Noise Events Detector has been generated in [?] to detect the sound sources from labeled recordings that are not related to the traffic component in order not to take them into account on the estimation of the traffic sound level. If the detection of the road traffic noise is good, the detection of these anomalous noise events stay weak and no information on the improvement on the estimation of the traffic nous level are presented.

Furthermore, this work and as well as the other works in the detection or recognition tasks, do not address the overlap of environmental sounds in an urban context. Although near major roads or ring roads traffic is predominant on all other sound sources, there are many places

---

<sup>1</sup><http://cense.ifsttar.fr/>

where road traffic overlaps with other sound sources that contribute significantly to the overall sound levels. In such case, the only detection of the traffic component does not make it possible to determine precisely its noise level. In consequence, to be effective on a wide range of sound environments, we propose in this paper to follow the blind source separation paradigm. That is, separating the contribution of the traffic from the other sources within a polyphonic scene.

One of the first and the most widely used techniques to do so is the Independent Component Analysis [?]. The principle is to decompose  $N$  recorded signals to a sum of  $P$  independent sound sources weighted by linear relations. This method is most of all suited for the 'cocktail party' issue where one tries to capture a signal among noise. However, ICA is limited to only over determined cases ( $N > P$ ). Furthermore, if it is suited for indoor environments where the number of sound sources is constant, it can not be fitted for an outdoor environment where the number of sources is unknown and variable and, moreover, it would be necessary to mount multiples sensors on one point to perform the source separation. A more convenient method is Non-negative Matrix Factorization (NMF) [?] which consists in approximating the magnitude spectrogram of an audio file from the product of two matrices. It has been widely used in the audio domain, [?] [?] [?], and has already been employed for the source separation task of monaural signals of speech and music [?] [?]. By design, this method deals with the overlapping sound sources as soon as the overlap can be resolved on the time/frequency plane. For the environmental sounds, the method has been used for the geo-localisation and classification of the sound environment, like in [?] where NMF is used to classify the audio files according to the 10 cities where they have been recorded. It has also been used by Innami and Kasai in the unsupervised case [?] for source separation. They proposed a source separation in two steps by separating the sound background from the events first and by separating the events between them. The audio files tested results of a simulation process where a sound background (river or wind) are adding to two sound events (school chime, announcement, frog croaking, dog barking and bell ringing). If the method proposed is interesting, the main issue here is the small size of the database (only 9 sounds) on which the algorithms are tested while some sounds (frog and river) are not representative of sounds that can be found in cities.

## 1.2 Proposed approach

We propose in this paper a method based on the Non-Negative Matrix Factorization (NMF) technique to estimate the global,  $\bar{L}_{p,traffic}$ , and the 1-s equivalent,  $\tilde{p}_{1s,traffic}$  sound level of the traffic through the super-

vised and the semi-supervised approaches as well as a method using thresholding. To validate these approaches, we consider a corpus of simulated scenes artificially created with the simulator software *simScene*. The use of simulated sound scenes is necessary as it offers a full control on the design of the scenes and the knowledge of the exact contribution of the traffic component which would hardly be extracted from a recording of an urban scene ( $L_{p,traffic}$  and  $p_{1s,traffic}$ ). Both the sound scene simulation and NMF require the creation of two sound databases (figure ??). In parallel, a baseline method built from a frequency low-pass filter is computed. This method considers that road traffic is mainly composed of low frequencies and therefore can be filtered by a low-pass filter at the cut-off frequency  $f_c$  (figure ??). The performance of the frequency low-pass filter and NMF are then compared with the calculation of two metrics (Mean Absolute Error, normalized Root Mean Square Error).

The remaining of the paper is organized as follows. Section 2 details the technical aspect of NMF. Section 3 described on the design of the environmental sound scene corpus and the experimental protocol setup. Then Section ?? shows and discusses the results obtained during the parametric study.

## 2 Non-negative Matrix Factorization

### 2.1 Description of NMF

Non-negative Matrix Factorization is a matrix approximation method introduced by Lee and Seung, [?], which can be used to approximate the spectrogram (obtained using a Short-Term Fourier Transform) of an audio file,  $\mathbf{V} \in \mathbb{R}_{F \times N}^+$  as :

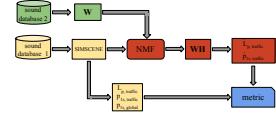
$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{W}\mathbf{H} \quad (1)$$

where  $\mathbf{W} \in \mathbb{R}_{F \times K}^+$  is the *dictionary* (or basis) matrix composed of audio spectrum and  $\mathbf{H} \in \mathbb{R}_{K \times N}^+$  is the *activation* matrix which summarizes the temporal evolution of each element of  $\mathbf{W}$  (fig. 2).

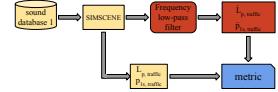
The choice of the dimensions is often made as that  $F \times K + K \times N < F \times N$ . NMF is then considered as a low rank approximation method. However, this constraint is not essential. To estimate the quality of the approximation, an objective function is used

$$\min_{\mathbf{H} \geq 0, \mathbf{W} \geq 0} D(\mathbf{V} || \tilde{\mathbf{V}}). \quad (2)$$

The operator  $D(x|y)$  is a divergence calculation such



(a) Frequency low-pass filter with  $f_c = 500$  Hz



(b) Block diagrams summed up the different step of the process for the frequency low-pass filter ?? and for NMF ??

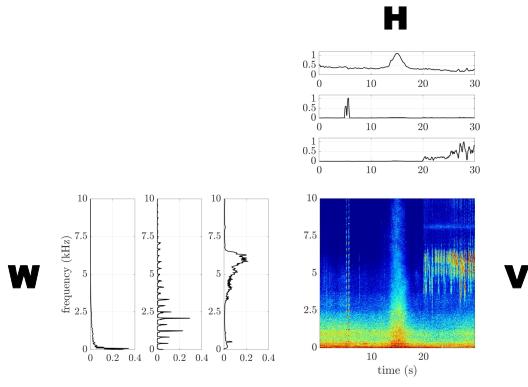


Figure 2: Example of a simple NMF for urban sound mixture,  $\mathbf{W}$  and  $\mathbf{V}$  are composed of 3 elements (car passages, car horn and bird's whistles)

as:

$$D(\mathbf{V} || \tilde{\mathbf{V}}) = \sum_{f=1}^F \sum_{n=1}^N d_\beta (\mathbf{V}_{fn} || [\mathbf{WH}]_{fn}) \quad (3)$$

and usually belongs to the  $\beta$ -divergence class [?] in which the well known Euclidean distance (eq. 4a) and the Kullback-Leibler divergence (eq. 4b) belong

$$d_\beta(x|y) = \begin{cases} \frac{1}{2}(x-y)^2, & \beta = 2, \\ x \log \frac{x}{y} - x + y, & \beta = 1. \end{cases} \quad (4a) \quad (4b)$$

Prior knowledge on the content can be adjusted with the addition of constraints (like the smoothness or the sparseness criteria [?]) in the objective function (equation (2)) to better take account prior knowledge of the sources.

Algorithms have been proposed to solve the minimization problem (2) iteratively such as the multiplicative update [?], the alternating least square method [?], the projected gradient [?] ... Here, the multiplicative update is chosen as it ensure non-negative results of which convergence has been proved [?].

## 2.2 Supervised NMF

First, supervised NMF is used: the *dictionary* includes audio spectrum of urban sound sources as, in the urban environments, a lot of different sound sources present are known and their spectrum can be obtained. The *basis* are then the unknown to estimate. In the first iteration,  $\mathbf{H}$  is initialized randomly, then it is updated by the generic algorithm

$$\mathbf{H}^{(i+1)} \leftarrow \mathbf{H}^{(i)} \cdot \left( \frac{\mathbf{W}^T \left[ (\mathbf{WH}^{(i)})^{(\beta-2)} \cdot \mathbf{V} \right]}{\mathbf{W}^T \left[ \mathbf{WH}^{(i)} \right]^{(\beta-1)}} \right)^{\gamma(\beta)} \quad (5)$$

with  $\gamma(\beta) = \frac{1}{2-\beta}$ , for  $\beta < 1$ ,  $\gamma(\beta) = 1$ , for  $\beta \in [1, 2]$  and  $\gamma(\beta) = \frac{1}{\beta-1}$  for  $\beta > 2$ . The product  $A \cdot B$  and  $A/B$  symbolized the Hadamard product and ratio. As in the supervised approach, the position in  $\mathbf{W}$  of traffic component is known, the source separation of this sound source is made by extracting the dictionary and basis elements related,

$$\tilde{\mathbf{V}}_{traffic} = [\mathbf{WH}]_{traffic}. \quad (6)$$

## 2.3 Semi-supervised NMF

One of the main issue with the supervised approach is the generalization issue: how to be adapted to different sound mixtures with a fixed dictionary ? To better take into account the diverse nature of urban scenes, semi-supervised NMF can be useful as it has been proposed [?] to offer more flexibility. This method consists in composing the *dictionary* with a fixed part  $\mathbf{W}_s \in \mathbb{R}_{F \times K}^+$ , composed in our case of spectrum representative of road traffic and with a mobile part,  $\mathbf{W}_r \in \mathbb{R}_{F \times J}^+$  with  $J \ll K$ , that is updated. Here,  $J = 2$ . The aim is to include in  $\mathbf{W}_r$  the element that are not related with the traffic. The problem (1) become

$$\mathbf{V} \approx \mathbf{W}_s \mathbf{H}_s + \mathbf{W}_r \mathbf{H}_r. \quad (7)$$

In a similar way as to solve the equation 2,  $\mathbf{W}_r$ ,  $\mathbf{H}_r$  and  $\mathbf{H}_s$  are successively updated with the relations (8):

$$\mathbf{W}_r^{(i+1)} \leftarrow \mathbf{W}_r^{(i)} \cdot \left( \frac{\left[ (\mathbf{W}_r \mathbf{H}_r^{(i)})^{(\beta-2)} \cdot \mathbf{V} \right] \mathbf{H}_r^T}{(\mathbf{W}_r \mathbf{H}_r^{(i)})^{(\beta-1)} \mathbf{H}_r^T} \right)^{\gamma(\beta)}, \quad (8a)$$

$$\mathbf{H}_r^{(i+1)} \leftarrow \mathbf{H}_r^{(i)} \cdot \left( \frac{\mathbf{W}_r^T \left[ (\mathbf{W}_r \mathbf{H}_r^{(i)})^{(\beta-2)} \cdot \mathbf{V} \right]}{\mathbf{W}_r^T (\mathbf{W}_r \mathbf{H}_r^{(i)})^{(\beta-1)}} \right)^{\gamma(\beta)}, \quad (8b)$$

$$\mathbf{H}_s^{(i+1)} \leftarrow \mathbf{H}_s^{(i)} \cdot \left( \frac{\mathbf{W}_s^T \left[ (\mathbf{W}_s \mathbf{H}_s^{(i)})^{(\beta-2)} \cdot \mathbf{V} \right]}{\mathbf{W}_s^T (\mathbf{W}_s \mathbf{H}_s^{(i)})^{(\beta-1)}} \right)^{\gamma(\beta)}. \quad (8c)$$

## 2.4 Thresholded constrained NMF

A last approach is tested based on unsupervised NMF. Usually,  $\mathbf{W}$  is learnt with the help of a learning corpus by initiated it randomly. Here, as the concerned sound source is known and audio samples of car passages are available, a initial dictionary,  $\mathbf{W}_0$ , is learnt by converting the audio files in the spectra domain (see part 3.2.1). Then NMF is performed where  $\mathbf{W}$  (eq. 9) and  $\mathbf{H}$  (eq. 5) are updated alternatively.  $\mathbf{W}$  is therefore updated by forcing its initiation with *a priori* knowledge.

$$\mathbf{W}^{(i+1)} \leftarrow \mathbf{W}^{(i)} \cdot \left( \frac{\left[ (\mathbf{W}^{(i)} \mathbf{H})^{(\beta-2)} \cdot \mathbf{V} \right] \mathbf{H}^T}{[\mathbf{W}^{(i)} \mathbf{H}]^{(\beta-1)} \mathbf{H}^T} \right)^{\gamma(\beta)} \quad (9)$$

After  $N$  iterations, a measure of similarity  $D_\theta(\mathbf{W}_0 || \mathbf{W})$  between  $\mathbf{W}_0$  and the get dictionary  $\mathbf{W}$  for each element  $k$  is computed through a cosine similarity,

$$\cos \theta = \frac{\mathbf{W} \cdot \mathbf{W}_0}{\|\mathbf{W}\| \cdot \|\mathbf{W}_0\|}. \quad (10)$$

$\cos \theta = 1$  means that the elements are identical (the  $k$ -th element of  $\mathbf{W}$  is then considered as traffic element) whereas  $\cos \theta = 0$  means that the elements are significantly different. This measure allows a bound between 1 and 0 and is an invariant scale estimation of the similarity. Then, the similarity is sorted in descending order. The elements in  $\mathbf{W}$  that can belong to  $\mathbf{W}_{traffic}$  are then selected by a *hard thresholding* method (Figure 3). It is defined as :

$$\mathbf{W}_k \in \mathbf{W}_{k,traffic} \text{ iff } D(\mathbf{W}_{0,k} || \mathbf{W}_k) > t \quad (11)$$

This approach is named *Thresholded constraint NMF* (ThC NMF). Other thresholding methods as the *soft* [?] and the *firm* [?] and multiples way to display the distance through a sigmoid or a Radial Basis Function have

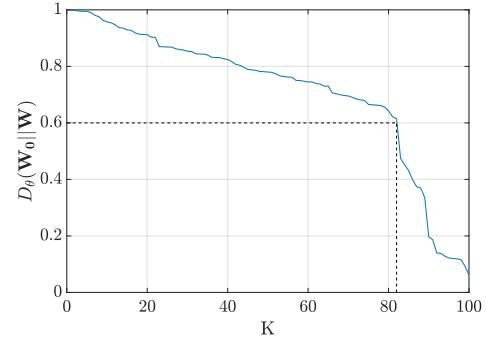


Figure 3: Example of the  $\mathbf{W}_{traffic}$  extraction from the sorted cosine similarity with a threshold  $t = 0.6$ . The 82-nd first elements are considered as traffic component.

been investigated. A fast parametric study has revealed that the *hard* thresholding method with a linear representation of the similarity according to  $K$  (see fig. 3) was the best way to get better performances.

## 3 Experimental protocol

In order to validate the usefulness of considering NMF framework to estimate the road traffic noise level, one need to have a reference level. It can hardly be measured or even annotated from real life recordings. Thus, simulated sound scenes are used to assess the performance of the proposed NMF. This offers a controlled framework to design specific sound environments in which all the traffic component is known. Then, the road traffic sound levels estimated with the method can be compared to the real ones, introduced within each simulated sound scene.

### 3.1 Environmental sound scene corpus

A corpus is designed with the *simScene* software<sup>2</sup>. *simScene* [?] is a simulator that creates sound scenes in a .wav format by summing audio samples that come from an isolated sound database.

This database is divided in two categories: *i*) the *event* category which are the brief sounds (from 1 to 20 seconds) that are considered as salient including 245 sound event samples divided in 19 sound classes (*ringing bell, birds, sweeping broom, car horn, car passages, hammer, drill, coughing, barking dog, rolling suitcase, closing door, plane, siren, footprint, storm, street noise, train, tramway, truck and voice*) and *ii*) the *background* category that includes all the sounds that are of long duration and whose acoustic properties

<sup>2</sup>Open-source project available at: <https://bitbucket.org/mlagrange/simscene>

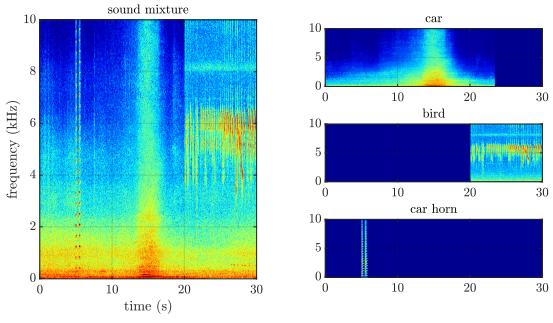


Figure 4: Example of a sound scene composed of 3 sound classes (car, bird, car horn)

do not vary with respect to time. 154 sound samples belong to this category divided in 9 sound classes (*birds, construction site noise, crowd, park, rain, children playing in schoolyard, constant traffic noise, ventilation, wind*). The sound class *car passages* comes from 60 recordings of 2 cars made on the Ifsttar's runway on different speeds with multiple gear ratio. The other audio files have been found online (*freesound.org*) and within the *UrbanSound8k* database [?]. Each sound classes is composed of multiples samples (*bird01.wav, bird02.wav ...*). The software allows the user to control some parameters (number of events of each class that appear in the mixture, elapsed time between each sample of a same class, presence of a fade in and a fade out ...) completed with a standard deviation that may bring some random behavior between the scenes. Furthermore, an audio file of each sound class present in the scene can be generated that allows to know its exact contribution as well as a text file that summarizes the time presence of all the events.

This database enables creating realistic urban sound scenes from the road traffic point of view [?]. A sound mixing corpus is composed of 6 sub-corpus of 25 audio files each lasting 30 seconds. Each sub-corpus is characterized by a specific generic sound class that summed with traffic will make the estimation of the traffic level more difficult. The classes are : *alert* (car horn, siren), *animals* (barking dog , whistling birds), *climate* (wind, rain), *humans* (crowd noise and voice), *mechanics* (different metallic and construction site noises) and *transportation* (train, tramway and plane). In each file, traffic component is present as the sum of the background and event traffic sounds and is mixed with the sound classes. The sound classes that are not related to the traffic component are summed up as the *interfering* sound class. To test different scenarios, each audio file is duplicated with the traffic sound level of the entire sound scene,  $L_{p,traffic}$ , fixed to a specific level according to the sound

level of the *interfering* class,  $L_{p,interfering}$  following the relation (12).

$$TIR = L_{p,traffic} - L_{p,interfering} \quad (12)$$

with the *Traffic Interference Ratio*  $TIR = [-12, -6, 0, 6, 12]$ . When  $TIR = -12$ , the traffic component is then less present than when  $TIR = 12$  where it is predominant on the *interfering* class. The 1 second equivalent sound pressure level,  $p_{1s,traffic}$ , is also calculated (figure 5). The total number of scenes designed is 750 (6 sub-corpus  $\times$  25 scenes  $\times$  5 TIR values).

## 3.2 Experiment

The experiment consists in estimating the traffic road sound level of the 6 environmental sound sub-corpus (*alert (al), animals (an), humans (hu), climate (cl), mechanics (me), transportation (tr)*) and for 5 *TIR* ([ $-12 -6 0 6 12$ ] dB). First, the spectrogram  $\mathbf{V}$  of each sound scene is built with a window size  $w = 2^{12}$  with a 50 % overlap and a number of point  $nfft = 2^{12}$ . Therefore, the dimensions of  $\mathbf{V}$  are  $F = 2049$  and  $N = 664$ .

The first estimator to determine the traffic sound level is a basic frequency low-pass filter which depend only on the cut-off frequencies  $f_c = [500 \text{ } 1\text{k} \text{ } 2\text{k} \text{ } 5\text{k} \text{ } 10\text{k} \text{ } 20\text{k}] \text{ Hz}$ . The spectrograms  $\mathbf{V}$  are filtered and the remaining energy is then considered as traffic component (eq. 13).

$$\tilde{\mathbf{V}}_{traffic} = \mathbf{V}_{f_c}. \quad (13)$$

The second estimator is NMF. Multiples parameters are involved here between the dictionary learning and the own parameters of the NMF.

### 3.2.1 Dictionary building

The dictionary is built from a second sound database dedicated specifically to this task. It is composed of 53 audio files of passing cars. These records have been made on the Ifsttar's runway with the same conditions that the records made for the *SimScene* database but with two different cars (). First, for each audio file, its spectrogram is calculated with fixed parameters ( $w$ , 50 % overlap,  $nfft$ ). Then time/frequency windows with  $w_t \times F$  dimension are applied without overlapping on the spectrogram in order to consider several spectrum for each audio file.  $w_t$  is fixed at  $w_t = [0.51]$  second. In each window, the root mean square value is calculated on each frequency bin to reduce the different spectrum in one spectra. Since the number of elements given by processing all the sound database is high, in order to reduce the computational and delete redundant information, a  $K$ -means clustering is applied to reduce the number of spectrum to  $K = [25, 50, 100]$ . A special case is added

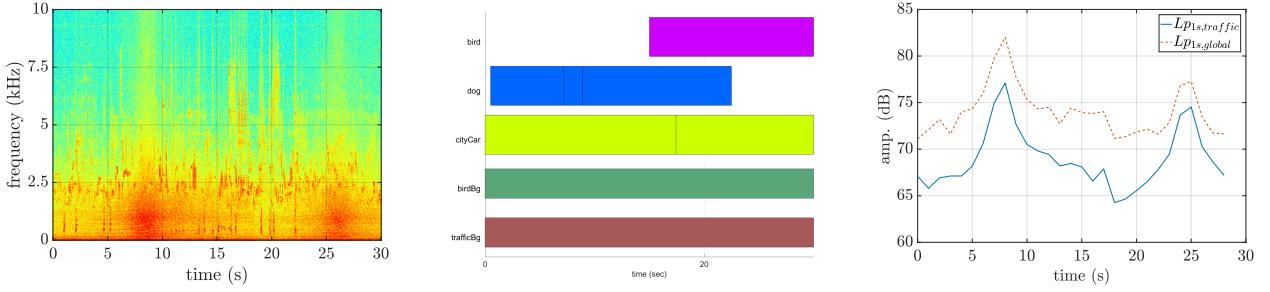
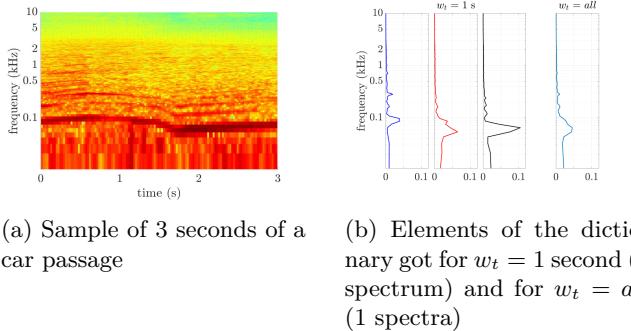


Figure 5: Example of a scene of the *animals* sub corpus. Spectrogram (on left), *Piano Roll* of the different sound classes (on the middle) and 1-s equivalent sound level of the traffic,  $L_{p1s, \text{traffic}}$  and of the global sound scene,  $L_{p1s, \text{global}}$  (on right)



where the root mean square of *all* the spectrogram is applied. Each audio file generates one element  $k$  of  $\mathbf{W}$ . An example that illustrates the process can be found on figure ?? on 3 seconds extract of a spectrogram of a car passage (fig. 6a). In the case where  $w_t = 1$  second , 3 elements are therefore extracted of the spectrogram while in the case where  $w_t = \text{all}$ , all the spectrogram is reduced to one element (fig. 6b).

Each  $k$  element of  $\mathbf{W}$  is normalized such as  $\|\mathbf{W}_k\| = 1$  with  $\|\bullet\|$  the  $\ell$ -1 norm.

Table 1 summarizes the parameters and their related values.

Parameter	value		
$K$	25	50	100
$w_t$ (s)	0.5	1	<i>all</i>

Table 1: Summary of the dictionary parameters

### 3.2.2 NMF parameters

Supervised and semi-supervised NMF are performed for 400 iterations which is sufficiently enough to get a stabilized reconstruction. ThC-NMF is performed on a lower number of iteration (60) to prevent  $\mathbf{W}$  to not deviate too much from the initial dictionary. The spectrogram

$\mathbf{V}$  and the dictionary  $\mathbf{W}$  are expressed on two different spectral representations: with a linear frequency scale ( $\Delta f \approx 10.8$  Hz) and with third octave bands (29 bands). These two methods are considered to compare a fine grain approach (the linear scale) with a coarser one (the third octave bands) as it reduces the number of frequency bins in the high frequencies where the traffic component is less present and generates less calculations. Furthermore, in the case of the linear frequency scale and for supervised and semi-supervised NMF,  $\mathbf{V}$  and  $\mathbf{W}$  are filtered at the frequencies  $f_c$  in order to focused the reconstruction of the signal of the low frequency bins. Nevertheless, if NMF is performed with filtered elements ( $\mathbf{V}_{f_c}$  and  $\mathbf{W}_{f_c}$ ) to determine  $\mathbf{H}_{f_c}$ , the traffic signal reconstruction is done with the original dictionary  $\mathbf{W}$ , as

$$\tilde{\mathbf{V}}_{\text{traffic}} = [\mathbf{W}\mathbf{H}_{f_c}]_{\text{traffic}}. \quad (14)$$

With the third octave spectral representation, only the case  $f_c = 20$  kHz is applied. For the ThC NMF, the threshold is define between 0.20 and 0.70 with a 0.01 step. Table ?? summarizes the parameters and their related values.

### 3.2.3 Metrics

The performances of the two estimators of the road traffic sound level are assessed through the calculation of one metric, the Mean Absolute Error (*MAE*). It expresses the quality of the long-term reconstruction of the signal. It consists in the average over the  $N$  sound scenes of the absolute difference between the exact and estimated traffic sound level in dB,

$$MAE = \frac{\sum_{n=1}^N |L_{p,\text{traffic}}^n - \tilde{L}_{p,\text{traffic}}^n|}{N}. \quad (15)$$

In all, according Tables 1 and ??, 24540 settings are performed between the different form of the dictionary  $\mathbf{W}$  (table 1) and the multiple parameters taken into account by NMF: for the filter estimator, between the