PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

# Performance Analysis of the Acoustic Event Detector in the DYNAMAP's Rome suburban area

Rosa Ma Alsina-Pagès[(1)], Francesc Alías[(1)], Joan Claudi Socoró[(1)], Ferran Orga[(1)]

[(1)]GTM - Grup de recerca en Tecnologies Mèdia, La Salle - Universitat Ramon Llull (URL), , C/Quatre Camins, 30, 08022 Barcelona (Spain) , rosamaria.alsina@salle.url.edu

**Abstract**

Environmental noise is increasing year after year and, besides annoyance, it causes harmful health effects on people according to last 2018 WHO report. The Environmental Noise Directive 2002/49/EC (END) is the main instrument of the European Union to identify and combat noise pollution, followed by the CNOSSOS-EU methodological framework. In order to apply the END legislation, the EU Member States have to publish noise maps and action plans every five years. The use of Wireless Acoustic Sensor Networks (WASNs) changes the paradigm that addresses the END regulatory requirements as they enable the dynamic ubiquitous measurement of environmental noise pollution. Following the END, the DYNAMAP project develops a WASN-based low-cost noise mapping system to monitor in real-time the impact of road infrastructures in two pilot areas: Milan and Rome. To avoid biasing the noise maps with noise levels unrelated to traffic noise, an Anomalous Noise Event Detector (ANED) is included to remove them from the corresponding $L_{Aeq}$. The paper reflects the adaptation of the ANED algorithm to the WASN of the suburban area of Rome, which requires a specific analysis of the particularities of the suburban audio database, as well as future challenges and research on the generalization of the WASN.

Keywords: Noise monitoring, Anomalous Noise Event, Road Traffic Noise

## 1 INTRODUCTION

The World Health Organization (WHO) has recently stated that environmental noise is increasing year after year [21]. Besides causing annoyance, it provokes harmful health effects on people, becoming one of the main environmental health concerns [13, 14], having also a significant social and economic impacts [6]. This is specially important in those cities with high population and consequently, with high density of traffic, since it makes the traffic noise a dramatic problem that negatively affects the quality of life of their inhabitants [12].

At the European level, The Environmental Noise Directive 2002/49/EC (END) [9] is the main instrument defined to identify and combat noise pollution, together with CNOSSOS-EU methodological framework [16]. In order to apply the END legislation, the EU Member States have to publish noise maps and action plans every five years in large agglomerations. The design and deployment of Wireless Acoustic Sensor Networks (WASNs) has allowed to start changing the previous paradigm to address the END regulatory requirements from expert-based noise map generation to the dynamic ubiquitous measurement of environmental noise pollution by means of dynamic noise maps in smart cities [17].

Among the different WASNs that can be found in the literature -see [1] for a recent review about WASNs-, the one envisioned by the LIFE DYNAMAP project has been designed to monitor in real-time the noise levels generated by road infrastructures by means of low-cost acoustic sensors. To that effect, two WASNs have been installed in two pilot areas [19]. The first has been deployed in the District 9 of Milan, as an urban environment, and the second in the A90 highway as a suburban area. In order to avoid biasing the noise maps with noise levels unrelated to Road Traffic Noise (RTN), an Anomalous Noise Event Detector (ANED) is included in the monitoring to remove all Anomalous Noise Events (ANE) them from the corresponding A-weighted equivalent noise level computation ($L_{Aeq}$) before updating the noise map [20].

The ANED algorithm belongs to the so called acoustic event detection and classification, which is typically based on the segmentation of the input acoustic data into slices to identify some predefined target class [11]. To do so, these algorithms are typically trained with ad-hoc designed databases, considering a finite set of predefined acoustic classes [10, 18, 6]. As a first attempt to create an acoustic dataset to model the acoustic environments of the urban and suburban pilot areas in the framework of the DYNAMAP project, an expert-based recording campaign was conducted before the two WASNs were deployed [2]. This dataset allowed us to evaluate the feasability of designing the ANED algorithm as a binary classifier (ANE vs. RTN), using representative set of ANEs collected in real-environment, but not exactly in the final real-operation conditions since it was created before the two WASNs had been deployed. This paper works with a wider dataset, recently published in [5], that has taken into account the recording of all day and night to increase the number of types and the total duration of ANEs collected and labelled. This paper reflects the adaptation of the ANED algorithm to the performance real-time in the WASN in the suburban area of Rome. To this aim a specific analysis of the particularities of the suburban audio database is conducted [5], evaluating the accuracy of the ANED after being trained with the real-operation data collected in a weekday. Finally, the paper also discuses future challenges and research on the generalization of the ANED in the framework of the entire WASN.

The remainder of this paper is the following. Section 2 describes the real-operation conditions environmental audio database in the suburban scenario. Section 3 analyses the ANED training with the WASN-based dataset and performance in terms of accuracy. Finally, Section 4 discusses in detail the results obtained in the analysis and the future improvements to be conducted in the algorithm.

## 2 METHODOLOGY AND APPLICATION CONTEXT

This section briefly describes the main characteristics of the WASN-based dataset recorded in the suburban area of Rome in real-operation during a weekday [5], which has been used for the training and the test of the ANED algorithm.

### 2.1 The Rome Pilot in the DYNAMAP Project

In a first approach to collect representative anomalous noise events and traffic noise samples in a real environment, the DYNAMAP team conducted a recording campaign in several points of the A90 highway surrounding Rome before the WASN was deployed. The total amount of recorded data was 4 h and 44 min, which included 12.2% of ANEs [2]. This first dataset was labelled and subsequently used to train the ANED algorithm designed as a two-class classifier (see [20] for further details). However, after the deployment of the sensor network, the recording locations and conditions were severely different from the first recording campaign spots. The WASN of the Rome pilot area is composed of 24 low-cost sensors, of which 19 are high-capacity nodes and 5 are low-capacity nodes [3]. The sensor network installed can provide recordings all along the day and from any of the node locations for a limited amount of data. Taking advantage of the features of the WASN, a new dataset has been recorded taking advantage of all the nodes final locations, recording homogeneously distributed time frames [5].

### 2.2 Analysis of the Real-Operation Recording Campaign

The recording campaign was performed using the 19-node WASN deployed along the A90 motorway surrounding Rome [7] during a weekday (the $2^{nd}$ of November 2017) [5]. The recordings consisted of gathering 20 minutes of consecutive audio data at the beginning of each hour along the entire day. After the recording, the manual labelling was done over only the odd hours (01:00, 03:00, 05:00, ..., 23:00) to reduce the burdensome of an exhaustive process (e.g. considering the full set of recordings), which supposed a subsampling that still maintained the observation of relevant day times based on the mean noise equivalent profiles [5]. Hence, labelling process was carried out to 76 hours of recordings and following an equivalent procedure as the one explained in [2]. This includes also the computation of a contextual Signal-to-Noise Ratio (SNR) for each anomalous noise event (in dB). The labelled audio corpus contains - in terms of total time recorded and labelled - a 98.3% of RTN and 0.7% of ANE while the remaining 1.0% was attributed to *others* (a dummy class excluded from the analyses that represents those audio passages that were difficult to categorize in one specific class due to the complexity of the audio scene). This numbers differ substantially from the ones obtained in the preliminary dataset designed [2] (e.g. 3.2% of ANE). In this preliminary corpus, most of the recordings were performed during the day, when more ANE occur.

Up to fourteen ANE subcategories were identified in [5] weekday dataset: *airp*: airplanes, *alrm*: sounds of cars and houses alarm systems, *bike*: noise of bikes, *bird*: birdsong, *brak*: noise of brake or cars' trimming belt, *busd*: opening bus or tramway, door noise, or noise of pressurized air, *door*: noise of house or vehicle doors, or other object blows, *horn*: horn vehicles noise, *inte*: interfering signal from ad industry or human machine, *musi*: music in car or in the street, *sire*: sirens of ambulances, police, fire trucks, etc, *stru*: noise of portals structure derived from its vibration, typically caused by the passing-by of very large trucks, *tran*: stop, start and pass-by of trains, *trck*: noise when trucks or vehicles with heavy load passed over a bump.

Table 1 shows the ANE occurrences and their respective accumulated lengths (which is referred hereafter as *Total duration*) aggregated in the recordings for each ANE subcategory. It can be appreciated that *brak*, *trck* and *horn* are the most frequent type of ANEs, while *brak*, *tran* and *trck* are the ones with the largest accumulated duration. Sounds from brakes are the ones that both occur more often and for a longer time, while sounds from bikes, alarms, sounds of airplanes and music were the least frequently observed noise events.

Table 1. Number of ANE occurrences for each ANE subcategory.

| **ANE subcategory** | **brak** | **trck** | **horn** | **bird** | **busd** | **tran** | **door** |
|---|---|---|---|---|---|---|---|
| **# occurrences** | 328 | 296 | 148 | 105 | 66 | 57 | 43 |
| **Total duration (s)** | 574.4 | 304.5 | 106.1 | 101.7 | 41.4 | 466.9 | 8.4 |
| **ANE subcategory** | **stru** | **sire** | **inte** | **bike** | **alrm** | **airp** | **musi** |
| **# occurrences** | 39 | 36 | 4 | 2 | 2 | 1 | 1 |
| **Total duration (s)** | 34.2 | 191.4 | 51.9 | 4.3 | 12.9 | 5.5 | 3.6 |

Figure 1 shows the boxplots of ANE durations by ANE subcategory. *inte*, *tran*, *alrm* and *sire* are the ANEs with larger durations while *door*, *brid*, *stru*, *horn* and *busd* present the shorter ones.

Figure 2 shows the ANE SNR boxplots in terms of each ANE subcategory. From the figure, sounds of trains, sirens, horns, bikes, aiplanes and trucks are the main ANEs with higher SNR values. However, considering both SNR and durations, ANEs with more relevance are trains, sirens and horns.

## 3 EXPERIMENTS AND RESULTS

In this section, we detail the experiments performed to assess the ANED trained with the weekday suburban dataset [5], and we evaluate their results.

### 3.1 Details of the Training and Test Dataset

The ANED training and test was conducted with the weekday WASN-based dataset detailed in section 2.2. Feature extraction consisted on obtaining the Mel-Frequency Cepstral Coefficients (MFCC) from short-term windowed acoustic signals using a Hamming window of $T_a = 30$ *ms* length, with a time overlap of 50 %.

Only the 0.7% of the time labelled in the database corresponded to ANE compared to the 98.3% of RTN, which denotes a high imbalance between the two acoustic classes. It is well known that this large class imbalance can bias the classifier towards the majority class [15]. Furthermore, performing the two-class classifier training using the complete set of parameterized audio frames can be unaffordable due to the extremely high demand of computational resources to complete it. For these reasons, the data clustering and reduction procedure explained in [4] was applied to the RTN class, obtaining a more balanced dataset to suit better the training of ANED.

For the data reduction process we define $R$ as the percentage of feature vectors that will be selected from the

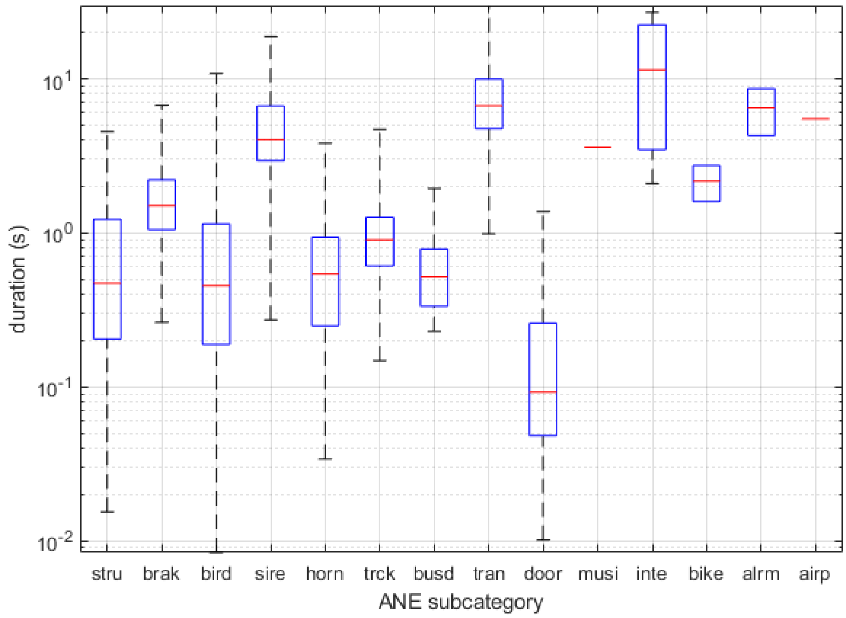


Figure 1. Duration boxplots for each ANE category. A logarithmic axis is applied to duration for illustration purposes.

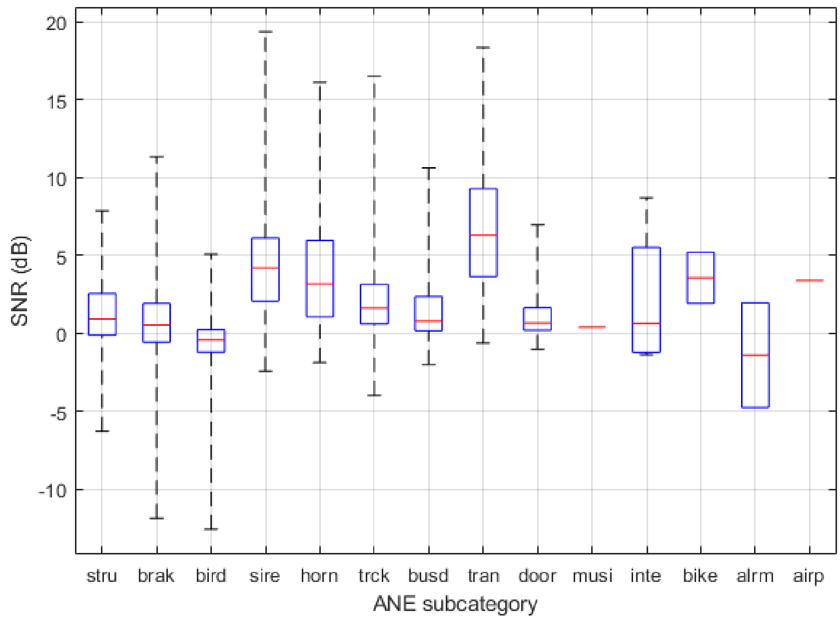


Figure 2. SNR Boxplots for each ANE subcategory.

original RTN feature subset. This process is based on the clustering of the RTN feature MFCC vectors using Davies-Bouldin criterion to automatically select the optimal number of clusters [8]. Next, $R\%$ of the audio frames weighted by their probability are selected for each cluster, according to their distance to the cluster centroid. The process is objectively validated comparing the normalized histograms between the audio frames and the cluster centroids before and after data reduction, which sould stay stable. In this work, the set of frames of audio features of the ANE class represented 0.915% of the total amount of frames of the RTN feature vectors - which is slightly higher due to the fact that the measure here is the window size with an overlap -, so the same value of $R = 0.915$ was applied to the majority class to obtain a balanced audio dataset, and 2 clusters were obtained from the Davies-Bouldin criterion.

Figure 3 shows the normalized histograms of the distances between audio frames and the two cluster centroids for the RTN class of the audio database. We can observe that histograms of the reduced data (c and d) coincide with those of the original data (a and b), which shows us that the data reduction maintains the distribution appropriately.

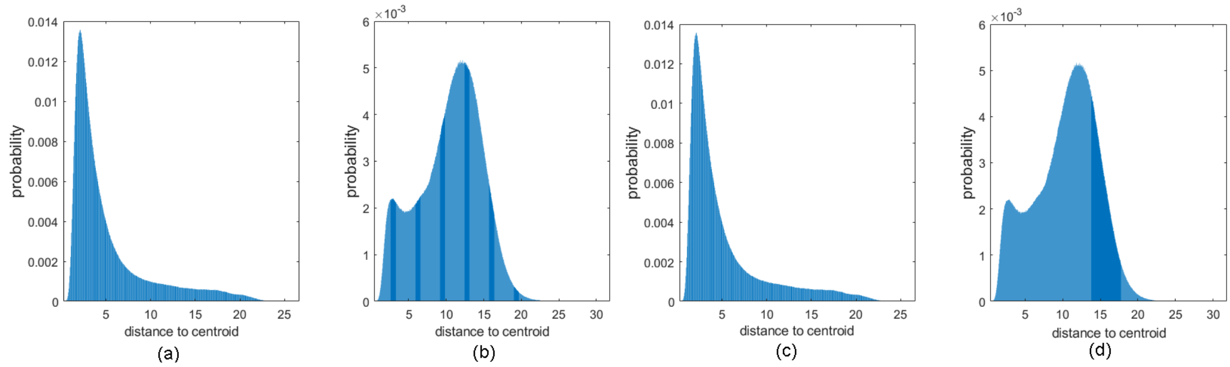


Figure 3. Normalized histograms of the distances between audio frames and the cluster centroids for the RTN class of the audio database. (a) and (b) correspond to the 2-cluster histograms for the original data while (c) and (d) correspond to the reduced data.

### 3.2 Training of the ANED with Balanced Dataset

As described in the Section 1, the ANED algorithm was designed as a two-class classifier using GMM as the core frame-based classifier, being the solution that obtains a proper balance between accuracy and real-time performance [20]. Moreover, a high-level classifier based on simple but yet effective majority vote of the low-level frame-based decisions was included to perform with higher stability at 1 s frame basis.

The assessment of the trained ANED with the audio database from the WASN real-operation has been conducted using a 4-fold cross-validation strategy as in [20]. Within each fold, the corresponding audio features have been randomized to provide the proper diversity for training and test data partitions. Moreover, the GMMs are trained using 32 Gaussians for both acoustic models (ANE and RTN), being the value that offers a good trade-off between performance and real-time behaviour [20]. Figure 4 shows the results obtained in terms of the F1 measures for each class (ANE and RTN), as well as the macro-averaged F1 value. These measures assess only the low-level frame-based ANED decisions, while high-level decisions have not been included in this study. As can be observed, the global performance attains a 78.01% , which constitutes an improvement of 17.38% in comparison with the performance obtained after training and testing the ANED algorithm with the preliminary recording campaign [20] .

Moreover, in Figure 4 we can observe that the ANE class obtains lower performance than the RTN class ($F1 = 75.08\%$ for the ANE class while $F1 = 79.46\%$ for the RTN class), but their difference is a moderate 4.38%. This states that ANED performance is nearly balanced for the two acoustic classes; the higher value obtained for RTN can be attributed to a higher homogeneity in the type of sounds [5].

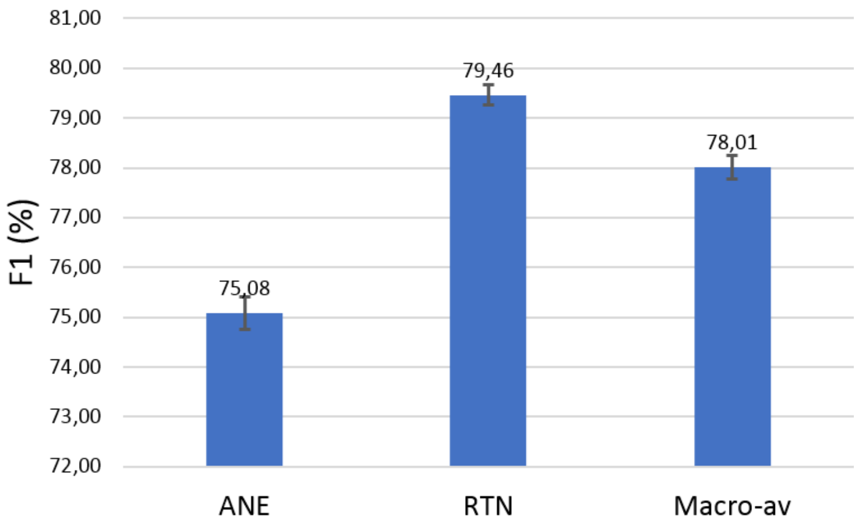


Figure 4. ANED performance assessment results using 4-fold cross-validation strategy. The standard deviations obtained from the 4-fold cross-validation scheme are shown as vertical lines in the top of each bar.

## 4 CONCLUSIONS

In this work, we have described the adaptation of the two-class ANED to the real-operation conditions of the suburban area of Rome in which a WASN has been deployed. The training has been performed using data from the 19-sensor network gathered during one weekday, by means of the 76 hours of raw audio data manually labelled in two categories (RTN and ANE). The distribution of the fourteen ANE subcategories found shows a high diversity of duration and SNR, and only three of them (trains, sirens and horns) present higher values on both characteristics. Before the adaptation, a clustering and data reduction process was applied to the majority class (RTN) to balance the two main categories of the dataset. The adapted ANED obtains a good performance results in terms of frame-level decisions, with an F1 macro-averaged value of 78.01% (17.38% higher than the performance obtained with the preliminary dataset).

As future work we plan to consider also weekend data from the 19-sensor WASN to work towards the generalization of the algorithm, with the aim of improving its performance and accuracy. Additionally, high-level 1 s-based decisions will be assessed considering other cross-validation schemes that combine maintaining original timing of signal frames while reducing the computational complexity of the balancing plus training processes. This way, a more reliable assessment could be obtained, closer to the real-life operation of the nodes of the WASN.

## ACKNOWLEDGEMENTS

The authors would like to thank Marc Hermosilla, Ester Vidaña, Alejandro González and Sergi Barqué for helping in the audio labelling. The research presented in this work has been partially supported by the LIFE DYNAMAP project (LIFE13 ENV/IT/001254). Francesc Alías thanks the Obra Social La Caixa for grant ref. 2018-URL-IR2nQ-029. Rosa Ma Alsina-Pagès thanks the Obra Social La Caixa for grant ref. 2018-URL-IR2nQ-038. Ferran Orga thanks the support of the European Social Fund and the Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement of the Catalan Government for the pre-doctoral FI grant

No. 2019_FI_B2_00168.

## REFERENCES


[1] F. Alías and R. M. Alsina-Pagès. Review of Wireless Acoustic Sensor Networks for Environmental Noise Monitoring in Smart Cities. *Journal of Sensors*, 2019(Article ID 7634860):13, 2019.

[2] F. Alías and J. C. Socoró. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Applied Sciences*, 7(2):146, 2017.

[3] R. M. Alsina-Pagès, F. Alías, J. C. Socoró, and F. Orga. Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN. *Sensors*, 18(4):1272, 2018.

[4] R. M. Alsina-Pagès, F. Alías, J. C. Socoró, F. Orga, R. Benocci, and G. Zambon. Anomalous events removal for automated traffic noise maps generation. *Applied Acoustics*, 151:183 – 192, 2019.

[5] R. M. Alsina-Pagès, F. Orga, F. Alías, and J. C. Socoró. WASN-based Suburban Dataset for Anomalous Noise Events Impact Analysis on Dynamic Road Traffic Noise Mapping. *Sensors*, 19(11):2480, 2019.

[6] J. P. Bello, C. Silva, O. Nov, R. L. Dubois, A. Arora, J. Salamon, C. Mydlarz, and H. Doraiswamy. SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution. *Communications of the ACM*, 62(2):68–77, 2019.

[7] P. Bellucci, L. Peruzzi, and G. Zambon. Life dynamap project: The case study of rome. *Applied Acoustics*, 117:193–206, 2017.

[8] D. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):224–227, 1999.

[9] EU. Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. *Off. Journal of the European Communities*, L189/12, 2002.

[10] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, and M. Vento. Reliable detection of audio events in highly noisy environments. *Pattern Recognition Letters*, 65:22 – 28, 2015.

[11] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. D. Plumbley. Detection and classification of acoustic scenes and events: An ieee aasp challenge. In *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 1–4, New Paltz, New York, October 2013.

[12] L. Goines and L. Hagler. Noise Pollution: a Modern Plague. *Southern Medical Journal-Birmingham Alabama-*, 100(3):287–294, 2007.

[13] H. Guite, C. Clark, and G. Ackrill. The impact of the physical and urban environment on mental well-being. *Public health*, 120(12):1117–1126, 2006.

[14] O. Hänninen, A. B. Knol, M. Jantunen, T.-A. Lim, A. Conrad, M. Rappolder, P. Carrer, A.-C. Fanetti, R. Kim, J. Buekers, et al. Environmental burden of disease in europe: assessing nine risk factors in six countries. *Environmental health perspectives*, 122(5):439, 2014.

[15] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, Sept 2009.

[16] S. Kephalopoulos, M. Paviotti, and F. Anfosso-Lédée. Common Noise Assessment Methods in Europe (CNOSSOS-EU). *Publications Office of the European Union*, Report EUR 25379 EN:1–180, 2002.

[17] D. Manvell. Utilising the strengths of different sound sensor networks in smart city noise management. In *Proceedings of EuroNoise 2015*, pages 2305 – 2308, Maastrich, Netherlands, 31 May - 3 June 2015. EAA-NAG-ABAV.

[18] A. Mesaros, A. Diment, B. Elizalde, T. Heittola, E. Vincent, B. Raj, and T. Virtanen. Sound event detection in the dcase 2017 challenge. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 2019. (In press).

[19] X. Sevillano, J. C. Socoró, F. Alías, P. Bellucci, L. Peruzzi, S. Radaelli, P. Coppi, L. Nencini, A. Cerniglia, A. Bisceglie, R. Benocci, and G. Zambon. DYNAMAP – Development of low cost sensors networks for real time noise mapping. *Noise Mapping*, 3:172–189, May 2016.

[20] J. C. Socoró, F. Alías, and R. M. Alsina-Pagès. An anomalous noise events detector for dynamic road traffic noise mapping in real-life urban and suburban environments. *Sensors*, 17(10):2323, 2017.

[21] E. van Kempen, M. Casas, G. Pershagen, and M. Foraster. Who environmental noise guidelines for the european region: a systematic review on environmental noise and cardiovascular and metabolic effects: a summary. *International journal of environmental research and public health*, 15(2):379, 2018.