DOI: 10.61782/fa.2025.0167

FORUM ACUSTICUM
EURONOISE 2025

AUTOMATIC DETECTION OF ROADWAY NOISE USING AI CLASSIFIERS AND SOUND LEVEL FEATURES

Erwann Betton-Ployon[1,2*] **Abbes Kacem**[1] **Jérôme Mars**[2] **Nadine Martin**[3]
[1] ACOUSTB, Saint-Martin-d'Hères, 38400, France
[2] Universite Grenoble Alpes, CNRS, Grenoble-INP, GIPSA-Lab, Grenoble, 38000, France
[3] ASTRIIS, Chambery, 73000, France

ABSTRACT

Noise management has become a major public health issue over several decades. To initiate protective measures against noise overexposure, it is essential to accurately evaluate annoyance. This involves detecting sound sources, their emission durations and associated sound levels. Among commonly encountered sources, road traffic is a prominent contributor to noise pollution, according to health organisation reports. This paper presents an automatic roadway noise detection system. The proposed method combines event detection and classification through a multi-layer approach. Sound event detection is ensured by distinct units. First, a sliding window enables signal preclassification by identifying specific patterns on its mel-spectrogram. Simultaneously, sound level features are used to detect prominent periods, often marking off sound events. Both units combined and sharpened provide detection of the most relevant sound events over the acoustic signal. Precise classification is issued by an additional AI model, dedicated to recognising roadway noise among various vehicle types. Our system operates in diverse soundscapes while maintaining a high level of roadway noise detection accuracy. It permits an automatic estimation of roadway noise contribution, which corresponds to equivalent sound level when roadway noise prevails. This estimation closely aligns with manual assessments from experts, validating the proposed system relevance.

Keywords: *roadway noise, sound event detection, acoustic signal processing, noise monitoring*

***Corresponding author**: *erwann.betton-ployon@egis-group.com.*

1. INTRODUCTION

Over the past decades, noise from transportation has emerged as a significant public health issue. Several World Health Organization reports point at induced health issues [1]. The European Union indicates that "approximately 1 in 5 people [...] are exposed to unhealthy levels of road traffic noise" among European citizens [2]. In such context, it is essential to provide adequate solutions to reduce noise overexposure. This involves first, a precise diagnosis of the impacted areas, by estimating roadway noise contribution. Roadway noise contribution corresponds to the equivalent sound level during periods of roadway noise prevalence over other sound sources. It is often evaluated thanks to a manual detection of roadway noise over acoustic signals. However, this is a tedious and time-consuming task. With the advent of Machine Learning and automatic Sound Event Detection models, a new perspective appears to automate this process. In that context, an automatic approach is proposed in this paper, through a three-staged method designed to handle long-term acoustic signals and separate roadway from non-roadway noise periods. Its detailed functioning is presented (Section 2) before discussing its performances in several contexts (Section 3) and concluding on various development perspectives.

2. METHODOLOGY

The presented process addresses the issue of automatic roadway noise contribution estimation, which requires the identification of roadway noise prevalence periods over other sound sources. The proposed method relies on successive units, presented in Fig. 1. *YAMNet sound ensemble preclassification* and *Final classification* operate at different levels to classify sounds. They are enhanced by two processing units, refining event detection and boundary definition thanks to the evolution of sound level.

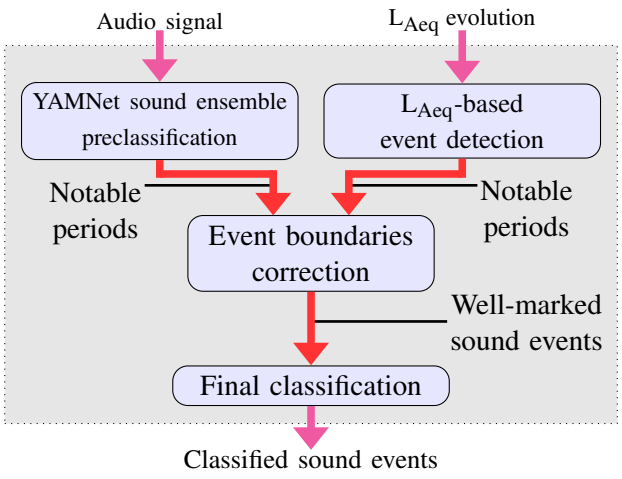

Figure 1. Description of the proposed automatic roadway noise detection method.

The first stage involves a coarse detection process, which utilizes two parallel analyses: one classification over the raw audio signal and another event detection system relying on sound level evolution. In the second stage, the *Event boundaries correction* algorithm processes the detected "Notable periods" from both units. This step provides a clear delimitation of each detected event, defining accurate start and end points, and identifying the monophonic or polyphonic nature of the event. In the last stage, a classifier processes each refined event to accurately classify the prominent sound source, determining whether it corresponds to roadway noise or not. These stages are now discussed in three dedicated subsections (2.1 to 2.3).

2.1 Long-term signal preclassification

This subsection details the two upper units of Fig. 1. The outcome of the open source environmental acoustic classifier YAMNet is exploited (section 2.1.1). In parallel, another algorithm uses A-weighted equivalent sound level (L_{Aeq}) evolution to detect singular events (section 2.1.2).

2.1.1 YAMNet sound ensemble preclassification

After the recording of a long-term acoustic signal, a broad preclassification stage helps understanding the various elements that form the complete soundscape. This analysis is carried out with a sliding window (0.96 s, 50%-overlap) over the full audio signal, to get an objective first view of the prevalent sources. Open-source environmental acoustic classifier YAMNet [3], developed by Google in 2020, was trained on more than 2 millions of audio segments, split into 521 labels [4]. As a CNN replicating MobileNet v1 architecture, it relies on mel-spectrogram representation of the audio signal [3]. Thus, characteristic spectral patterns are identified over the signal and associated with sound sources. However, data imbalance and label imprecision make the classifier unreliable in complex soundscapes. It is better to interpret its prediction as a hint on the involved sounds. Thus, several sound ensembles are defined: *Vehicles, Animals, Music or Speech, Alarms, Loud construction sounds* and *Background noise*. Each initial YAMNet class is associated with one of our sound ensembles. We look into the predicted sound ensembles instead of the YAMNet class, to prevent most misclassifications. This helps defining notable periods over the signal, where detected sounds clearly differ from background noise.

2.1.2 L_{Aeq}-based event detection

To complement sound ensemble preclassification, the equivalent sound level (L_{Aeq}, Eq. (1)) evolution is used to detect notable periods on the long-term signal. In this context, "notable periods" refer to energetic parts of the signal, having the most impact on the perceived nuisances.

$$L_{Aeq,T} = 10\log_{10}\left[\frac{1}{T}\int_0^T \frac{p_A^2(t)}{p_{ref}^2}\,dt\right] \qquad (1)$$

Two different properties or patterns are sought in the L_{Aeq} signal to detect these notable periods. Medium or long-term L_{Aeq} leaps signify the appearance of a loud sound source. Also, a high L_{Aeq} degree of variance is often inconsistent with roadway noise, and rather caused by speech, construction or animal sounds.

Two criteria are introduced to detect these two phenomena. For detecting medium and long-term L_{Aeq} leaps, we compare short-term $L_{Aeq,T=2\ min}$ to long-term $L_{Aeq,T=3\ hrs}$ (Eq. (1)) along the signal. Each time short-term exceeds long-term levels, a notable period is defined. On the other hand, detection of periods with a high degree of variance is performed with a threshold on $L_{Aeq,1s}$ standard deviation over a 20s-long sliding window.

2.2 Event boundaries correction

The stage role is to gather information from the initial preclassification and event detection stages (Notable periods in Fig. 1). Outcomes of both units are subject to sound emergence and duration criteria before they get merged. Thus, each event is defined with its starting time, ending time and one or several prevalent sound ensemble(s). These information are transmitted to the final classifier, aimed at improving the sound sources classifications.

2.3 Final classification

At this stage, a complete overview of the major sounds that prevail over the long-term acoustic recording is available. Nonetheless, numerous use cases require more precision than the defined sound ensembles (section 2.1.1).

For example, the *Vehicles* ensemble may refer to roadway noise as well as aircraft or railway noise. Therefore, we choose to use another classifier to bring in-depth sound source prediction. With such expectations, close attention should be paid to the chosen dataset and classifier model.

2.3.1 *Training dataset*

In order to build a performing supervised classifier, an appropriate training dataset is essential. It should cover most environmental sounds with few classes, while keeping consistency in defined classes. Our dataset architecture is inspired from other major environmental sounds datasets (AudioSet [4] and ESC-50 [5]). 13 classes are used, split into 4 categories, as depicted in Tab. 1.

Table 1. Categorisation of the classifier training dataset.

Roadway sounds	Human sounds, Animal & Music	Other vehicles sounds	Natural sounds
Dense road traffic noise	Human Voice	Airplanes & Helicopters	Wind
Cars	Music & Alarms	Railway	Rain
Motorcycles	Animals (Birds)	Construction machines	
Trucks & Buses	Animals (Insects)		

Internally collected signals, completed with few samples from ESC-50 [5] and Urbansound8k [6] make up for a total of 3500 files and around 10h of labelled data.

2.3.2 *Classifier model*

Regarding the classification model, the goal is to maximize the separation between roadway-related sounds and non-roadway noise. To build it, we drew inspiration from Contrastive Language-Audio Pretraining (CLAP) model, recently proposed in [7]. This model has already shown its effectiveness to encode audio waveforms and text signals. The same audio encoder architecture is trained on our data and used as input to a two-layer deep neural network. These last layers provide a 1D-classification vector corresponding to our dataset classes (Tab. 1).

3. DATASET AND RESULTS

3.1 Step-by-step application on a custom example

To ease the evaluation, a synthetic audio signal is built in mixing sound events from various sites over 30 minutes. Added sound events include train pass-bys, construction sounds, music, alarms, birds, insects and human voice, from 3 dB(A) to 30 dB(A) above residual sound level. L_{Aeq} evolution, reference labels and mel-spectrogram of this signal are shown in the three upper subplots of Fig. 2.

The detection part (2 preclassification stages and event boundaries correction) is first evaluated on its own. A 99 % recall and a 73 % precision are reached when comparing frames from reference and automatically detected events. This means that almost the entirety of reference events have been detected with proper boundaries. Meanwhile, some frames belonging to background noise were detected as sound events, explaining the 73 % precision.

Once the events are properly detected, the final classification unit is called. For each event, it predicts whether it belongs to roadway noise or not. This prediction is depicted in the last subplot of Fig. 2. To evaluate it, Average Precision (AP) [4] is computed regarding the *Roadway sounds* category. On this short example, a 0.92 AP is obtained. It shows that the detector recognizes various environmental sources emerging from background roadway noise. However, this small synthetic test sample is not large enough to conclude on the system performances.

3.2 Application on a large dataset

To carry on with the evaluation, real acoustic signals from 3 different locations were recorded and strongly labelled. Along each 24-hour long signal, roadway noise is manually detected and established as a reference. It permits the estimation of roadway noise contribution : equivalent sound level over all roadway noise-labelled periods. For each site, roadway noise contribution (denoted here as $L_{road,ref}$) is given in Tab. 2. It can be compared to $L_{glob,ref}$, equivalent sound level over the whole signal. The closer both indicators are, the more prevalent roadway noise is.

Table 2. Soundscapes and noise contribution of the 3 evaluation dataset sites.

Site	Major non-roadway source	$L_{road,ref}$	$L_{glob,ref}$
1	Construction machines	53.3 dB(A)	55.9 dB(A)
2	Birds	51.3 dB(A)	51.8 dB(A)
3	Railway	45.8 dB(A)	56.0 dB(A)

Chosen sites cover various use cases: prevalent roadway noise in sites 1 & 2, with disturbance from construction machines or birds. On site 3, railway noise prevails and road traffic is the secondary noise disturbance source.

Audio recordings from each site are supplied to our automatic detector, whose outcome is used to compute AP (Tab. 3). Also, automatic roadway noise contribution estimation ($L_{road,auto}$) is compared to the manual one ($L_{road,ref}$). This second comparison focuses on noisier periods, that matter most for noise contribution estimations.

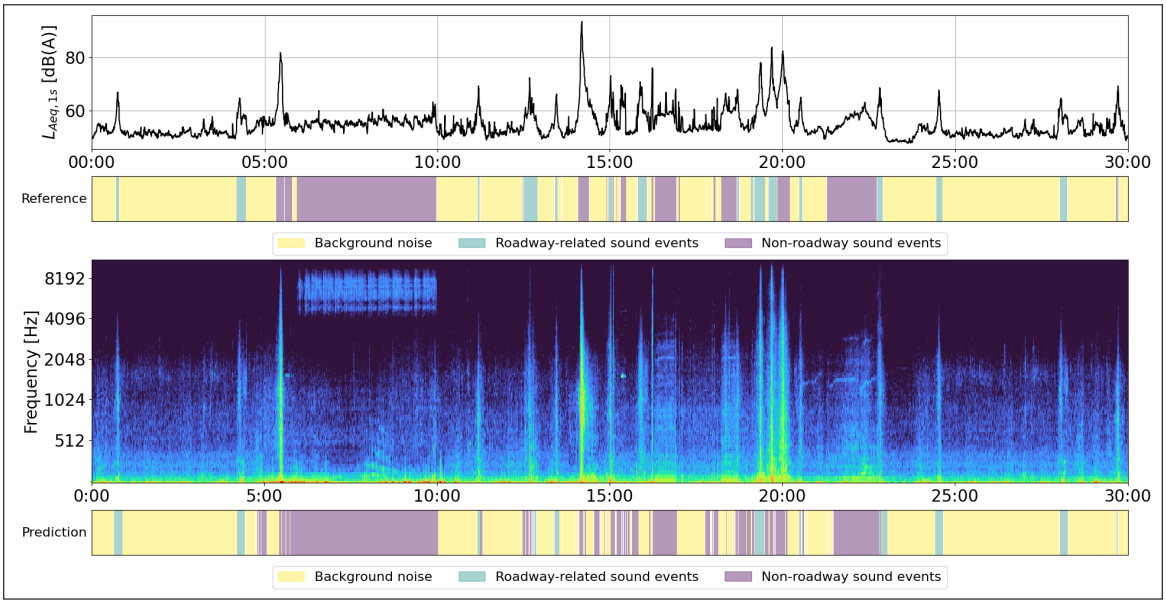

Figure 2. $L_{Aeq,1s}$ evolution, manually detected events, mel-spectrogram and predicted events for a 30-minute signal.

Table 3. Average Precision (AP) of the automated roadway noise detection and comparison between automated and manual roadway noise contributions.

Site	AP	$\lvert L_{\mathbf{road,auto}} - L_{\mathbf{road,ref}} \rvert$
1	0.88	0.1 dB(A)
2	0.91	0.1 dB(A)
3	0.96	0.7 dB(A)

On each site, AP stands above 0.88 and even reaches 0.96 for Site 3. Regarding automatic roadway noise contribution estimations, it also stands within a 0.7 dB(A) margin compared to the one obtained with manual labelling. Therefore, in three various contexts, our automatic achieves a high roadway noise detection rate, especially for noisier periods, allowing an accurate roadway noise contribution estimation.

4. CONCLUSION

In this paper, an automatic roadway noise detection method is proposed. With two classifiers enhanced by sound level analysis units, a robust sound event detection method is provided for road-impacted soundscapes. On 3 long-term signals from various recording sites, a minimum 0.88 Average Precision is reached. Above all, the proposed method aims at automating roadway noise contribution estimations. To this matter, it achieves satisfactory results on each site, with predictions standing within a 0.7 dB(A) margin compared to manual estimation. Future work on dataset or final model may improve accuracy. Still, this method already efficiently handles the automatic roadway noise detection tasks in various contexts.

5. REFERENCES

[1] World Health Organization *et al.*, "Burden of disease from environmental noise: Quantification of healthy life years lost in Europe," 2011.

[2] European Environment Agency, "Reported data on noise exposure covered by directive 2002/49/EC," 2024.

[3] E. Tsalera, A. Papadakis, and M. Samarakou, "Comparison of pre-trained CNNs for audio classification using transfer learning," *Journal of Sensor and Actuator Networks*, vol. 10, no. 4, p. 72, 2021.

[4] J. F. Gemmeke, D. P. Ellis, *et al.*, "Audio set: An ontology and human-labeled dataset for audio events," in *2017 IEEE international conference on acoustics, speech and signal processing*, pp. 776–780, 2017.

[5] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *Proceedings of the 23rd Annual ACM Conference on Multimedia*, pp. 1015–1018, 2015.

[6] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *Proceedings of the 22nd ACM International Conference on Multimedia*, p. 1041–1044, 2014.

[7] Y. Wu, K. Chen, *et al.*, "Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023.