ESTIMATING TRAFFIC NOISE LEVELS USING ACOUSTIC MONITORING: A PRELIMINARY STUDY

Jean-Rémy Gloaguen, Arnaud Can

Ifsttar - LAE
Route de Bouaye - CS4
44344, Bouguenais, FR
jean-remy.gloaguen@ifsttar.fr

Mathieu Lagrange, Jean-François Petiot

IRCCyN, UMR CNRS 6597
École Centrale de Nantes
1 rue de la Noe
44321, Nantes, FR

ABSTRACT

In this paper, Non-negative Matrix Factorization is applied for isolating the contribution of road traffic from acoustic measurements in urban sound mixtures. This method is tested on simulated scenes to enable a better control of the presence of different sound sources. The presented first results show the potential of the method.

***Index Terms*—** Non-negative Matrix Factorization, road traffic noise mapping, urban measurements

1. INTRODUCTION

Noise in cities is one of the main sources of annoyance essentially caused by road, air an rail traffic. To know better the noise spatial distribution, the number of people impacted and to preserve quiet areas, the European Directive 2002/49/EC [1] requires that cities over 250 000 inhabitants produce noise maps for road, air and rail traffic. Road traffic noise maps are produced based on a census of the traffic volumes and mean speeds along the main roads which allows estimating their acoustic emission. Assuming knowledge of the city topography, the acoustic propagation within the streets is then calculated. In addition, noise observatories are being deployed in some agglomerations. They aim to facilitate both the mandatory five year update of maps and the validation of the simulated noise maps. Combining classical noise maps with measures would also be a promising approach to go towards more accurate noise maps [2] [3].

However, to achieve those important goals, we have to isolate the road traffic contribution from measurements of the sound mixture that contain many other sources. Indeed, urban sound environments are composed of a large variety of sounds as traffic noise, horn, bird whistles, foot steps, construction sound noise, voices . . . Each has its own spectral properties and temporal structure and may overlap with the other sound sources. Without distinction between these, the traffic noise level estimation is calculated with some sources which do not belong to a traffic car class and is then overestimated. In this study car horn and braking noise are not considered as a traffic car noise as they are not taken into account in traffic noise map.

Different techniques exist and were shown relevant for recognition or detection in urban environment [4] [5] but they do not take into account the overlap between the sources. Methods for source separation, such as Computational Auditory Scene Analysis [6] or Independent Component Analysis [7] are efficient but are, to the best of our understanding, not suitable for urban applications. Indeed, the first one has been primarily developed to simulate the human auditory system whereas the second one requires as many sensors as sound sources, which is unrealistic in a urban context.

Non-negative Matrix Factorization (NMF) [8] has the advantage to deal with the overlap between the sound sources. It has been used for many applications in audio domain such as polyphonic music transcription [9] or for source separation of musical content [10]. Thus the NMF seems to be a suitable method for the isolation of the contribution of road traffic from measurements. We propose to apply an NMF scheme on a corpus of urban sound mixtures to validate its ability to estimate the noise level of road traffic. The specificity of urban sound environments, and the fact that the method has, to the best of our knowledge, never been used in this setting, stands as a challenge and requires specific adaptations.

In this paper, we present the implementation of our experimental plan and some first results. Section 2 exposes the structure of the proposed system based on the NMF framework. Then the experimental protocol is presented in Section 3 and preliminary results are discussed in section 4.

2. PROPOSED APPROACH

The aim of the system is to estimate the level of some predefined sources in the mixture coming from measurements of the urban scene. As can be seen in Figure 1, the signal is

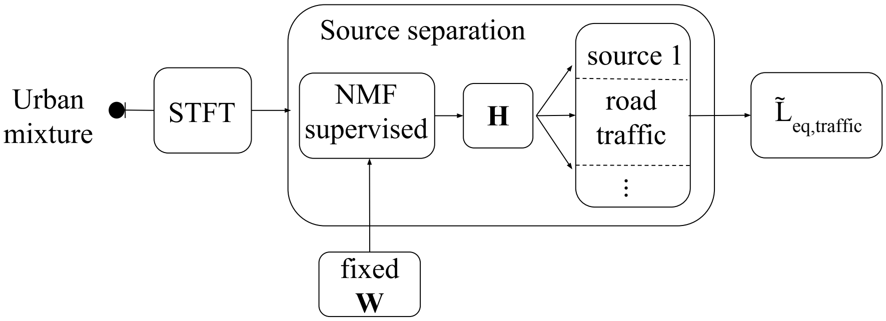

Figure 1: Block diagram of the proposed method

first mapped to a time-frequency plane using the Short Time Fourier Transform. Using the NMF framework, the contribution of the road traffic is isolated and its level, $\tilde{L}_{eq,tr}$, is estimated.

2.1. Non-Negative Matrix Factorization

Non-negative Matrix Factorization is a dimension-reduction technique expressed by

$$\mathbf{V} \approx \tilde{\mathbf{V}} = \mathbf{WH} \tag{1}$$

where $\mathbf{V}_{F\times N}$, is the power spectrogram of an audio, $\tilde{\mathbf{V}}$ is the approximate power spectrogram determined by the NMF, $\mathbf{W}_{F\times K}$, is the basis matrix (called dictionary), in our case, representing a set of sound spectra usually found in urban areas. $\mathbf{H}_{K\times N}$ is the feature matrix standing for the temporal variation of each spectrum. All these elements are constrained to be positive leading to additive combinations only. The approximation (1) is determined by a minimization problem

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

$D(\mathbf{V}||\mathbf{WH})$ is called cost function, a dissimilarity measure usually belonging to the β-divergence for the NMF. 3 popular expressions are compared in this study namely the Euclidean distance ($\beta = 2$),

$$D_{EUC}(\mathbf{V}||\mathbf{WH}) = ||\mathbf{V} - \mathbf{WH}||, \tag{3}$$

the Kullback-Leibler divergence, ($\beta = 1$),

$$D_{KL}(V||WH) = \mathbf{V}\log\frac{\mathbf{V}}{\mathbf{WH}} - \mathbf{V} + \mathbf{WH}, \tag{4}$$

and the Itakura-Sato divergence, ($\beta = 0$),

$$D_{IS}(V||WH) = \frac{\mathbf{V}}{\mathbf{WH}} - \log\frac{\mathbf{V}}{\mathbf{WH}} - 1. \tag{5}$$

Note that decimal β values between 0 and 2 will be investigated in a further study. Here, the supervised NMF is considered where $\mathbf{W}$ is fixed and only $\mathbf{H}$ is updated iteratively. The choosen algorithm is the maximisation-minimisation algorithm proposed by Févotte and Idier [11].

$$\mathbf{H}^{k+1} \longleftarrow \mathbf{H}^{k}.\left(\frac{\mathbf{W}^{T}\left[(\mathbf{WH}^{k})^{\beta-2}.\mathbf{V}\right]}{\mathbf{W}^{T}\left[\mathbf{WH}^{k}\right]^{\beta-1}}\right)^{\gamma(\beta)} \tag{6}$$

where $\gamma(\beta) = 1$ for $\beta \in [1\ 2]$ and $\gamma(\beta) = \frac{1}{2}$ for $\beta = 0$.

2.2. Method

Our approach consists in considering an audio signal recorded in an urban context, sampled at $44,1$ kHz and expressed in the time-frequency plan using a Short Time Fourier Transform. The size of the Hanning window is 5000 points with an overlap of 50 % and $NFFT = 4096$ points. The temporal resolution chosen is, for the moment, very low ($\Delta t \approx 0,05$ s).

The supervised NMF is then performed with the spectrogram $\mathbf{V}$ in the input, a fixed dictionary $\mathbf{W}$ and $\tilde{\mathbf{V}}$ in the output. Currently, $\mathbf{H}$ is updated for a number of iterations fixed at 100. When the iteration is over, it is possible to estimate the level of the elements of interest. In the case of road traffic, $\tilde{\mathbf{V}}_{tr} = [\mathbf{WH}]_{tr}$ which allows to calculate the sound pressure level $\tilde{L}_p$ for each temporal frame

$$\tilde{L}_{p,tr,n} = 20\log\frac{\sum_f \tilde{\mathbf{v}}_{\mathbf{n},\mathbf{tr}}}{p_0} \tag{7}$$

with $\tilde{\mathbf{v}}_{\mathbf{n},\mathbf{tr}}$, the n-th temporal frame of the matrix $\tilde{\mathbf{V}}_{tr}$ and $p_0 = 2\times 10^{-5}$ Pa, the reference sound pressure. The equivalent traffic sound level estimated, $\tilde{L}_{eq,tr}$, is then determined by

$$\tilde{L}_{eq,tr} = \frac{1}{T}\sum_{n} 10\log\left(10^{\tilde{L}_{p,tr,n}/10}\right) \tag{8}$$

where T is the duration of $\mathbf{V}$.

3. EXPERIMENT

To evaluate the ability of the NMF framework to estimate the road traffic level, we consider simulated sounds mixtures where the actual level of contribution of the traffic is known. This solution ensures controlling the road traffic level, $L_{eq,tr}$, relatively to the other sources in comparison to real recordings where it would not be correctly determined. Furthermore, working on simulated sound mixtures will create a controlled framework where the time of presence of each source is exactly known. Thus allows the production of specific sound environments (animated streets, parks ...).

The mixtures are simulated with *simScene* software developed by Mathias Rossignol and Gregoire Lafay [12][1] which synthesizes sound mixtures from a sound database of isoled sound events. This tool can control multiple parameters as the event/background ratio, the sample duration, the time between samples ... Each of these parameters is coupled with a standard deviation to bring some variability between the scenes produced. In the output, an audio file of each sound class is created that allows us to compute the specific contribution of each class present in the scene. The sound database we use is composed of sound samples provided with the software and completed by others sounds found online[2]. The scenes are built with the first half of the database, the second half being considered as the dictionary $\mathbf{W}$. For tests of feasibility, the first constructed scenes are simple but more realistic scenes fully consistent will be soon produced.
For this preliminary study, 20 scenes are created with a duration of 15 s. Each one is composed of 3 classes of sounds that can typically be heard in urban areas: *car*, *bird* and *car horn* and a noise background (voice hubbub). Currently, our dataset for creating these scenes is composed of 30 audio samples for the *car* class and 3 samples for *bird* and *horn* classes. The dictionary W is then composed of the same number of samples but extracted form the second half of our database. The aim of this preliminary study is to see the influence of some parameters of the NMF (such as the divergence calculation or the number of iteration) on the quality of the traffic noise levels estimation. The NMF is performed on each scene i and $L^i_{eq,tr}$ is compared with the computational level $\tilde{L^i}_{eq,tr}$ to evaluate the performance of the method by computing the error,

$$RSME = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(L^i_{eq,tr} - \tilde{L}^i_{eq,tr})^2} \tag{9}$$

where N, the number of scenes created.

4. RESULTS

Figure 2 presents the spectrograms obtained by *simScene* (on left) and by the NMF (on right) for one scene with the Euclidean distance (3) after 100 updates of $\mathbf{H}$. We can observe the bird on the frequency range $[3000 - 6000]$ Hz, the horn is characterized by its harmonic content whereas the car is mainly composed of low frequencies with a slower temporal evolution. From each sound mixture, comparison between $L_{p,n}$ and $\tilde{L}_{p,n}$ for Euclidean distance (EUC), Kullback-Leibler (K-L) and Itakura-Sato (I-S) divergences can be made (Figure 3 for the scene presented in Figure 2).

For this scene, in the time interval $[1.5 - 4.5]\ s$, there is no traffic, the actual sound level is then zero. But we can see in Figure 3 that the class *car* contributes to describe the noise background level. This result is the consequence of the minimization problem (2) where this sound class is activated to reduce the cost function even though there is no traffic. Nevertheless, the noise background is low enough in comparison with the other class sounds to not distort the estimations.

Let us now consider the error RMSE with respect to the number of iterations of the NMF computed on the N scenes for the three β-divergences on Figure 4. The error between $L_{eq,tr}$ and the equivalent sound pressure of the global mixture, L_{eq} (global error), is added. This corresponds to the error that would be made if no source separation was done and all the sound sources were taken into account without distinction.

Even if the global error is low ($\approx 2dB$), the use of the NMF to compute the traffic noise level produces a better estimation than taking the sound mixture with all the sound source. The Kullback-Leibler divergence produced the most interesting results with the lowest and the most stable RMSE. Surprisingly, the Itakura-divergence, despite its scale invariant property [11], has an error similar to the Euclidean distance. This result may change in the future with more complex and more realistic scenes.

5. CONCLUSION

In this article, we proposed to use the supervised NMF framework to estimate the road traffic noise levels based on acoustic measurements achieved in an urban context. In

[1] Open-source project available at: `https://bitbucket.org/mlagrange/simscene`

[2] `www.freesound.org`

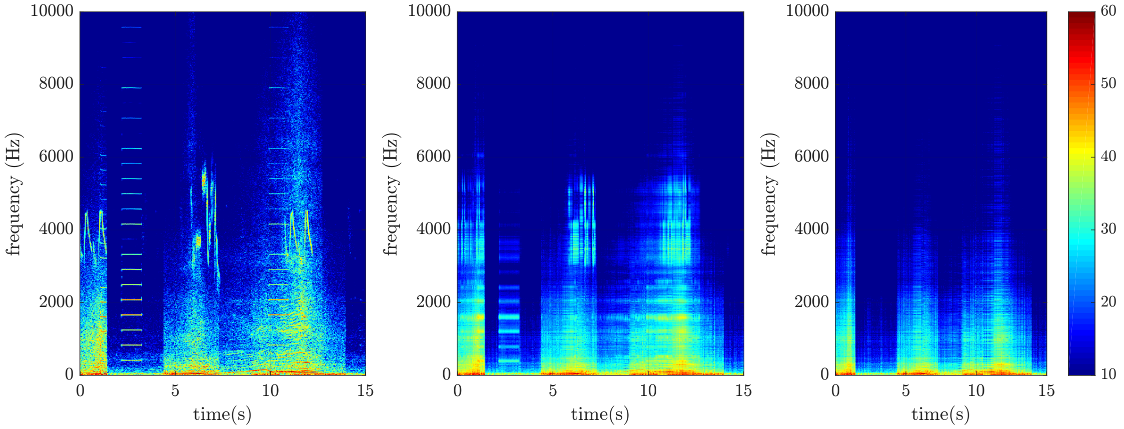

Figure 2: Spectrograms of a sound mixture composed with 3 sound classes (*car*, *horn*, *bird*). On the left, the initial audio spectrogram given by *simScene*, in the middle, the estimation $\tilde{\mathbf{V}}$ given by the NMF, on the right, the traffic car noise estimated $\tilde{\mathbf{V}}_{\mathbf{tr}}$ after the source separation.

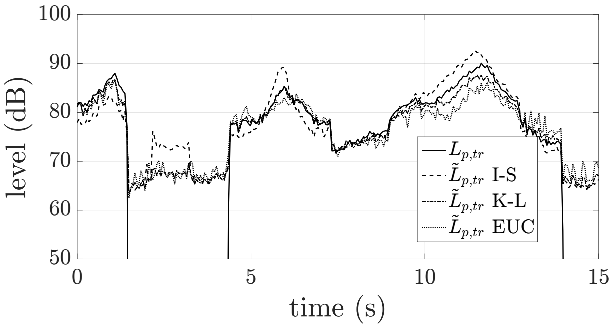

Figure 3: Evolution, according to time, of the actual sound pressure level, $L_{p,tr}$, and the estimated levels .

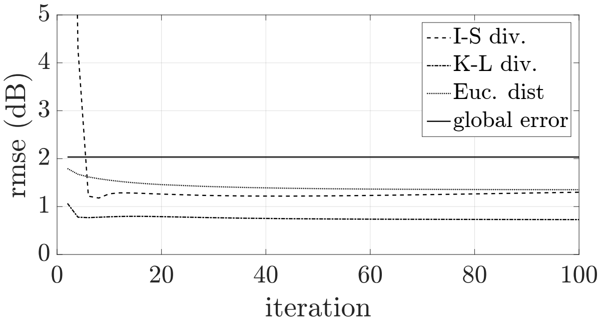

Figure 4: RMSE evolution

our opinion, such approach would find many applications in the environmental acoustics field such other than improving noise maps with acoustic measurements, for example acoustic biodiversity monitoring.

This method is tested on sound mixtures simulated using the *simScene* software which allows us to get the exact traffic contribution separately from the other sounds. The method is tested by comparing the equivalent sound level between the traffic element of *simScene* and the estimation given by the NMF for three cost functions. The first results show that this method gives a better estimation of the sound level than if the source separation is not done, thus demonstrating its interest. Both the road traffic time of presence and amplitude are accurately estimated, advocating for the use of the NMF for isolating the road traffic contribution. The Kullback-Leibler divergence results in the lowest errors and will therefore receive specific attention for future work.

Further investigations with more realistic and complex scenes are now required to confirm the behavior of the Kullabck-divergence. Then, some refinements of the NMF including acoustics considerations should improve the goodness of the road traffic noise levels estimation. For instance, the addition of some temporal constraints with a smoothness constraint within the NMF framework such as [13] [14] [15] to better model the temporal evolution of the traffic elements is currently investigated.

6. REFERENCES

[1] “Directive 2002/49/EC relating to the assessment and management of environmental noise.” [Online]. Available: http://ec.europa.eu/environment/noise/directive_en.htm

[2] A. Can, L. Dekoninck, and D. Botteldooren, “Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation ap-

proaches," *Applied Acoustics*, vol. 83, pp. 32–39, Sept. 2014.

[3] 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*, vol. 101, pp. 127–140, Jan. 2016.

[4] J.-J. Aucouturier, B. Defreville, and F. Pachet, "The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music," *The Journal of the Acoustical Society of America*, vol. 122, no. 2, pp. 881–891, 2007.

[5] B. Defreville, F. Pachet, C. Rosin, and P. Roy, "Automatic Recognition of Urban Sound Sources." Audio Engineering Society, 2006.

[6] G. J. Brown and M. Cooke, "Computational auditory scene analysis," *Computer Speech & Language*, vol. 8, no. 4, pp. 297–336, Oct. 1994.

[7] P. Comon, "Higher Order Statistics Independent component analysis, A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[8] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[9] P. Smaragdis and J. Brown, "Non-negative matrix factorization for polyphonic music transcription," pp. 177–180, Oct. 2003.

[10] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," in *Proc. EUSIPCO2005.*, 2005.

[11] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[12] M. Rossignol, G. Lafay, M. Lagrange, and N. Misdariis, "SimScene: a web-based acoustic scenes simulator," in *1st Web Audio Conference (WAC)*, 2015.

[13] C. Févotte, "Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 International Conference on IEEE*. IEEE, 2011, pp. 1980–1983.

[14] S. Essid and C. Févotte, "Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 415–425, 2013.

[15] T. Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.