

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

The road traffic noise is one of the main annoyance perceived by the urban dweller. This boredom is not harmless and becomes a public problem because too long exposure to high sound levels can generate health problems [1]. Accordingly, European directive 2002/EC/49 [2] has been developed to determine the impact of this noise on the population. For cities with a population of more than 100,000, it requires the production of traffic noise maps, which consist in calculating the sound level of traffic noise in the streets. These maps come from a simulation process resulting from the estimation of the traffic density on the main roads and using the acoustic propagation models. The maps express the average sound levels over 24 hours, L_{DEN} (*Day-Evening-Night*), and for the night over a period of 8h, L_N (*Night*). These maps make it possible to estimate the number of city dwellers impacted by the high noise levels and to draw up action plans to reduce them. However, these maps have been criticized because a lot of uncertainty can be generated because of the errors produced by the numerical tools [3], by the different calculation methodologies used [4] or even by the calculation procedure of the number of inhabitants exposed to noise [5]. In addition, these maps offer restrictive information with only two calculated global sound levels that need to be updated every five years while road traffic varies throughout the day. Since the creation of the traffic noise maps is a long process, it is inconceivable to update them every hour, the use of measurements could facilitate their updating or even the generation of dynamic maps [6]. These measurements can be performed on fixed points spreading all over the cities [7] [8], which allow to get the long-term evolution of the traffic noise levels, or can be performed with mobile points [9] [10], which have the advantage of being able to cover a larger area with fewer sensors. Also, the clustering between mobile and fixed measure-

ments have been studied in [11]. 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 [12] project. It focuses on sensor installations on specific roads in all cities of Milan and Rome [13]. The principle is, since the measurements made on these roads, to correct sound emissions on similar roads that are supposed to follow the same evolution than the measured ones. In a similar way, but reduced to few neighborhoods, the CENSE project¹ [] plains to combine *in situ* observations, from a sensor network, and numerical data, from noise modelling, by using data assimilation techniques. In addition to improving the traffic noise maps, the measurements will be used to lead perceptual studies on the urban soundscapes and to develop sound recognition tools. 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 [7]. 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. Our study proposed to applied a source separation tool, Non-Negative Matrix Factorization, on simulated sound mixtures to succeed on the estimation of the traffic sound level. Part 2 sums up on the different approaches in signal processing to isolate the traffic component of a monaural signal. Part 3 details the technical aspect of the chosen method. Part 4 is focused on the experimental protocol set up. Then part 5 reveals the results obtained during the parametric study.

¹<http://cense.ifsttar.fr/>

2 State of the art in the detection and separation tasks for environmental sounds

Many references have focused on the classification of the sound environments [14], on the recognition [15] or one the detection of the environmental sound events [16]. This category of sound is not specific to the urban environment but includes all sound sources that are not related to music or speech. The algorithms are first learned on a known database to evaluate the performance of the algorithm then is applied to a second database simulating its application in real cases. To achieve these tasks, many approaches have the same generic pattern: a first extraction phase followed by a classification phase. In the first phase, the parameters of the characteristic are applied to an audio file to describe it, from the simplest, such as the sound level, the spectrum Gravity Centrums, the harmonicity to the most complex (Mel-Frequency Cepstral Coefficient, Perceptual Linear Prediction). In the second step, the characteristics are classified according to their evolution making it possible to distinguish the different categories of sounds. Common classification techniques are Hidden Markov Models, Artificial Neural Network, Gaussian Mixture Models and Support Vector Machines. Cowling and Sitte [17] summarize multiples of these extraction features and classification techniques that are used in [18], [19], [20]. Even if each of them tested their algorithm on their own isolated sound database, specific database have been proposed to test and compare the performances of the algorithms as the DCASE database [21] or the TUT database [22]. Currently, the interest is focused on the use of the Neural Network in the case of the detection of acoustics events [23] [24].

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. Consequently, the only detection of the traffic component does not make it possible to determine precisely its noise level. In consequence, to effective on all the different sound environment, 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 [25]. 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 environ-

ments 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. In consequence, the use of Non-negative Matrix Factorization (NMF) is chosen has it allows to take into account easily the overlapping between the different sound sources and the time-frequency evolution of the sound sources by its spectrogram representation. This method has been widely used in the sound domain, [26] [27] [28], and has already been employed for the source separation task dedicated for speech [27] and musical content [29] on monaural signals.

3 Non-negative Matrix Factorization

3.1 Supervised NMF

Non-negative Matrix Factorization (NMF) is an approximation method introduced by Lee and Seung, [30], which estimates the spectrogram of an audio file, \mathbf{V} , with $F \times N$ dimension and which belong in \mathbb{R}^+ as :

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

where $\mathbf{W}_{F \times K} \in \mathbb{R}^+$ is the *basis* matrix composed of audio spectrum and $\mathbf{H}_{K \times N} \in \mathbb{R}^+$ is the *activation* matrix which sum up the temporal evolution of each element of \mathbf{W} (fig. 1).

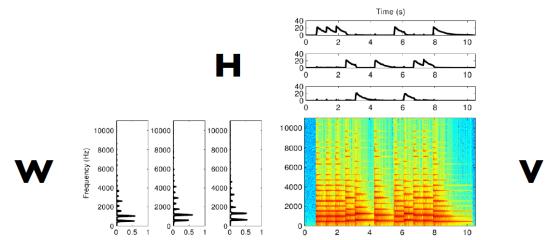


Figure 1: Example of a simple NMF for a musical content [31]

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

$$\min D(\mathbf{V} || \tilde{\mathbf{V}}) \quad \text{w.r.t. } \mathbf{H} \geq 0, \mathbf{W} \geq 0 \quad (2)$$

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 [32] in which the well known Euclidean distance (eq. 4a) and the Kullback-Leibler divergence (eq. 4b) belong

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

The prior knowledges on the content can be adjust with the add of constraints (like the smoothness or the sparseness criteria [33]) in the objective function 2.

NMF can be employed in an unsupervised case where nor \mathbf{W} or \mathbf{H} is known. The *dictionary* is learned first during the training phase with different sound samples, the *basis* got are not retained. \mathbf{W} is afterwards used on the test phase where a new *basis* is got. For the supervised case, \mathbf{W} is already known. In the urban environments, a lot of different sound sources present in the cities are known and their spectrum can be obtained. Multiple methods exist to solve the minimization problem (2) (Alternating Least Square Method [34], Projected Gradient [35] ...). Here, the multiplicative update is the chosen method [36] as it assures to have a non-negative results where the convergence of the results have been proved [37]. The minimization problem (2) is solved iteratively. As it is the supervised method that is consider first, only the \mathbf{H} is updated. 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.

The source separation is made after I iterations by generating a filtering mask, M_s ,

$$\mathbf{M}_s = \frac{[\mathbf{WH}]_{traffic}}{\sum_{k=1}^K [\mathbf{WH}]}, \quad (6)$$

which allows to determine the spectrogram of the traffic component, $\tilde{\mathbf{V}}_{traffic}$

$$\tilde{\mathbf{V}}_{traffic} = \mathbf{M}_s \cdot \mathbf{V}. \quad (7)$$

For the environmental sounds, the method has been used for geo-localisation and classification of the sound environment like in [38] 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. They used the MFCC features to composed the dictionary by adding a time variant gain features (a calculation of variance and a ratio to nearly zero values on the *basis* vectors). 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.

3.2 NMF Semi-supervised

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 [39]. This method consists in composing the dictionary $\mathbf{W}_{F \times K+J}$ with a fixed part $\mathbf{W}_{sF \times K}$ (composed here of traffic audio spectrum) and with a mobile part, $\mathbf{W}_{rF \times J}$ 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

$$\mathbf{V} \approx \mathbf{W}_s \mathbf{H}_s + \mathbf{W}_r \mathbf{H}_r \quad (8)$$

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 (9):

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

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

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

This approach has been used TO COMPLETE.

4 Experimental protocol

NMF are not tested on urban sound recordings: the traffic component is here unknown. Therefore, the estima-

tion of the traffic sound level cannot be compare. In consequence, NMF is applied on simulated sound mixtures as it offers a controlled framework where the contribution of each sound source (and especially the traffic) is known and offers an exact solution to compare with the estimated one.

4.1 Environmental sound scene corpus

A first corpus is design with the *simScene* software². *simScene* [40] 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 sum up the time presence of all the event.

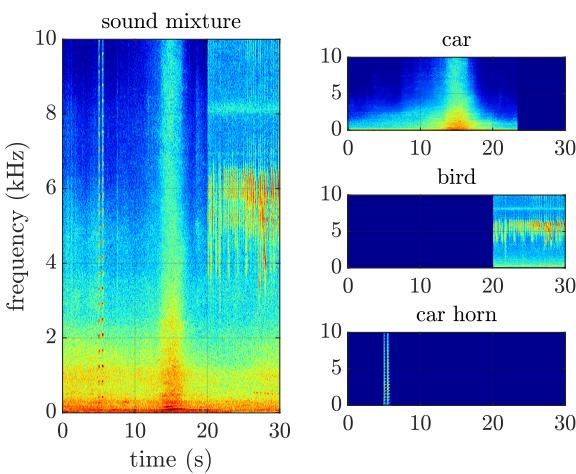


Figure 2: Example of a sound mixture composed of 3 sound classes (car, bird, car horn)

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

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*, *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 [41]. The other audio files are sampled at 44.1 kHz and have been found online (freesound.org) and with the help of the *UrbanSound8k* database [42]. 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 ?) 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 (10).

$$TPR = L_{p,traffic} - L_{p,perturbator} \quad (10)$$

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).

4.2 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, the low-pass filter is applied with NMF with the aim to focus the reconstruction of the signal on the frequencies where the traffic components are. The figure 4 sum

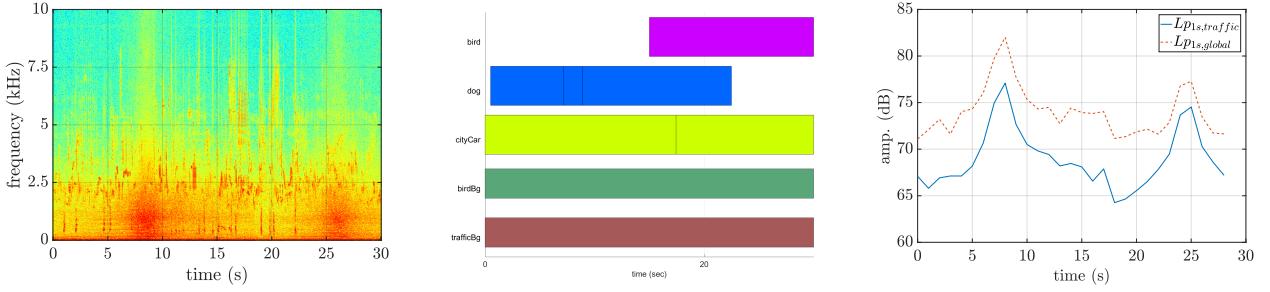


Figure 3: Example of a scene of the *animals* sub corpus. Spectrogram (on left), *Piano Roll* of the different sound classes (on the middle) and 1-s equivalent sound level of the traffic, $L_{p1s, traffic}$ and of the global sound mixture, $L_{p1s, global}$ (on right)

up the different step 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.

4.2.1 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 spectrum is applied on all the spectrogram. In consequence, the audio file is summed up in one spectra.

4.2.2 Estimation of the traffic sound level

4.2.3 Low-pass filter

The first estimator to determine the traffic sound level is a basic low-pass filter with the cut-off frequencies

$f_c = [500, 1k, 2k, 5k, 10k, 20k]$ Hz. 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.

4.2.4 NMF

The second estimator is NMF with the supervised and the semi-supervised approaches (see part 3). The spectrogram \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. The source separation is performed following equations (6) and (7) according two ways: based on the filtered elements (\mathbf{V}_{f_c} and \mathbf{W}_{f_c}) or on the entire spectrogram and dictionary (\mathbf{V} and \mathbf{W}). Then for each spectrogram, the estimated equivalent traffic sound level in dB of the entire scene, $\tilde{L}_{p, traffic}$, is calculated as well as the 1 second equivalent sound pressure level, $\tilde{p}_{1s, traffic}$.

4.2.5 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 global RMSE, $gRMSE$, which consists in calculate for each combination of factors the error be-

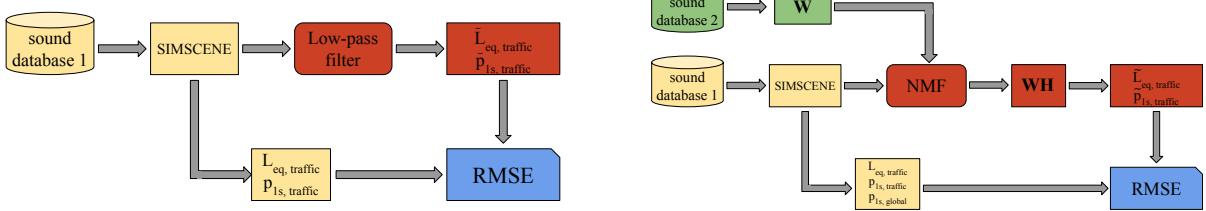


Figure 4: Block diagrams summed up the different step of the process. On top, for the low-pass filter, on bottom, for NMF.

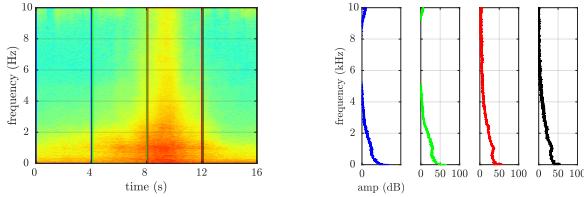


Figure 5: Example of the extraction of the spectrum with a 4 seconds window. On left, the original spectrogram with the temporal window, on the left, the 4 spectrum obtained (dB scale)

tween the exact and estimated global traffic sound level in dB of the N scenes.

$$gRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(L_{p, traffic}^n - \tilde{L}_{p, traffic}^n \right)^2} \quad (11)$$

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,

$$nRMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{p_{1s, traffic}^t - \tilde{p}_{1s, traffic}^t}{p_{1s, global}^t} \right)^2} \quad (12)$$

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.

The calculation are performed in 3 steps. The first set up the dictionary with two modalities : K value (25 50 100) and description of the audio files (in one fft or with a temporal window of 0.5 second). The second step include modalities : the sub-corpus (*alert* (al), *animals* (an), *humans* (hu), *climate* (cl), *mechanics* (me), *transportation* (tr)), the estimator (low-pass filter, NMF), the cut-off frequency f_c ([500 1k 2k 5k 10k 20k] Hz), the *TPR* ([-12 -6 0 6 12] dB), $\beta = [1 2]$, the approach NMF (supervised or semi-supervised) and the frequency scale (linear or third octave). In all, 5940 settings are performed.

5 Results

In a first time, the low-pass filter estimator is performed on all the scene. The table 2 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)	$nRMSE$	$gRMSE$ (dB)
500	1.48 ± 1.10	5.96 ± 6.70
1k	1.52 ± 1.34	6.52 ± 7.73
2k	1.57 ± 1.54	6.80 ± 8.32
5k	1.82 ± 1.65	7.44 ± 8.89
10k	1.89 ± 1.70	7.56 ± 8.99
20k	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.

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

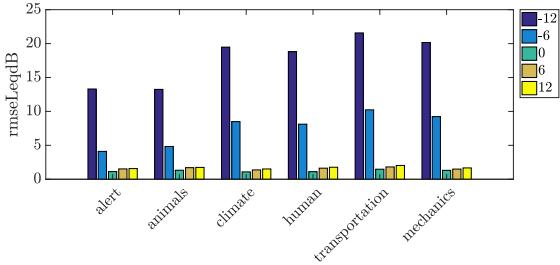


Figure 6: Bar plot for the low-pass filter according to the sub-classes and the TPR at 500 Hz cut-off frequency

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.

The first results averaged on all the TPR and the sub-classes for the supervised NMF is sums up on table ??, the frequency domain is, for now, limited to *spectra*

Table 2: RMSE error for the low pass filter averaged on all the TPR and sub-classes

Résultat global pour le NMF supervisée en tableau puis en bar pour chaque ambiance et chaque TPR Tracer l'évolution 1s du signal sonore pour 2 ambiances et 2 TPR résultat semi-supervisée beta = 2, = 1

6 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] N. Garg and S. Maji. A Critical Review of Principal Traffic Noise Models: Strategies and Implications. *Environmental Impact Assessment Review*, 46, April 2014.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] L. Bellucci, P. and Peruzzi and G. Zambon. LIFE DYNAMAP project: The case study of Rome. *Applied Acoustics*, Part B(117):193–206, 2017.
- [14] Silvia Allegro, Michael Büchler, and Stefan Launer. Automatic sound classification inspired by auditory scene analysis. In *in Consistent and Reliable Acoustic Cues for Sound Analysis*, 2001.
- [15] B. Defreville, F. Pachet, C. Rosin, and P. Roy. Automatic Recognition of Urban Sound Sources. Audio Engineering Society, 2006.
- [16] 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.
- [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] D. Giannoulis, D. Stowell, E. Benetos, M. Rossignol, M. Lagrange, and M. D. Plumley. A database and challenge for acoustic scene classification and event detection. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5, September 2013.

- [22] A. Mesaros, T. Heittola, and T. Virtanen. TUT database for acoustic scene classification and sound event detection. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 1128–1132, August 2016.
- [23] N. Genaro, A. Torija, A. Ramos-Ridao, I. Requena, D. P. Ruiz, and M. Zamorano. A neural network based model for urban noise prediction. *The Journal of the Acoustical Society of America*, 128(4):1738–1746, October 2010.
- [24] G. Parascandolo, H. Huttunen, and T. Virtanen. Recurrent neural networks for polyphonic sound event detection in real life recordings. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6440–6444, March 2016.
- [25] P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, April 1994.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] B. Wang and M. D. Plumley. Musical audio stream separation by non-negative matrix factorization. *Proc. DMRN summer conf*, pages 23–24, 2005.
- [30] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] C. J. Lin. Projected Gradient Methods for Nonnegative Matrix Factorization. *Neural Computation*, 19(10):2756–2779, October 2007.
- [36] D. Lee and H. Seung. Algorithms for Non-negative Matrix Factorization. In *In NIPS*, pages 556–562. MIT Press, 2000.
- [37] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the β -divergence. *Neural Computation*, 23(9):2421–2456, 2011.
- [38] Anurag Kumar, Benjamin Elizalde, and Bhiksha Raj. Audio Content based Geotagging in Multimedia. *arXiv:1606.02816 [cs]*, June 2016. arXiv: 1606.02816.
- [39] H. Lee, J. Yoo, and S. Choi. Semi-Supervised Nonnegative Matrix Factorization. *IEEE Signal Processing Letters*, 17(1):4–7, January 2010.
- [40] M. Rossignol, G. Lafay, M. Lagrange, and N. Misdariis. Sim-Scene: a web-based acoustic scenes simulator. In *1st Web Audio Conference (WAC)*, 2015.
- [41] 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.
- [42] 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.