

Estimation of the traffic sound level from measurements made in an urban environment

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

Road traffic noise is one of the main annoyance perceived by urban dwellers and a major societal issue because too long exposures to high sound levels can generate health problems [1]. Accordingly, the European Directive 2002/EC/49 [2] has been developed to determine the exposure of the populations. Cities with a population of more than 100 000 must produce road traffic noise maps. These maps come from a simulation process resulting from the estimation of the traffic density on the main roads and calculating sound propagation. The maps express L_{DEN} and L_N , respectively, *Day-Evening-Night* and *Night* equivalent A-weighted sound levels. These maps make it possible to estimate the number of city dwellers exposed to highly annoying noise levels and to draw up action plans to reduce them. However, these maps introduce lot of uncertainty generated by the numerical tools [3], by the different calculation methodologies used [4][5] or even by the calculation procedure of the number of inhabitants exposed to noise [6]. 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 measurements could facilitate their updating or even the generation of dynamic maps [7]. These measurements can be performed at fixed stations spread all over the cities [8] [9], which allow to get the long-term evolution of the traffic noise levels, or can be performed with mobile stations [10] [11] covering a larger area with fewer sensors but also sparse time periods. The clustering between mobile and fixed measurements have been studied in [12].

Currently, we are seeing the diffusion of sensor networks in cities for multiple applications (air quality assessment, measurement of meteorological parameters),

including assessment of urban noise levels. Various projects exist to study the establishment and feasibility of such installations as the DYNAMAP [13] project. It focuses on sensor installations on specific roads in all cities of Milan and Rome [14]. In a similar way, but reduced to few neighborhoods, the CENSE project¹ [] plans to combine *in situ* observations, from a sensor network, and numerical data, from noise modeling, by using data assimilation techniques.

If these installations, from their measurements, allow to have a more precise evaluation of the traffic noise in cities than the simulated maps, the issue of the correct estimation from measurement of the traffic sound level is still unsolved [8]. Indeed, the urban sound environment is a complex environment gathering lots of different sounds both from a temporal and a frequency point of view (car passages, voices, bird's whistles, car horn ...) that can overlap. In consequence, the traffic sound level estimation is not trivial task.

Many works have been focused on the detection [15] or on recognition [16] task of environmental sounds. These methods use a two step process : describe the audio files with a set of features (Spectrum Gravity Spectrum, harmonicity, Mel-Frequency Cepstral Coefficient ...) and classified them with the help of classifiers (Support Vector Machines, Gaussian Mixture Models, Hidden Markov Model, Artificial Neural Networks). A description of these features and classifiers can be found on [17] and their application can be found in [18], [19], [20]. However, many of these results in the detection or recognition tasks, do not address the overlap of environmental sounds in an urban context. Although in some places, near major roads or ring roads, traffic is predominant on all other sound sources, there are many places where there is little traffic and overlaps with other sound sources. The only detection of the traffic component does not make it possible to determine precisely its noise level. In consequence, to be effective on all the

¹<http://cense.ifsttar.fr/>

different sound environments, it seems more suited to consider the issue as the one of blind source separation.

One of the first and the most widely used techniques is the Independent Component Analysis [21]. The principle is to decompose N recorded signals in 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 install multiples sensors on one point to realize the source separation which is not feasible. A second method is Non-negative Matrix Factorization (NMF) [22] which consists in approximate the non-negative spectrogram of an audio file from the product of two matrices. It has been widely used in the audio domain, [23] [24] [25], and has already been employed for the source separation task of a monaural signal included speech and musical contents [26] [24]. This method has the advantage to easily deals with the overlap of the sound sources. For the environmental sounds, the method has been used for geo-localisation and classification of the sound environment like in [27] where NMF allows to classify the audio files according to the 10 cities where they have been recorded. For the source separation, it has been used by Innami and Kasai in the unsupervised case. They proposed a source separation in two step 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.

Our study proposed to applied Non-Negative Matrix Factorization on simulated sound mixtures to succeed on the estimation of the traffic sound level. The use of simulated sound mixtures 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 what recordings do not allow. Part 2 details the technical aspect of NMF. Part 3 is focused on the experimental protocol set up. Then part 4 reveals the results obtained during the parametric study.

2 Non-negative Matrix Factorization

2.1 Supervised NMF

Non-negative Matrix Factorization (NMF) is an approximation method introduced by Lee and Seung, [22], which estimates the spectrogram (get from a Short-Term Fourier Transform) of an audio file, $\mathbf{V} \in \mathbb{R}_{F \times N}^+$ as :

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

where $\mathbf{W} \in \mathbb{R}_{F \times K}^+$ is the *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. 1).

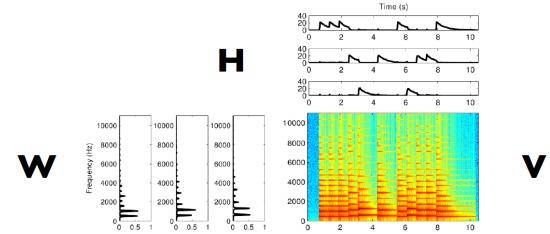


Figure 1: Example of a simple NMF for a musical content [28]

The choice of the dimensions is done in a such way that $F \times K + K \times N < F \times N$. 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 as

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

and usually belongs to the β -divergence class [29] in which the well known Euclidean distance (eq. 4a) and the Kullback-Leibler divergence (eq. 4b) belong

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

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

The prior knowledges on the content can be adjust with the add of constraints (like the smoothness or the sparseness criteria [30]) in the objective function (equation (2)).

Multiple methods exist to solve the minimization problem (2) (Alternating Least Square Method [31], Projected Gradient [32] ...). Here, the multiplicative update is the chosen method [33] as it assures to have a non-negative results where the convergence of the results have been proved [34]. The minimization problem (2) is solved iteratively. NMF is employed here in the supervised case: the *dictionary* is already known as in the urban environments, a lot of different sound sources present are known and their spectrum can be obtained. The *basis* are the unknown to find. 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. The source separation is made after I iterations either by extracting the dictionary and basis elements related to the traffic

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

2.2 Semi-supervised NMF

One of the main issue with the supervised approach is this is not always adapted to different sound environments. Because in the supervised case, the fixed dictionary has to sum up all the sound sources present in the different sound environment. In consequence, as the dictionary dimension is limited and cannot include all the urban sounds, to offer more flexibility, semi-supervised NMF has been proposed [35]. This method consists in composing the *dictionary* with a fixed part $\mathbf{W}_s \in \mathbb{R}_{F \times K}^+$ (composed here of traffic audio spectrum) and with a mobile part, $\mathbf{W}_r \in \mathbb{R}_{F \times J}^+$ that is updated. 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):

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

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

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

This approach has been used TO COMPLETE.

3 Experimental protocol

To assess the performance of NMF, simulated sound mixtures are used as it offers a controlled framework to design specific sound environments where all the traffic component is known. The estimation of the traffic sound level can be compare to an exact solution.

3.1 Environmental sound scene corpus

A first corpus is design with the *simScene* software². *simScene* [36] is a simulator that creates sound mixtures in a .wav format by superposing audio samples that come from an isolated sound database. This database is divided in two categories: the *event* category which are the brief sounds (from 1 to 20 seconds) that are considered as salient whereas all the sounds that are of long duration and whose acoustic properties do not vary with respect to time belong to the *background* category. Inside each category, the sound samples are grouped in sound classes (*bird*, *car*, *foot steps* ...), each of them being composed of multiples samples (bird01.wav, bird02.wav ...).

The software allows the user to control some parameters as the number of events of each class that appears in the mixture, the elapsed time between each sample of a same class or the presence of a fade in and a fade out Each parameter is completed with a standard deviation that may brings some random behavior between the scenes. Furthermore, with the global scene mixture, an audio for each sound class presented in the scene can be generated that allow to know its exact contribution and a text file is created that summarizes the time presence of all the events.

With this software, a sound database has been built including 245 sound event samples (with the sound classes *ringing bell*, *birds*, *sweeping broom*, *car horn*,

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

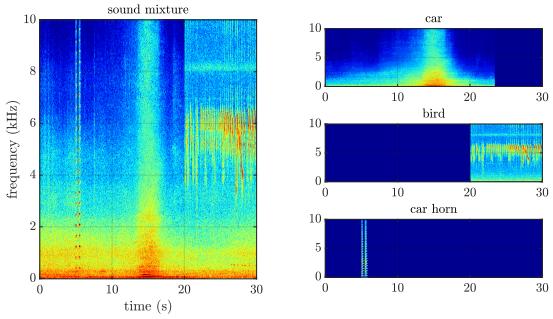


Figure 2: Example of a sound mixture composed of 3 sound classes (car, bird, car horn)

car passages, hammer, drill, coughing, barking dog, rolling suitcase, closing door, plane, siren, footstep, storm, street noise, train, tramway, truck and voice) and 154 sound background samples (birds, construction site noise, crowd, park, rain, children playing in schoolyard, constant traffic noise, ventilation, wind). The sound class *car passages* comes from recordings of 4 cars made on the Ifsttar's runway on different speeds with multiple gear ratio [37]. The other audio files have been found online (freesound.org) and with the help of the *UrbanSound8k* database [38]. A sound mixture corpus is designed composed of 6 sub-corpus of 25 30 seconds audio files ($N = 25$). Each sub-corpus is characterized by a specific generic sound class : *alert* (car horn, siren), *animals* (barking dog , whistling birds), *climate* (wind, rain), *humans* (crowd noise and voice), *mechanics* (different metallic and construction site noise) and *transportation* (train, tramway and plane). In each file, traffic component is present as sound background and event and is mixed with one or multiple sound classes. The sound classes that are not related to the traffic component are summed up as the *perturbator* (n'existe pas comme terme, disruptive à la place ?) class. To test different scenarios, each audio file is duplicated with the traffic sound level of the entire sound mixture, $L_{p,traffic}$, fixed to a specific level according to the sound level of the *perturbator* class, $L_{p,perturbator}$ following the relation (9).

$$TPR = L_{p,traffic} - L_{p,perturbator} \quad (9)$$

with the *Traffic Perturbator Ratio* $TPR = [-12, -6, 0, 6, 12]$. When $TPR = -12$, the traffic component is then less present than when $TPR = 12$ where it is predominant on the *perturbator* class. The 1 second equivalent sound pressure level, $p_{1s,traffic}$, is too calculated (figure 3). Finally, the number of scenes designed is 750 (6 sub-corpus \times 25 scenes \times 5 TPR values).

3.2 Experiment

The experiment consists in applied NMF on the environmental sound corpus to estimate the sound level of the traffic road. This method is compared to a simple approach which is a low-frequency filter. This method considers, as the traffic component is mostly composed of low-frequencies, that the sound level can be calculated from the band-pass filter. Similarly, for NMF, the low-pass filter is applied on \mathbf{V}_{f_c} and \mathbf{W}_{f_c} with the aim to focus the reconstruction of the signal on the frequencies where the traffic components are. The figure 4 summarizes the different steps of the process depending on the chosen method.

3 calculation steps are set up :

- the dictionary building for NMF
- the estimation of the traffic sound level according the chosen method (NMF or low-pass filter)
- calculation of the error between the exact and the estimated sound levels.

3.2.1 Dictionary building

The dictionary is built from a sound database dedicated specially to this task. It is composed of 53 audio files of car passages. Multiple version of dictionaries are tested. First, for each audio file, its spectrogram is calculated with fixed parameters ($w = 2^{14}$, $noverlap = 75\%$, $nfft = 2^{14}$). Then a temporal rectangular window of 0.5 second is applied without overlapping on the spectrogram. In each window, the root mean square value are calculated. This window allow to get multiples spectrum from an audio file. An example, to illustrate the process, can be found on figure with a large window of 4 seconds.

As the number of spectrum got on all the sound database does not respect the constraint imposed about the dimension of NMF, a K -medoid clustering is applied to reduce the number of spectrum to $K = [25, 50, 100]$. A special case is added where the root mean square of the entire spectrum is applied on all the spectrogram of the audio file. In consequence, one spectra is generated by audio file.

3.2.2 Estimation of the traffic sound level

The traffic sound level is estimated according to the choice of the estimator (low-pass filter or NMF). These methods are applied on the totality of the corpus (*alert* (al), *animals* (an), *humans* (hu), *climate* (cl), *mechanics* (me), *transportation* (tr)), for all the cut-off frequencies

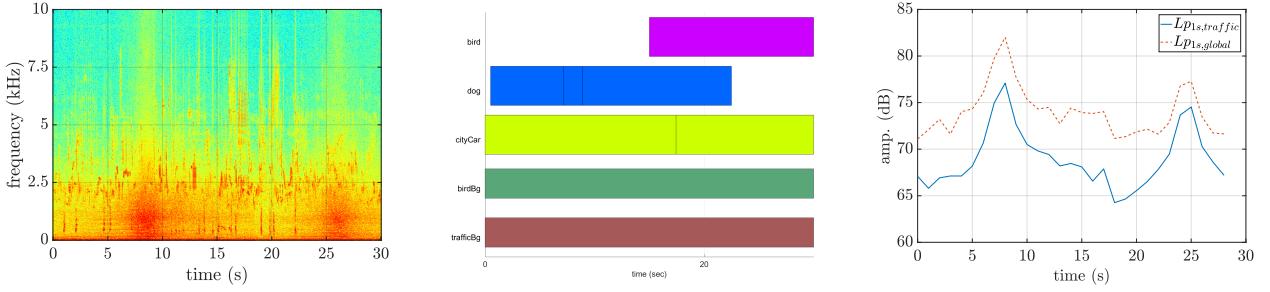


Figure 3: Example of a scene of the *animals* sub corpus. Spectrogram (on left), *Piano Roll* of the different sound classes (on the middle) and 1-s equivalent sound level of the traffic, $L_{p1s, \text{traffic}}$ and of the global sound mixture, $L_{p1s, \text{global}}$ (on right)

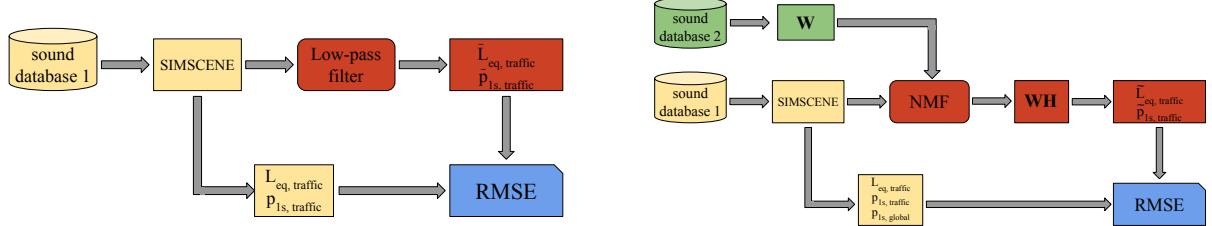


Figure 4: Block diagrams summed up the different step of the precess. On top, for the low-pass filter, on bottom, for NMF.

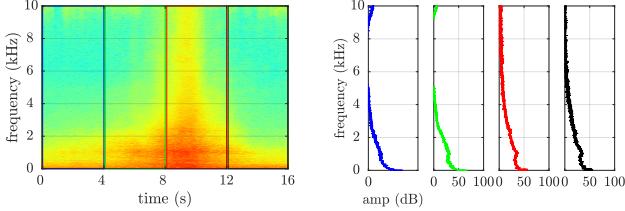


Figure 5: Example of the extraction of the spectrum with a 4 seconds window. On left, the original spectrogram with the temporal window, on the left, the 4 spectrum obtained (dB scale)

f_c ([500 1k 2k 5k 10k 20k] Hz) and for 5 TPR ([-12 -6 0 6 12] dB).

3.2.2.1 Low-pass filter

The first estimator to determine the traffic sound level is a basic low-pass filter which depend only on the cut-off frequencies f_c . As the road traffic is mainly composed of low frequencies (< 5000 Hz), this method considers that all the components located on the pass-band can be considered as road traffic.

3.2.2.2 NMF

The second estimator is NMF with the supervised and the semi-supervised approaches (see part 2). The spec-

rogram \mathbf{V} and the dictionary \mathbf{W} are expressed on two different formats: with a linear frequency scale ($\Delta f \approx 2.8$ Hz) and with third octave bands. Theses two methods enables to compare a fine approach (the linear scale) with a degraded one (the third octave bands) as it reduces the number of frequency bins (27 bands) and allows to reduces the number of bands in the high frequencies where the traffic component is less present. Furthermore, in the case of the linear frequency scale, \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. The β -divergence is reduced to the Kullback-Leibler divergence ($\beta = 1$) and to the Euclidean distance ($\beta = 2$). 400 iterations are performed. As only one sound source is considered in the dictionary, the source separation is performed following the equation (6). Also, 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 made with the original dictionary \mathbf{W} , as

$$\tilde{\mathbf{V}}_{\text{traffic}} = [\mathbf{WH}_{f_c}]_{\text{traffic}} \quad (10)$$

Then for each spectrogram, the estimated equivalent traffic sound level in dB of the entire scene, $\tilde{L}_{p,\text{traffic}}$, is calculated as well as the 1 second equivalent sound pressure level, $\tilde{p}_{1s,\text{traffic}}$.

3.2.3 Metric

The performances of the two estimators to estimate the sound level of the traffic correctly are estimated through the calculation of two metrics based on the Root Mean Square Error (RMSE) :

- the Mean Absolute Error, MAE , which consists in calculate for each combination of factors the absolute difference between the exact and estimated global traffic sound level in dB of the N scenes.

$$MAE = \frac{\sum_{n=1}^N |L_{p,traffic}^n - \tilde{L}_{p,traffic}^n|}{N} \quad (11)$$

This metric expresses the quality of the long-term reconstruction of the signal.

- the global RMSE, $gRMSE$, which consists in calculate for each combination of factors the error between the exact and estimated global traffic sound level in dB of the N scenes.

$$gRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(L_{p,traffic}^n - \tilde{L}_{p,traffic}^n \right)^2} \quad (12)$$

This metric expresses the quality of the long-term reconstruction of the signal.

- The normalized short-term RMSE, $nRMSE$, consists in calculate, for each sound mixture, the error between the 1-s equivalent sound level exact and estimated of the traffic of each file normalized by the 1-s equivalent sound level of the global mixture in the linear scale,

$$nRMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{p_{1s,traffic}^t - \tilde{p}_{1s,traffic}^t}{p_{1s,global}^t} \right)^2} \quad (13)$$

with T is the number of temporal bin in the signal. The linear scale is here more relevant than the dB scale as it is more sensitive to the error on the high sound levels. Then for one combination of factors, the N $nRMSE$ calculated are averaged.

In all, 5940 settings are performed.

4 Results

4.1 Low-pass filter results

In a first time, the low-pass filter estimator is performed on all the scene. The table 1 sums up the mean error on the totality of the 750 scenes to find the cut-off frequency the most efficient on all the cases.

f_c (Hz)	MAE (dB)	$nRMSE$	$gRMSE$ (dB)
500	5.66 ± 6.59	1.48 ± 1.10	5.96 ± 6.70
1000	6.31 ± 7.64	1.52 ± 1.34	6.52 ± 7.73
2000	6.65 ± 8.24	1.57 ± 1.54	6.80 ± 8.32
5000	7.42 ± 8.90	1.82 ± 1.65	7.44 ± 8.89
10000	7.55 ± 9.00	1.89 ± 1.70	7.56 ± 8.99
20000	7.59 ± 9.02	1.89 ± 1.72	7.59 ± 9.02

Table 1: RMSE error for the low pass filter averaged on all the TPR and sub-classes

According the table, the cut-off frequency at 500 Hz is the most efficient on all the TPR and all the sub-classes. The error for a 500 Hz cut-off frequency according to the sub-classes and the TPR is sums up in the figure 6.

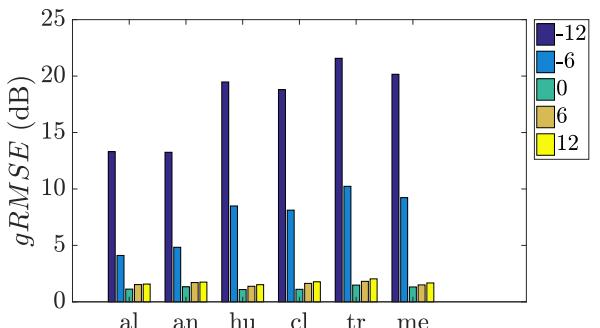


Figure 6: Bar plot for the low-pass filter according to the sub-classes and the TPR at 500 Hz cut-off frequency

The error for all the sub-corpus is important for the low TPR (-12 and -6) due to the confusion of the *perturbator* class as the traffic component whereas it is not. The error is less for the *alert* and *humans* sub-corpus as it is composed of higher frequencies while for the *climate* and *transportation* the error is far more important. For the other TPR, the traffic is more present and become dominating on the *perturbator* class. Then the error is due to the suppression of the traffic energy by the low-pass filter.

4.2 Supervised NMF results

As it is not possible to summarize all the parameter combinations, just the best results according to the domain and β are presented in table 2. The bar plot in figure 7 display the $gRMSE$ error for the best combination.

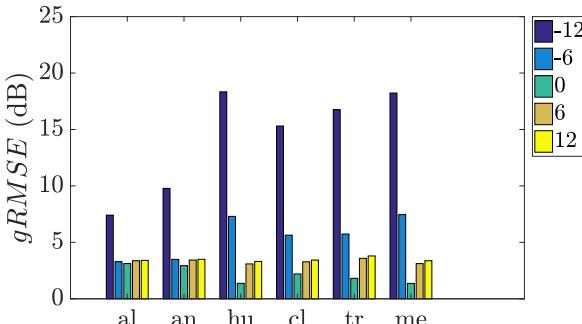


Figure 7: bar plot for the best parameter combination of the supervised NMF ($K = 25$, temporal window = 0.5 s, domain = spectra, $\beta = 2$, $f_c =$)

The best combination for the supervised NMF is the one with described in the *spectra* domain and for the euclidean distance with $K = 25$, the temporal window of 500 ms and $f_c = 500$ Hz. The errors produced by NMF is smaller than the filter approach for all the metrics. The figure 7 allows to see that the supervised NMF improves the estimation of the road traffic for $TPR = [-12 \ 6]$ whereas the error for the higher TPR is superior.

The evolution of the error for 3 sub classes (*alert*, *climate* and *transportation*) and 2 TPR (-6 and 6) are displayed in figures ?? and ??.

We observe that the error if growing for $TPR = -6$, with an minimum in . The phenomena of early stop is present with few iteration . In contrary, for $TPR = 6$, the error is decreasing constantly.

4.3 Semi-supervised NMF results

The error produced for the semi-supervised approach is summarize in table 3.

The best combination is for the third octave domain with $f_c = 20$ kHz, $K = 50$ and a temporal window null. The error, compared to the supervised approach, is lower. Furthermore, the standard deviation is lower too meaning that the semi-supervised approach offer a much more stable error than the previous method. The figure ?? summarize the error for the best scenario expanded for the sub-classes and the TPR.

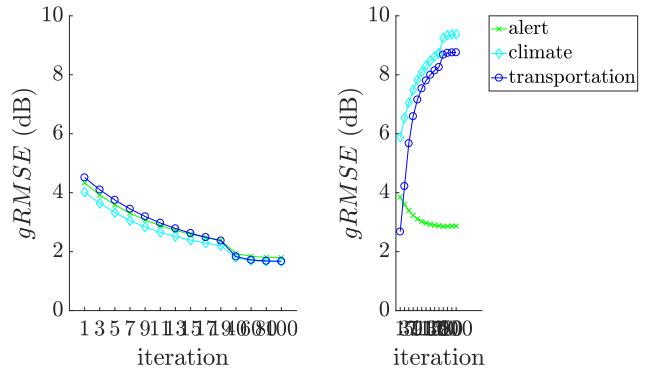


Figure 8: Evolution of the $gRMSE$ for 3 sub-classes for 2 TPR

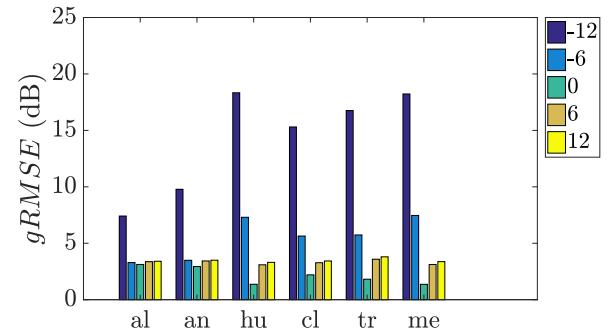


Figure 9: bar plot for the best parameter combination of the supervised NMF ($K = 25$, temporal window = 0.5 s, domain = spectra, $\beta = 2$, $f_c = 20$ kHz)

The mobile part of the dictionary, \mathbf{W}_r , is displayed in two particular sub-classes (*alert*, which include harmonic sound sources, and *climate*, composed of low frequencies) for two TPR (-6 and 6).

For $TPR = -6$, for the *alert* sub-class, \mathbf{W}_r enables to include the harmonic component which are not present in \mathbf{W}_s , the improvement bring by the semi-supervised approach is then remarkable. Whereas, in *climate* sub-class, the *perturbators* sound classes are composed of low-frequencies that are present in \mathbf{W}_s . This class might then be describe with the help of \mathbf{W}_s .

In the opposite, for the high TPR , for the *climate* sub-class, the harmonic content disappears in \mathbf{W}_r and is replace by low frequencies content that might be traffic component. In this approach, as the objective function of to minimize the distance between the spectrogram \mathbf{V} and \mathbf{WH} , the semi-supervised approach is free to include traffic component in \mathbf{W}_r to do it. This freedom deteriorates the reconstruction of the traffic signal. In the supervised approach, as \mathbf{W} is only limited to the traffic spectrum. In consequence, NMF is constrained to use

temporal							
K	window (ms)	f_c (Hz)	domain	β	MAE (dB)	nRMSE	gRMSE (dB)
25	500	spectra	2000	1	6.18 ±7.49	1.48 ±1.20	6.36 ±7.55
25	500	spectra	500	2	5.32 ±6.28	1.40 ±0.93	5.66 ±6.38
25	0	thirdOctave	20000	1	6.82 ±8.16	1.60 ± 1.34	6.88 ±8.15
25	500	thirdOctave	20000	2	6.26 ±7.69	1.36 ±1.19	6.36 ±7.73

Table 2: RMSE error for supervised NMF averaged on all the TPR and sub-classes

temporal							
K	window (ms)	f_c (Hz)	domain	β	MAE (dB)	nRMSE	gRMSE (dB)
			spectra	1			
100	0	20000	spectra	2	4.02 ±3.48	1.14 ±0.54	4.41 ±3.69
50	0	20000	third octave	1	3.88 ±2.33	1.29 ± 0.69	4.43 ±2.71
25	0	20000	third octave	2	4.29 ±3.58	1.21 ± 0.61	4.74 ±3.84

Table 3: RMSE error for semi-supervised NMF averaged on all the TPR and sub-classes

them to approximate \mathbf{V} , which improve the estimation of the traffic sound level for the high TPR.

PLOT COST PLOT RMSE PLOT Lp 1s pour 2 ambiance 2 TPR

5 Conclusion

References

- [1] World Health Organization. Burden of disease from environmental noise. Quantification of healthy life years lost in Europe. <http://www.euro.who.int/en/home/copyright-notice>. visité le 24/08/2017.
- [2] Commission in the Conciliation Committee on the directive relating to the assessment and management of environmental noise. Directive 2002/49/ec relating to the assessment and management of environmental noise. visité le 24/08/2017.
- [3] H. Van Leeuwen and S. Van Banda. Noise mapping - State of the art - Is it just as simple as it looks? *EuroNoise*, 2015.
- [4] O. Leroy, B. Gauvreau, F. Junker, E. De Rocquigny, and M. Berengier. Uncertainty assessment for outdoor sound propagation. In *20th International Congress on Acoustics, ICA 2010*, page 7p, France, August 2010.
- [5] N. Garg and S. Maji. A Critical Review of Principal Traffic Noise Models: Strategies and Implications. *Environmental Impact Assessment Review*, 46, April 2014.
- [6] E. A. King, E. Murphy, and H. J. Rice. Implementation of the EU environmental noise directive: lessons from the first phase of strategic noise mapping and action planning in Ireland. *Journal of Environmental Management*, 92(3):756–764, March 2011.
- [7] W. Wei, T. Van Renterghem, B. De Coensel, and D. Botteldooren. Dynamic noise mapping: A map-based interpolation between noise measurements with high temporal resolution. *Applied Acoustics*, Complete(101):127–140, 2016.
- [8] P. Mioduszewski, J. A. Ejsmont, J. Grabowski, and D. Karpinski. Noise map validation by continuous noise monitoring. *Applied Acoustics*, 72(8):582–589, july 2011.
- [9] Christophe Mietlicki, Fanny Mietlicki, and Matthieu Sineau. An innovative approach for long-term environmental noise measurement: Rumeur network. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 2012, pages 7119–7130. Institute of Noise Control Engineering, 2012.
- [10] A. Can, T. Van Renterghem, and D. Botteldooren. Exploring the use of mobile sensors for noise and black carbon measurements in an urban environment. In Société Française d'Acoustique, editor, *Acoustics 2012*, Nantes, France, April 2012.
- [11] D Manvell, L Ballarin Marcos, H Stapelfeldt, and R Sanz. Sadmam-combining measurements and calculations to map noise in madrid. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 2004, pages 1998–2005. Institute of Noise Control Engineering, 2004.
- [12] A. Can, L. Dekoninck, and D. Botteldooren. Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation approaches. *Applied acoustics*, 83:32–39, 2014.
- [13] S. Xavier, S. J. Claudi, A. Francesc, B. Patrizia, and al. DYNAMAP – Development of low cost sensors networks for real time noise mapping. *Noise Mapping*, 3(1), May 2016.
- [14] L. Bellucci, P. and Peruzzi and G. Zambon. LIFE DYNAMAP project: The case study of Rome. *Applied Acoustics*, Part B(117):193–206, 2017.
- [15] T. Heittola, A. Mesaros, T. Virtanen, and A. Eronen. Sound event detection in multisource environments using source separation. In *in Workshop on Machine Listening in Multisource Environments, CHiME2011*, 2011.
- [16] B. Defreville, F. Pachet, C. Rosin, and P. Roy. Automatic Recognition of Urban Sound Sources. Audio Engineering Society, 2006.
- [17] M. Cowling and R. Sitte. Comparison of techniques for environmental sound recognition. *Pattern Recognition Letters*, 24(15):2895–2907, November 2003.
- [18] G. Shen, Q. Nguyen, and J. Choi. An Environmental Sound Source Classification System Based on Mel-Frequency Cepstral Coefficients and Gaussian Mixture Models. *IFAC Proceedings Volumes*, 45(6):1802–1807, May 2012.
- [19] F. Beritelli and R. Grasso. A pattern recognition system for environmental sound classification based on MFCCs and neural networks. In *2008 2nd International Conference on Signal Processing and Communication Systems*, pages 1–4, December 2008.

- [20] L. Couvreur and M. Laniray. Automatic Noise Recognition in Urban Environments Based on Artificial Neural Networks and Hidden Markov Models. In *The 33 rd International Congress and Exposition on Noise Control Engineering*, Prague, August 2004.
- [21] P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, April 1994.
- [22] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.
- [23] P. Smaragdis and J.C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on*, pages 177–180, October 2003.
- [24] K. W. Wilson, B. Raj, P. Smaragdis, and A. Divakaran. Speech denoising using nonnegative matrix factorization with priors. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4029–4032, March 2008.
- [25] A. Mesaros, T. Heittola, O. Dikmen, and T. Virtanen. Sound event detection in real life recordings using coupled matrix factorization of spectral representations and class activity annotations. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 151–155, April 2015.
- [26] B. Wang and M. D. Plumley. Musical audio stream separation by non-negative matrix factorization. *Proc. DMRN summer conf*, pages 23–24, 2005.
- [27] A. Kumar, B. Elizalde, and B. Raj. Audio Content based Geotagging in Multimedia. *arXiv:1606.02816 [cs]*, June 2016. arXiv: 1606.02816.
- [28] N. Bertin. *Les factorisations en matrices non-négatives : approches contraintes et probabilistes, application à la transcription automatique de musique polyphonique*. Paris, Télécom ParisTech, January 2009.
- [29] C. Févotte, N. Bertin, and J. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: with application to music analysis. *Neural Computation*, 21(3):793–830, March 2009.
- [30] T. Virtanen. Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3):1066–1074, March 2007.
- [31] A. Cichocki and R. Zdunek. Regularized Alternating Least Squares Algorithms for Non-negative Matrix/Tensor Factorization. In *Advances in Neural Networks – ISNN 2007*, Lecture Notes in Computer Science, pages 793–802. Springer, Berlin, Heidelberg, June 2007.
- [32] C. J. Lin. Projected Gradient Methods for Nonnegative Matrix Factorization. *Neural Computation*, 19(10):2756–2779, October 2007.
- [33] D. Lee and H. Seung. Algorithms for Non-negative Matrix Factorization. In *In NIPS*, pages 556–562. MIT Press, 2000.
- [34] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the β -divergence. *Neural Computation*, 23(9):2421–2456, 2011.
- [35] H. Lee, J. Yoo, and S. Choi. Semi-Supervised Nonnegative Matrix Factorization. *IEEE Signal Processing Letters*, 17(1):4–7, January 2010.
- [36] M. Rossignol, G. Lafay, M. Lagrange, and N. Misdariis. Sim-Scene: a web-based acoustic scenes simulator. In *1st Web Audio Conference (WAC)*, 2015.
- [37] J.-R. Gloaguen, A. Can, M. Lagrange, and J.-F. Petiot. Creation of a corpus of realistic urban sound scenes with controlled acoustic properties. *The Journal of the Acoustical Society of America*, 141(5):4044–4044, May 2017.
- [38] J. Salamon, C. Jacoby, and J. P. Bello. A dataset and taxonomy for urban sound research. In *22st ACM International Conference on Multimedia (ACM-MM'14)*, Orlando, FL, USA, Nov. 2014.