

Article

BCNDataset: Description and Analysis of an Annotated Night Urban Leisure Sound Dataset

Ester Vidaña-Vila , Leticia Duboc , Rosa Ma Alsina-Pagès * , Francesc Polls and Harold Vargas

GTM—Grup de Recerca en Tecnologies Mèdia, La Salle—Universitat Ramon Llull, c/Quatre Camins, 30, 08022 Barcelona, Spain; ester.vidana@salle.url.edu (E.V.-V.); l.duboc@salle.url.edu (L.D.); francesc.polls@gmail.com (F.P.); haroldvb91@gmail.com (H.V.)

* Correspondence: rosamaria.alsina@salle.url.edu

Received: 27 July 2020; Accepted: 24 September 2020; Published: 2 October 2020



Abstract: Acoustic pollution has been associated with adverse effects on the health and life expectancy of people, especially when noise exposure happens during the nighttime. With over half of the world population living in urban areas, acoustic pollution is an important concern for city administrators, especially those focused on transportation and leisure noise. Advances in sensor and network technologies made the deployment of Wireless Acoustic Sensor Networks (WASN) possible in cities, which, combined with artificial intelligence (AI), can enable smart services for their citizens. However, the creation of such services often requires structured environmental audio databases to train AI algorithms. This paper reports on an environmental audio dataset of 363 min and 53 s created in a lively area of the Barcelona city center, which targeted traffic and leisure events. This dataset, which is free and publicly available, can provide researchers with real-world acoustic data to help the development and testing of sound monitoring solutions for urban environments.

Keywords: road-traffic noise; leisure noise; Barcelona; acoustic dataset; urban sound; annoyance

1. Introduction

More than four billion people (55% of the world population) live in urban areas, and the projection is that by 2050, this number will increase to seven billion, or two-thirds of the world population [1]. Barcelona, for example, has a population of over 1.6 million inhabitants (<https://www.idescat.cat/> (Population: 2019)) and receives nine million tourists every year [2]. Big cities like Barcelona combine a large range of industrial, business, and leisure activities, which can cause several environmental problems. Among these, acoustic pollution has gained increased attention over the last few years, as research has related the urban noise with adverse effects on the life expectancy and health of people [3–5]. In particular, the exposure to nocturnal noise was found to have a greater negative impact than daytime noise on long-term cardiovascular health, probably due to the repeated autonomic arousal during sleep [6,7]. Many of such studies focus on the negative health effects of traffic noise [8,9]. However, leisure noise is being increasingly recognized as an important challenge in cities with a high number of tourist and cultural offers, where the needs of residents and visitors need to be balanced [10,11], to the extent that the World Health Organization (WHO) has included recommendations on leisure noise in its recent Environmental Noise Guidelines for the European Region [12].

In order to address these problems, the European Commission has created the Environmental Noise Directive 2002/49/EC (END) [13] and the Common Noise Assessment Methods in Europe (CNOSSOS-EU) [14]. The former requires from Member States the development of separate strategic noise maps and noise management plans every five years for major roads, airports, railways,

and agglomerations of more than 100,000 inhabitants. The latter provides common methods that Member States are expected to use for such purposes. Historically, such maps have been manually built to ensure this separation of noise sources, which is a laborious and slow process that requires human intervention.

With the advances in sensor and network technologies, many cities have now deployed Wireless Acoustic Sensor Networks (WASNs). These networks have the potential to represent a paradigm change for city managers and the population alike. WASNs can enable the automatic and real-time generation of strategic noise maps and, consequently, the creation of more efficient policies and technical solutions for managing urban noise pollution and for designing sustainable urban and suburban soundscapes.

Such technical solutions often use machine learning (ML) algorithms for the automatic identification of noise sources [15,16], many of which require supervised training. That is, they use structured environmental audio datasets to train these ML algorithms. However, the creation of reliable environmental audio datasets normally involves the manual tagging of many hours of audio data, which is very labor- and time-consuming. In order to avoid this work, several ML solutions are being developed using datasets created by artificially mixing sounds from online digital repositories, such as FreeSound (<http://www.freesound.org>), Soundcloud (<http://www.soundcloud.com>), and AudioHero (<http://www.audiohero.com>). While these datasets allow algorithms to be trained with very large amounts of data, most of them gather sounds collected from several places and devices, which could make training more difficult. A real-life dataset recorded under controlled conditions and devices is closer to real operation conditions of the nodes of the WASN and allows for data augmentation when the classification algorithm requires a larger dataset [17,18]. Therefore, developers of sound solutions based on machine learning can benefit from free and publicly available real-world environmental audio datasets.

In this paper, we report on an environmental audio dataset created from 6 h of recording in a lively area of Barcelona city center, recorded as a joint collaboration with the Barcelona City Council. The recordings took place in a selected neighborhood that had produced a high number of complaints from residents. In particular, given the impact of nocturnal traffic and leisure noise on people's health, the spot chosen for the recording is well known for having both types of noise during the night.

Therefore, the contribution of this work is a precisely labeled night urban traffic and leisure sound dataset and its analysis, which is open and freely available to researchers and technicians. The analysis includes the duration of the events, the signal-to-noise ratio, the number of occurrences, the impact of each occurrence on the background noise L_{Aeq} , and the intermittency ratio (IR) of the entire data sample [19,20], which are metrics that have been related to healthy effects in different studies [20]. We envision this dataset being used, extended, and combined with others for different purposes, such as the development of noise identification and monitoring solutions, the creation of guidelines for designing sustainable urban and suburban soundscapes, and the comparison with other datasets for health impact studies.

2. Related Work

The environmental acoustic databases described in the literature, which are used by the machine listening research community to train and test different types of algorithms, are normally generated by artificially mixing sounds or from real-life recordings. The former allows the control of the signal-to-noise ratio (SNR) of the synthetic mixtures [21], also dealing with data scarcity by means of data augmentation [17,18]. Nevertheless, for this contribution, data augmentation techniques are not a key issue because the goal is to obtain and analyze a dataset exclusively from real operation data. We next review the literature on datasets recorded in real operation environments.

Valero [22] presents an automatic approach for the classification of road vehicles by means of their pass-by signatures. The team recorded a dataset with six categories (light vehicles, heavy vehicles, motorcycles, aircrafts, trains, and industrial noise), resulting in 90 sound samples for each category,

with a duration of 4 s each. Heittola [23] published a 1,133 min audio dataset that includes 10 different acoustic environments from indoor and outdoor recordings. Their goal is to detail how contextual information can be used in automatic sound event detection. The work attempts to simulate human behavior when detecting and identifying sound events by means of a two-stage process that includes the automatic recognition of the context and the subsequent detection of the sound event. Despite the presence of the temporal component in this analysis, their work does not refer to the dynamics and the location of the sound. Instead, it uses the surrounding events to improve its classification results. Foggia [24] presents a large dataset of audio events for a surveillance application using acoustic event detection. The dataset includes both long and short sounds and presents pieces of background noise with a significant noise level. The training dataset contains about 20 h, while the test set has around 9 h. In this sense, the goal of the dataset generation is completeness in terms of events coexisting with diverse background noise. Moreover, the same research laboratory developed a smaller dataset of about 1 h—also for surveillance purposes—focused on road acoustic events, which contains sound events from tire skidding and car crashes [25].

Alías [21] presents a 9 h and 8 min real-life acoustic database collected from the urban and suburban environments of the pilot areas of the LIFE+ DYNAMAP project [26]. This expert-based recording was carried out for discriminating Road-Traffic Noise (RTN) from Anomalous Noise Events (ANEs) through the Anomalous Noise Event Detection (ANED) algorithm running on low-cost acoustic sensors [15,27]. The ANEs, which correspond to 7.5% of the labeled data, were classified into 19 different subcategories after expert annotation, and the SNR levels were evaluated, taking as a reference the background noise. The SNR results ranged from -10 to $+15$ dB, also showing a wide heterogeneity of intermediate SNR levels. It is worth mentioning that the recordings in the urban area were conducted at the street level at pre-selected locations within District 9 of Milan [28], while the recordings in the suburban area were conducted on the A90 ring-road portals surrounding Rome (see [29] for further details). In the final stage of the LIFE+ DYNAMAP project, in [30], the same authors presented the production and analysis of a real-operation environmental audio database collected through the 19-node WASN of a suburban area of Rome. As a result, 156 h and 20 min of labeled audio data were obtained, differentiating among RTN and ANEs (classified in 16 subcategories). The preliminary suburban expert-based dataset contained 3.2% of the ANEs of the total recorded time, whereas this new dataset contains only 1.8%. A possible explanation to these differences is that the expert-based dataset recording was centered in daytime, and this WASN-based dataset was recorded day and night, where night shows low presence of ANEs with respect to the day. A complementary analysis to these works can be found in [31], which is focused on evaluating the aggregate impact of the ANEs occurring in the acoustic environments where the sensors of WASNs are installed.

Another WASN-based project that has collected real operation acoustic samples is the SONYC (Sounds of New York) project [32]. Bello provides a simplified taxonomy of the sounds of the city by means of a two-level hierarchy, dividing them into eight coarse categories and 23 fine labels [33]. The generated dataset is composed of 2351 recordings in the train split and 443 in the validation split, making a total of 2794 audio samples of 10 s each. The full taxonomy and details of the SONYC project dataset can be found in a previous work from the same authors [34]. The most innovative proposal of the SONYC project is that by means of the deployed network, the researchers can locate the distribution of the outdoor noise complaints and identify whether there have been, e.g., after-hours construction noise [32]. This identification can be done by means of the occurrence time of the group of annoying events, also allowing the retrieval and visualization of the data streams obtained for each complaint location. Nevertheless, their use of deep learning models requires a large amount of labeled data, which are unavailable for environmental sound; for this reason, the data necessary for the training of the model are obtained by means of an audio data augmentation, which deforms the data using audio transformations [32]. The final dataset used to train and test the network contains both real-life audio pieces and other synthetically mixed samples.

Mesaros [35] details an acoustic dataset recorded in multiple cities in Europe, which is an extension of the TUT 2018 Urban Acoustic Scenes dataset [36]. The original dataset contains recordings from Barcelona, Helsinki, London, Paris, Stockholm, and Vienna, and TAU 2019 adds Lisbon, Amsterdam, Lyon, Madrid, Milan, and Prague. The recordings were conducted with four devices simultaneously: (i) Soundman OKM II Klassik/studio A3 electret binaural microphone, (ii) Samsung Galaxy S7, (iii) iPhone SE, and (iv) GoPro Hero5 Session. Taking into account this variety of recording devices, the scenes were manually labeled to enable training and testing of machine learning algorithms. The dataset was used in one of the DCASE 2019 Challenges (<http://dcase.community/challenge2019/>), which included data from different recorded acoustic scenes and where the acoustic raw pieces were used together despite their different locations and origins.

3. Location Selection

Since our focus is the noise pollution caused by traffic and leisure activities, we chose to study large streets with large influxes of vehicles and with nighttime leisure activities. Our contacts in the Environmental Quality Department from the Barcelona City Council provided us with maps of the most problematic places in the district of Eixample based on the noise-related complaints from neighbors. The maps, which cannot be published for confidentiality reasons, contrast the areas with the greatest numbers of noise-related complaints about bars, restaurants, and music venues, many of which have terraces. After an analysis of the maps with the Department of Environmental Quality, we chose to focus our study in the following four parallel streets: Muntaner, Aribau, Enrique Granados, and Balmes, between the streets of Consell de Cent and Mallorca. All these streets have acoustic sensors or sound level meters located in light posts at around 4 m from the ground in places that were of interest for the Barcelona city council, as shown by the numbered circles in Figure 1.

We next summarize the analysis we conducted around these sensors:

1. Sensor 1 is a TA120 from CESVA with protection against external agents (such as birds, wind, rain, insects, etc.), which is a Class 1 precision sensor with programmable noise measurement integration time ranging from 1 s to 60 min, and is connected via optic fiber with the city council. It is located at Balmes street close to the corner with Consell de Cent street. Balmes is a street with a heavy traffic flow, but few leisure activities. The street has a sidewalk of two meters and four lanes dedicated to vehicles that circulate downwards, from the mountain to the sea. This is an important street to access the city center. Consell de Cent street, on the other hand, connects the city from west to east. Even with three lanes intended for traffic, it not a busy street. Since this area does not have many leisure venues, such as bars or nightclubs, in the street, most noise will be generated by vehicles, which also tend to be fewer than in upward streets. See picture 1 in Figure 2 for a photo of this street near the sensor.
2. Sound Level Meter 2 is TA025 from CESVA with an outdoor cabinet AR054 and an SC420 sound level meter, and is connected to the city council via 3G. It is located in Enric Granados, between the streets of Mallorca and Valencia. Enric Granados Street is one of the few pedestrian streets in the area, where the movement of vehicles is limited to only one lane and is at a reduced speed. It is also a street with many entertainment venues. These leisure activities are basically concentrated in bars and restaurants that have a closing time between 0:00 and 2:00 a.m. See picture 2 in Figure 2 for a photo of this street near the sensor.
3. Sound Level Meter 3 is TA025 from CESVA with an outdoor cabinet AR054 and a SC420 sound level meter, and is connected to the city council via 3G. It is located in Aribau, between València and Mallorca. This street connects the city center with the northern part of Barcelona and has three lanes dedicated to vehicles. Furthermore, traffic circulates uphill (from the sea to the mountain), which increases the noise from vehicles, which have to use more engine power to get around. This street also has a very active night life, with many bars and restaurants that close between 2:00 and 3:00 a.m. As such, the noise in this street is caused both by heavy traffic and by leisure activities.



Figure 2. Photos from the studied streets in Eixample, Barcelona, with numbers 1–4 representing the views of the streets close to the acoustic sensors. Picture 1 is Balmes street, picture 2 is Enric Granados street, picture 3 is Aribau street, and picture 4 is Muntaner street. Source: Google StreetView.

4. Recording Campaign

The recording campaign took place in two stages between March and June 2018. Since our goal was to create a dataset of raw acoustic data to distinguish between traffic and leisure noise, and since traffic is constant in the chosen location, we chose to carry out the study during the peak of the leisure activity hours; that is, on Saturdays between 22:00 and 03:00. Therefore, the first campaign was carried out on 17 March 2018 and resulted in two audio files. The first audio file has a duration of 124 min and 13 s, and the second audio file has a duration of 115 min and 27 s. The second campaign took place on 9 June 2018 and resulted in a single audio file of 124 min and 13 s, just like in the first campaign. Hence, the presented dataset has a total duration of 363 min and 53 s.

Despite having acoustic sensors in all the locations mentioned above, these had technical limitations. They were unable to make recordings for long periods of time (just for a few seconds) and could only store the sound level and frequency, as well as the time when a noise event went over a particular dB level. For this reason, we chose to use a ZOOM H5 [37] recorder with an attached microphone working at 44,100 Hz and with a microphone sensitivity of -45 dB/Pa that saved the recordings in .WAV format instead, as shown in Figure 3a. The recorder was placed close to the location of the acoustic sensor 3 on a first-floor balcony at 4.5 m above the street level (see Figure 3b). This also allowed us to record the sound without intervening in the street and altering people's or car's behaviors by placing the equipment at the sidewalk. Finally, two technicians standing in the street under the recorder observed the area and took independent notes about the noises and activities throughout the recording campaign in order to facilitate data labeling and analysis.

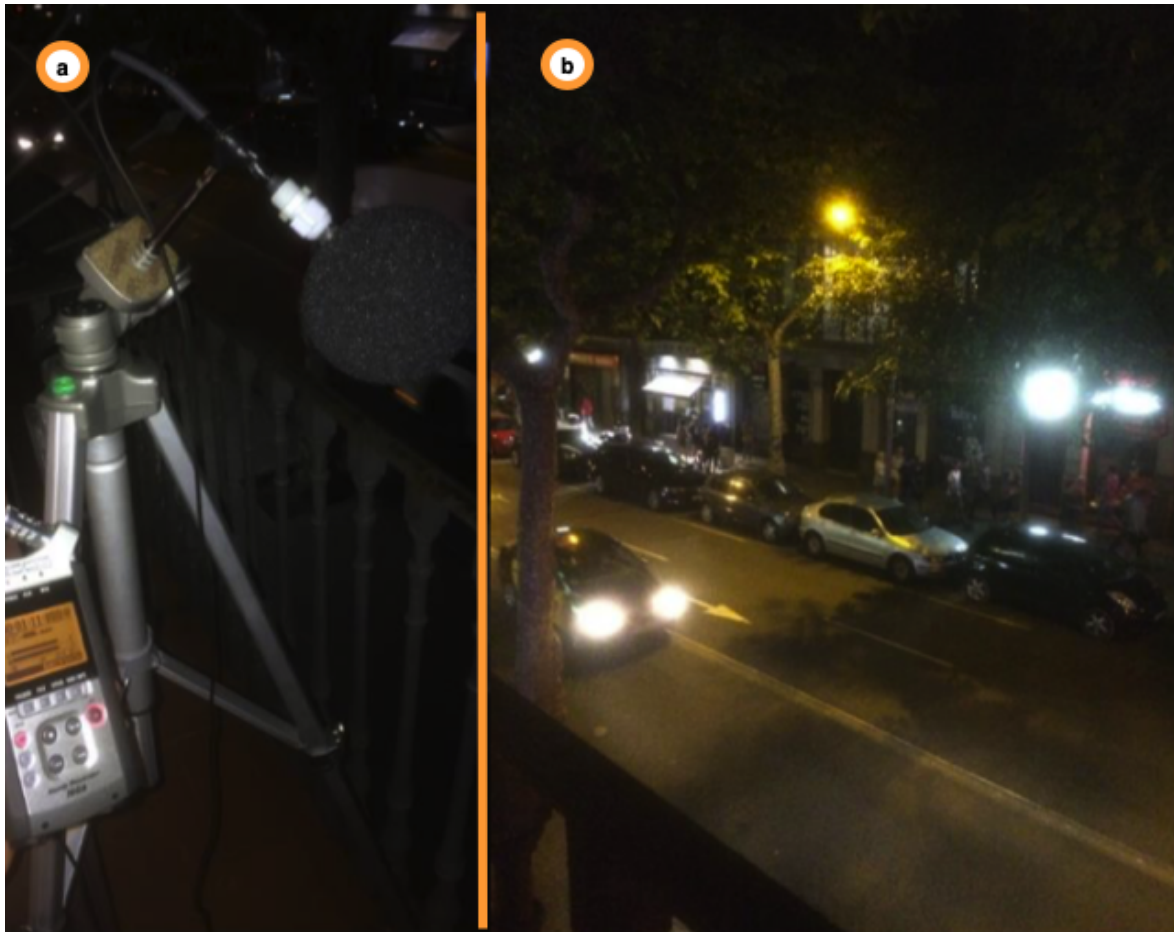


Figure 3. Photos of the recording device and its relation to the street level. (a) Shows the Zoom recorder on a first-floor balcony, while (b) shows the view of the street from this balcony.

5. Data Labeling

In order to create a dataset that could be used to train artificial intelligence (AI) algorithms, we labeled each audio event using the Audacity program. This is an audio recording and editing software that allows one to name sections of sound (i.e., noise events) and associate a text label to them, as shown in Figure 4. The result for each of the audio files recorded was a text file (.txt) containing the beginning and end of each section as well as their corresponding labels, with the following structure: “starting_time_event(seconds) ending_time_event(seconds) label”. The seconds are always referenced to the beginning (second 0) of each individual audio file.

The labeling process was independently carried out by two technicians. Careful and consistent labeling is very important to ensure an effective training of AI algorithms. For this reason, fragments of the labeled audio were cross-checked by experts of the LIFE+ DYNAMAP project [26], who have extensive experience in labeling similar recordings. The labeling process was repeated up to three times, until experts confirmed that the labels from the different fragments were consistent. The resulting dataset is composed of the events and their respective labels, which are shown in Table 2.

Since the main goal for which this dataset was created was to study the distinction between leisure and traffic in the city of Barcelona, the third column in Table 2 also shows how each event was classified between the leisure and traffic categories. We have classified as leisure all sounds related to people, blinds (bars and restaurants), and music. The traffic category, on the other hand, contains sound events related to vehicles. The authors would like to highlight that the *rtn* event represents the road traffic noise produced by different vehicles. The sound of vehicles could be considered background noise in cities instead of an acoustic event, as it does not have a clear start and stop time and it is more or

less stationary. However, as the purpose of this dataset is to compare the impact of traffic noise and leisure noise in the city center of Barcelona, only the road traffic noise that presented a noise level high enough to mask any other events occurring simultaneously has been tagged as *rtn*. This will be compared to the rest of the acoustic events in future analyses in Section 6.

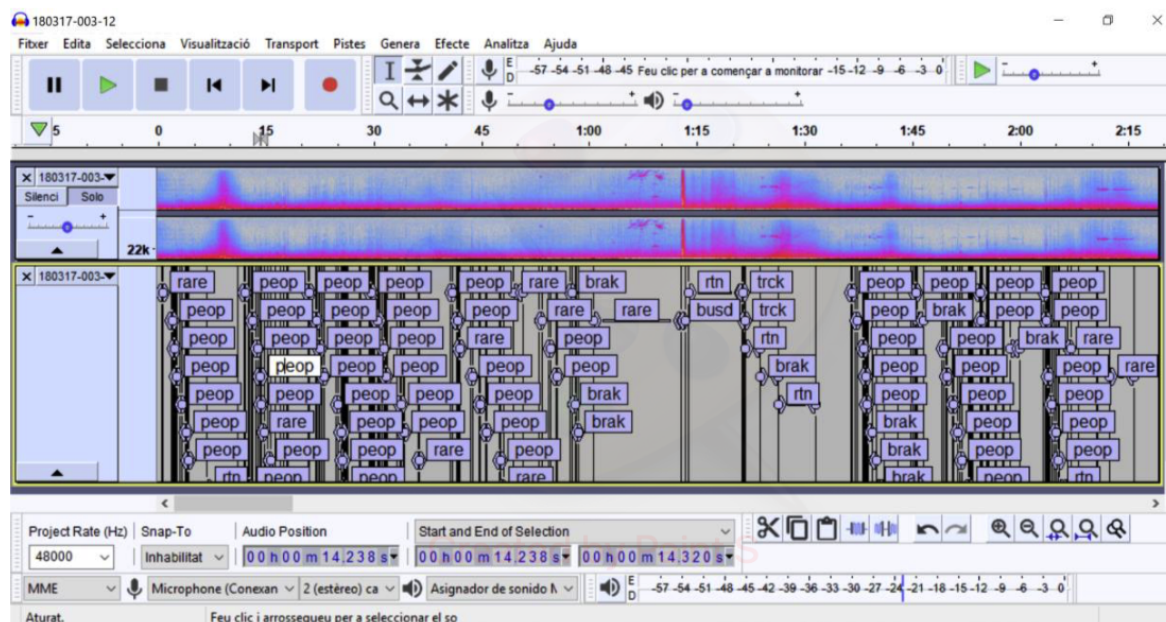


Figure 4. Screenshot of the Audacity program showing a labeled audio fragment.

It is also worth mentioning the *rare* noise event, which does not belong to any of the aforementioned categories. Such events are sounds not easily recognized by a human—their source was not possible to determine, even after consulting the notes taken by the observers during the recording campaigns—or a mix of sound events that cannot be classified in a single category, such as two events occurring at the same time (e.g., high-level *rtn* and high-level *peop* events occurring at the same time).

Table 2. Event types considered for the dataset and their respective descriptions and categories (leisure/traffic).

Label	Event	Category
bkmu	A mix of background city noise and music	Leisure
blin	The opening and closing of blinds	
coug	Person coughing	
door	Door or knock noise (house, car, or object)	
musi	Music in a car or in the street	
peop	People talking	
troll	Sound of wheels of suitcases (trolley)	
whtl	Whistle	Traffic
brak	Noise of brake or car's timing belt	
busd	Opening of a bus or tramway door	
horn	Horns of vehicles (cars, motorbikes, or trucks)	
rtn	High-intensity road traffic noise	
sire	Sirens of ambulances or police cars	None
rare	Unrecognizable noise	

6. Dataset Analysis

After recording and labeling the audio files, a detailed analysis was performed in the dataset to determine the main features of the sounds. Table 3 shows the numbers of events detected on each of the audio files of the recording campaigns. As can be observed, the class that presents the most events is—by far—*peop*, followed by *rtn*, *brak*, *door*, and *rare*. As *peop* and *door* are categorized as leisure sounds, we can deduce that the zone where the audio files were recorded contains mostly leisure-time noises. In order to be able to confirm this deduction, deeper analyses were carried out.

Table 3. Number of events labeled on each audio file and their durations in seconds.

Event	File #1	File #2	File #3	Total
bkmu	11 events, 49.65 s	0 events, 0 s	7 events, 25.61 s	18 events, 75.27 s
blin	1 event, 2.66 s	10 events, 6.34 s	9 events, 29.41 s	20 events, 38.42 s
brak	520 events, 428.51 s	134 events, 144.68 s	156 events, 239.57 s	810 events, 812.77 s
busd	15 events, 8.62 s	36 events, 14.22 s	12 events, 9.42 s	63 events, 32.27 s
coug	15 events, 5.50 s	0 events, 0 s	5 events, 4.14 s	20 events, 9.65 s
door	455 events, 108.85 s	138 events, 52.69 s	138 events, 95.46 s	731 events, 257.01 s
horn	36 events, 29.17 s	21 events, 22.46 s	36 events, 37.54 s	93 events, 89.19 s
musi	12 events, 51.88 s	2 events, 7.95 s	2 events, 9.15 s	16 events, 68.95 s
peop	810 events, 578.93 s	1068 events, 803.82 s	578 events, 2383.50 s	2456 events, 3765.86 s
rare	404 events, 787.65 s	135 events, 319.68 s	132 events, 474.50 s	671 events, 1581.84 s
rtn	574 events, 2132.89 s	343 events, 3789.41 s	233 events, 1968.52 s	1150 events, 7889.90 s
sire	4 events, 36.25 s	2 events, 16.41 s	3 events, 23.74 s	9 events, 76.42 s
troll	8 events, 25.38 s	1 event, 0.56 s	2 events, 8.35 s	11 events, 34.30 s
whtl	2 events, 0.91 s	0 events, 0 s	6 events, 7.13 s	8 events, 8.05 s
Total	2867 events, 4245.81 s	1890 events, 5178.34 s	1319 events, 5316.11 s	6076 events, 14,739.95 s

Apart from the number of occurrences of each type of sound, the duration of each of the events is important when considering the noise impact of each class. Hence, a boxplot displaying the duration of each labeled situation is presented in Figure 5. As can be observed in the figure, on average, the class that presents the greatest duration is *sire*, but considering that there are only nine samples of this type of sound, this event type may not be as relevant as other classes with greater numbers of occurrences. However, it is worth noting that the class that contains more events (i.e., *peop*) is usually short in time (less than 1 s on average) in comparison to other events that also appear several times, such as *rtn*. The higher number of occurrences of the *peop* class and the short duration of each of the occurrences are balanced with the fewer number of occurrences of the *rtn* class and the longer duration of each of the occurrences of this class. This fact is explained by the characteristics of each noise source: *peop* is labeled each time anybody speaks, as a conversation is mainly not considered a continuous event; *rtn* is usually considered a continuous event, despite that it contains several passes or other vehicles.

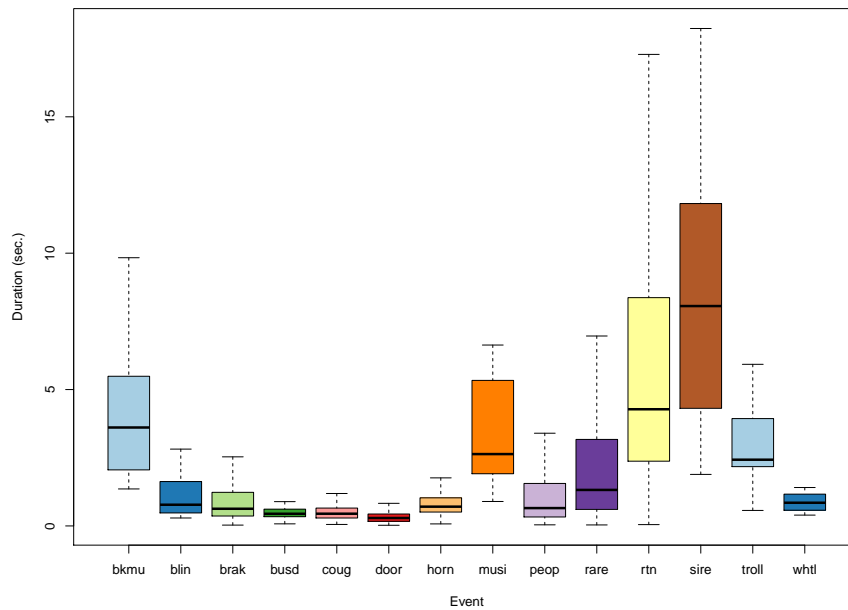


Figure 5. Boxplot of the durations (in seconds) of the labeled events for each of the classes of the dataset.

6.1. Signal-to-Noise Ratio Calculation

In fact, calculating the duration of an audio event is meaningless without considering the level of noise that it produces. Indeed, short events with a high level of noise (impulse noises) are usually perceived as more annoying by neighbors in comparison to long events with low noise level [38]. Hence, two important parameters to be taken into account are the SNR (signal-to-noise ratio) and the impacts of the different events. For this reason, and to be able to compare the traffic noise against the leisure noise, a boxplot of the SNR is shown in Figure 6.

To calculate the SNR, and as the background noise of the different labeled events is not stationary, we applied the methodology detailed in [39]. That is, we first calculated the power of the spectrum of the labeled event (considering that the event is the “Signal”), and then obtained the power of the background noise by getting samples from before and after the event. After that, we divided the power of the signal and the power of the noise to obtain the final value of the SNR in dB. This means that the obtained value is always relative to the sounds that happen right before and right after the event. In the case that an event is followed by signal with more power, the SNR would have a negative value, indicating that the event is less noisy than its environmental noise before and after the event.

As an example, Figure 7 depicts the spectrogram of an event labeled as *door*. In the figure, samples used as “Signal” or “Noise” have been marked with arrows. The N central samples (labeled as *door*) were the ones used to calculate the power of the signal, and the $\frac{N}{2}$ samples before and after the event were used to calculate the power of the noise. This means that the SNR was computed as:

1. Power of the event:

$$PS = \frac{\sum_{i=1}^N \text{Signal sample}_i^2}{N} \quad (1)$$

2. Power of the background noise around the event, considering the $\frac{N}{2}$ samples before and after the event depicted in Figure 7:

$$PN = \frac{\sum_{i=1}^{N/2} \text{Noise 1 sample}_i^2 + \sum_{i=1}^{N/2} \text{Noise 2 sample}_i^2}{N} \quad (2)$$

3. Finally, the ratio is calculated and converted to dB:

$$SNR = 10 \log_{10} \left(\frac{PS}{PN} \right) \quad (3)$$

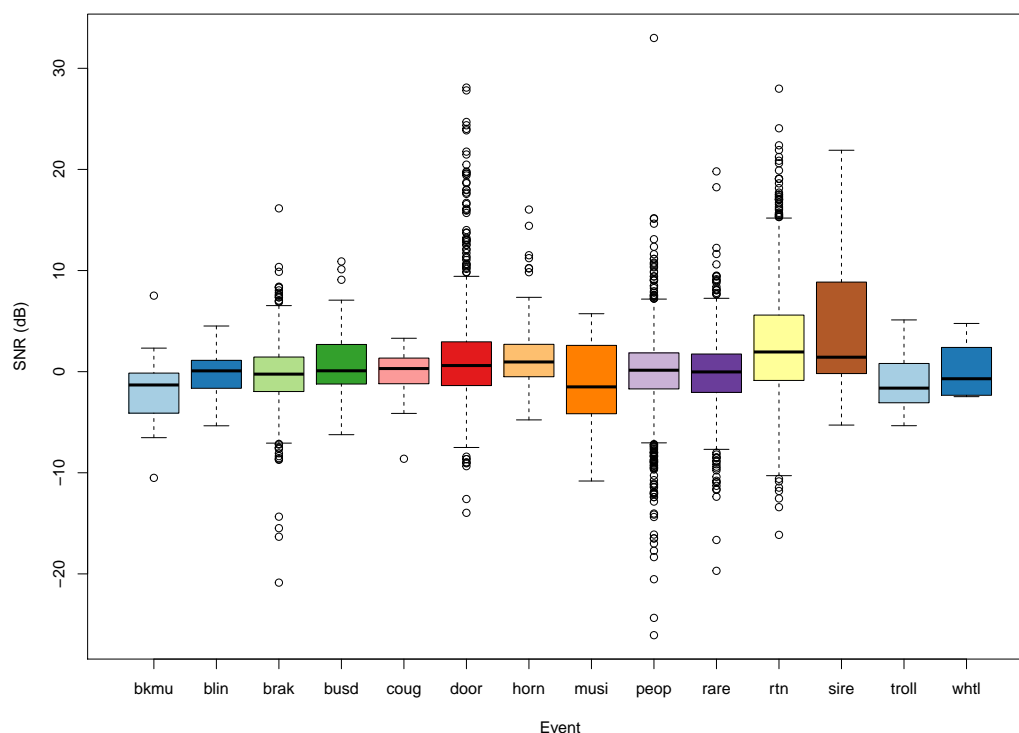


Figure 6. Boxplot of the signal-to-noise ratio (SNR; in dB) of the labeled events for each of the classes of the dataset.

From this analysis, we can conclude that the “traffic” events have, on average, a higher value of SNR. Concretely, the events that have, on average, the highest SNR values are *sire*, *horn*, and *rtn* (classes included in the traffic category). Considering together the durations and SNRs of these events, we can see that, whereas road traffic noise and siren sounds typically present a duration of few seconds and a high value of SNR, the *horn* event is almost an impulse noise (very short in time and with a high level of SNR). However, we can see in the boxplot that a few occurrences of *rtn* and *sire* events have SNR values of around 20 dB, meaning that independently of their duration, they present an extremely high noise level in comparison to their surrounding environment.

Regarding the *door* events, we can see that, whereas some occurrences have a high SNR (reaching a maximum of about 30 dB), some other occurrences present negative values. The main reason behind this phenomenon is that the dataset contains two main door types tagged with the same label. On the one hand, the closing of car doors has been labeled as *doors*. Because of the materials of the car, and as they are very heavy, when people close these types of doors, they make an impulse sound (very short in time and with a lot of energy). One example of this type of door event is shown in Figure 7. On the other hand, doors related to leisure places, such as such as bars, have also been tagged as *doors*. These doors are lighter and typically present lower levels of energy.

Another interesting observation is that all the musical events (*musi* and *bkmu*) present SNR values smaller than 0 dB on average, meaning that they are less noisy than their surrounding environmental noise. However, something that must be taken into account is that, when labeling the recorded audio files, the authors noticed that all the musical sounds originated from cars passing by with their windows opened, so they were all surrounded by *rtn*. Hence, as musical sounds are always surrounded by an event that typically presents a positive SNR by itself, they are partially masked in the recordings.

Finally, analyzing the most common event in the dataset, *peop*, we can see that some occurrences present a positive SNR and some other occurrences present a negative value. The main reason for this is that, during the recording campaign, there were two types of people in the street. On the one side, there were people walking by the street and talking normally to each other. On some occasions,

these occurrences were masked by other events or could not be distinguished from background noise, so they present negative values of SNR. On the other side, there were people standing in the street and having loud conversations—that even included a few shouts—close to the recording sensor. These are the occurrences that present high SNR values.

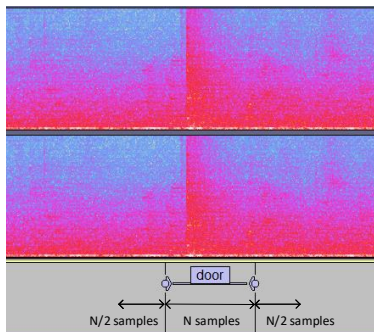


Figure 7. Spectrogram of a *door* event indicating which samples were used as signal or noise for the SNR calculation.

6.2. Event Impact Analysis

Apart from the SNR calculation, we also calculated the impact of each event. The impact measures the contribution of the labeled event over the equivalent level of a certain period of time after applying the A-weighting filter [40]. This indicator was calculated following the methodology explained in [39]. As in the cited work, the impact is relative to the 5 min of L_{Aeq} measured surrounding the event. To obtain the final impact value, the L_{Aeq} of the signal is obtained by first applying the A-weight filter and then obtaining the equivalent level. Then, the labeled event is removed from the audio file and replaced by an interpolated value of the background noise to maintain a continuous energy of the signal. Finally, the impact is measured as the subtraction between the initial L_{Aeq} and the L_{Aeq} without the labeled event. For more details about this procedure, the reader is referred to [39].

The value of impact of an event is highly related to the type of event that is being measured. If it is an event that presents a high value of SNR or its duration is long, the impact will have a high value. If both conditions are met (the SNR is high and the duration is long), the impact will be extremely high. As there are some events that usually have similar durations (e.g., a *door* event is not likely to last more than 1 s, while a *sire* event will often last for several seconds), the impacts of the events from the same class may have similar values. Impact values can be deduced by looking at the boxplots presented in Figures 5 and 6. Events that present smaller boxes in the boxplots (such as the *busd*, *cough*, or *door* will typically have similar values of impact. However, events that present bigger boxes, such as the *rtn* or *sire*, will have a wider range of impact values, as the duration and SNR can be very different when comparing events belonging to the same class.

Figure 8 shows the impact of all the labeled events divided among the three recording campaigns. As expected, the Figure suggests that the events that have longer durations also present the highest impact values. It can also be observed that both the traffic sounds and the leisure sounds have similar values of impact (the circles presented in the Figure have similar sizes). Actually, on the one hand, only a few events present an increase of L_{Aeq} greater than 0.01, which means that all the events have similar contributions to the noisiness of the environment. On the other hand, there are several events—which usually have a duration close to milliseconds—that present a negative impact, which means that their noisiness is lower than the average background noise. The results observed are highly correlated with the two previous boxplots (Figures 5 and 6). More precisely, the most remarkable event is the *rtn*, which, apart from being the class that presents occurrences with the highest durations, also presents the biggest circle, meaning that the impact is more notable.

Figure 8 is also useful to see the duration, SNR, and impact of the events for each individual audio recording, as there are several notorious differences between the features of the events of the different

classes. Whereas in the first and second files, the *peop* class typically has a duration smaller than 10 s (with only two exceptions in audio file #1), the third audio file presents several samples of *peop* talking in the street with longer duration. Given that the recording campaigns took place on different days (audio files #1 and #2 were recorded one day and audio file #3 was recorded on a another day), it is normal that the number of people in the street standing close to the sensor is slightly different.

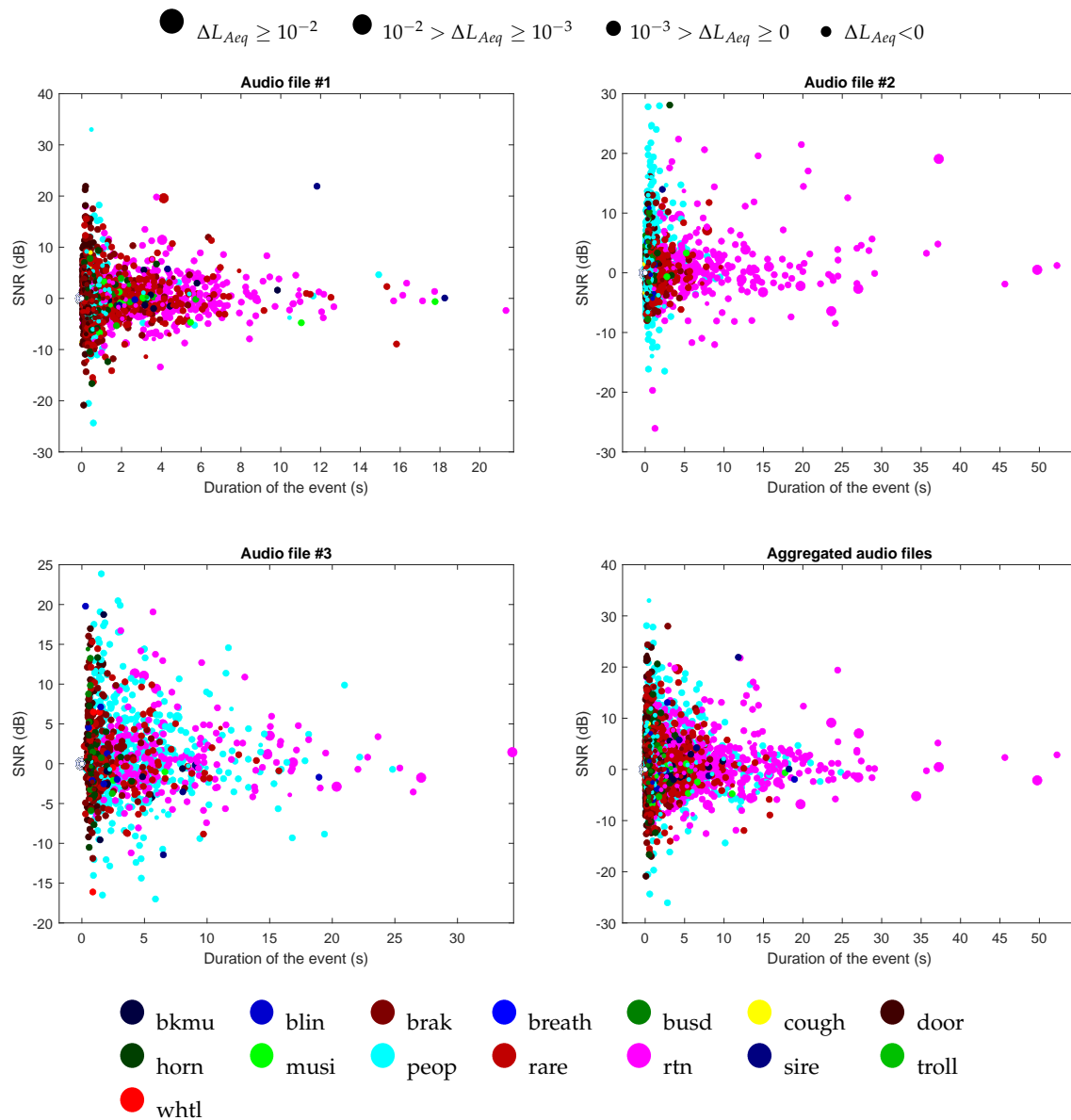


Figure 8. Analysis of the impacts of the audio events.

6.3. Analysis of the Time–Event Distribution

Figure 9 depicts the occurrence of the events in time. The figure shows three sub-plots, each one representing one of the audio files used to generate the dataset. The x-axis of each sub-plot represents the time of the audio file in minutes, and the y-axis contains the 14 different possible categories of the labels shown in Table 2. To display each event’s occurrence, a colored dot has been drawn at its starting second in the x-axis and at the height of its label. For example, if a *whtl* event happened at minute 0 in the third audio file, a dot would be drawn around the top-left corner of the last sub-plot.

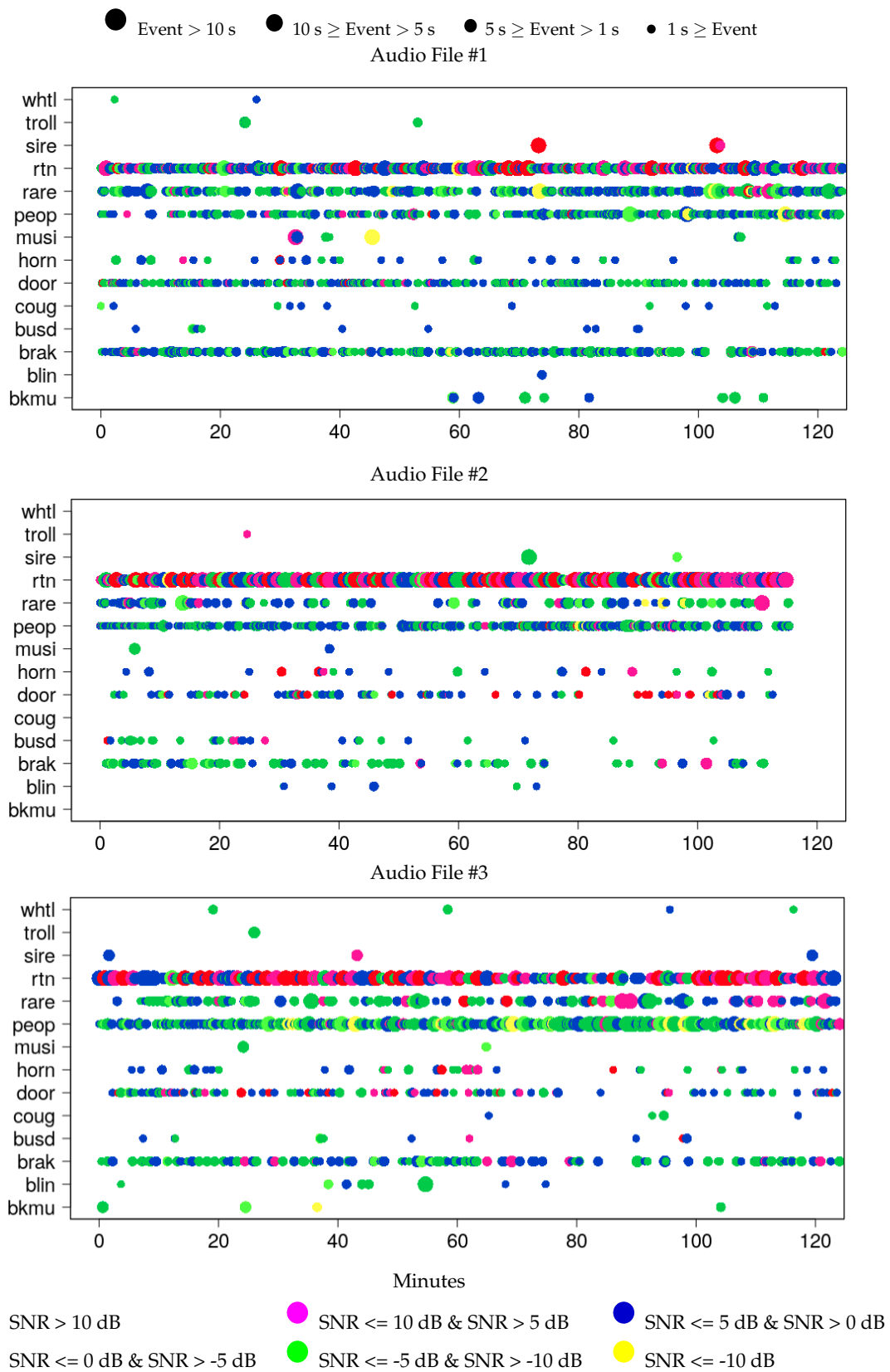


Figure 9. Distribution of the labeled events of each audio file in time. Each subplot stands for the results of an audio file. The x-axis is the time in minutes and the y-axis represents the different labeled categories that can be found in the dataset. Each dot corresponds to an event of the y-axis type starting at the moment indicated in the x-axis. The color of a dot represents the SNR of that concrete event and the size of the dot represents the duration of that event.

The color of the dot represents the SNR value of that event, calculated as explained in Section 6.1. The red, purple, and blue dots represent the events that have a positive SNR, and the rest of dots represent the events that have a negative SNR. As shown in the plots, typically, the events that present higher SNR in the three audio files are *rtm* and *sire*, as well as some *brak* and *door* events. Whereas *peop* is the category that has more occurrences, there are just a few events with SNR values greater than 10 dB. The size of each dot represents the duration of the event. Longer events are associated to bigger dots, whereas the events shorter than 1 s are represented with the smallest dots.

Regarding the time distribution of the events, we can see that there is not an accurate pattern for the occurrences of events of different classes or among the different audio files. Both traffic and leisure noises happen all along the audio files in a uniform distribution. The events that are more stable in the dataset are *peop* and *rtm*; they are present during all the audio files (we can see there are dots over all the horizontal axis, creating almost a constant line for these two categories). Then, audio files number 1 and 3 present higher occurrences of *door* and *brak* events, and they are also distributed across the audio files. In the second audio file, however, the events of those types are present mainly at the beginning, and there are just a few occurrences at the end. A remarkable fact is that, concretely, the few *door* events happening at the end of the second audio file (from minute 90 to minute 110), are the ones that present higher SNR values in that category.

Looking at the sizes of the dots, we can observe that even though the number of occurrences of *peop* and *rtm* seems to be constant, there is a considerable difference regarding the duration of these two events. In the three audio files, the *rtm* category has more occurrences of long events, and the SNR value is greater, too. This is consistent with the results previously observed in Figure 8.

6.4. Analysis of the Intermittency Ratio

As a final analysis, the intermittency ratio (IR) has been calculated and is presented in Figure 10. The intermittency ratio is a metric that was first introduced by Wunderli et al. in [20] and measures the “eventfulness” of a traffic environment. Concretely, it reflects the contributions of events that surpass a certain threshold to the total amount of energy in a certain period of time, measuring the impact on the total L_{Aeq} of all the individual loud events. Concerning the focus of the contribution of this work, with a standard metric, the IR supports the idea that the events detected and labeled (e.g., *rtm*) present a clear impact on the global value of the equivalent level measured in the street.

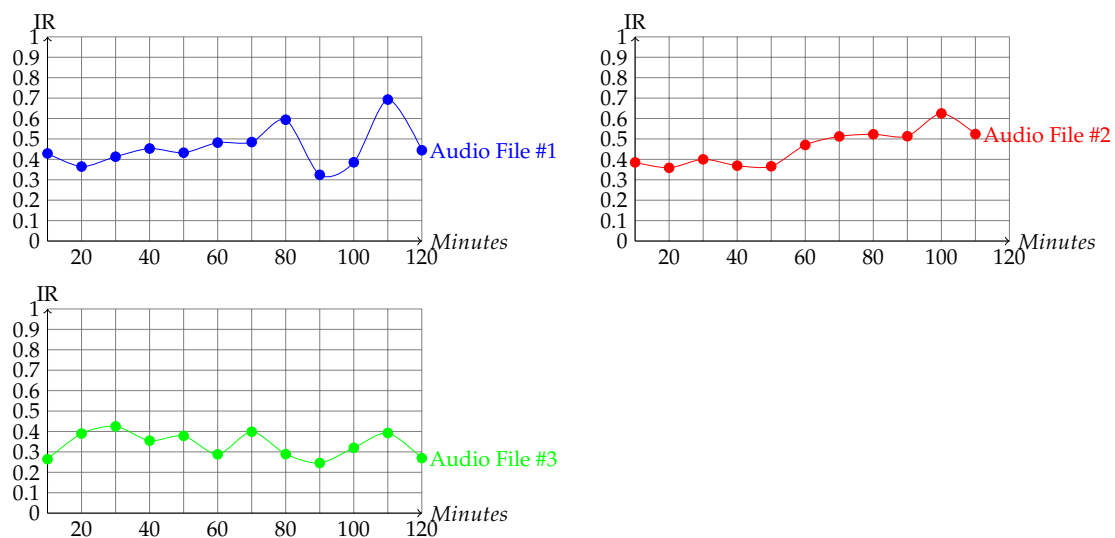


Figure 10. Intermittency Ratio of the three audio files presented in the dataset calculated with windows of 10 min. The y-axis represents the IR of each audio file and the x-axis represents the time evolution (in minutes) of the audio file that is being evaluated.

The procedure to calculate the ratio is as follows [19,20]: First, the equivalent level of energy of a window of size T is calculated as $L_{eq,T,tot}$. This is the amount of energy contained inside the window. For this study, we chose a window of 10 min, following the fact that the L_{Aeq} is mainly stationary, and we intend to define the impacts of the events with the shortest window frames possible to evaluate the differences in the axes of the time series. This trade-off time window allows us to have 12 IR values for each audio file (except in the case of Audio File #2, where we only have 11 values because the audio file is shorter). Then, the equivalent level of energy of each 1-s fragment inside the window was also calculated. In order to follow the methodology of the previous works [19,20] and to be able to obtain comparable results, those 1-s fragments presenting a L_{eq} greater than $L_{eq,T,tot} + 3$ dB were considered as “events”, independently of their labels in the dataset. Then, considering all the 1-s windows that surpassed the +3 dB threshold, the Heaviside step function was applied to remove the non-event sounds inside the 10-min window, and a new $L_{eq,T,events}$ was calculated to obtain the energy of only those 1-s fragments presenting a L_{eq} level greater than $L_{eq,T,tot} + 3$ dB. Finally, the ratio was calculated by dividing the $L_{eq,T,events}$ by $L_{eq,T,tot}$.

Summarizing, we used the next three equations to obtain the IR of each 10-min window of the three audio files of the dataset:

$$L_{eq,T,tot} = 10 \log_{10} \left(\frac{\sum_{n=0}^N X[n]^2}{N} \right) [\text{dB}], \quad (4)$$

where $X[n]$ are the samples of the audio file and N is the number of samples of a window (in our case, 10 min times the sampling frequency of the audio file).

$$L_{eq,T,events} = 10 \log_{10} \left(\frac{\sum_{n=0}^N H[X[n] - K] X[n]^2}{N} \right) [\text{dB}], \quad (5)$$

where $H[X[n] - K]$ is the Heaviside step function and K is the $L_{eq,T,tot}$ plus the threshold—in our case, set to 3 dB, as in [19].

$$IR = \left(\frac{10^{0.1 L_{eq,T,events}}}{10^{0.1 L_{eq,T,tot}}} \right) \quad (6)$$

To interpret the results of this ratio, we have to consider that, on the one hand, a ratio higher than 0.5 indicates that more than half of the energy of the signal is due to events (understanding events as parts of the signal with L_{eq} greater than K). This situation occurs when the events clearly stand out from the parts of the signal that are not considered as events (i.e., background noise), meaning that the ratio is high when the amount of energy of the events is significantly higher than the background noise. On the other hand, a small ratio value means that the amount of energy of the events is considerably low compared to the background noise. This situation occurs when the noise level of the events is close to the threshold. Hence, a small ratio in a street does not mean that the general noise level is lower than the noise level in a street with a higher ratio, but that the noise level of the events is similar to the background noise.

Evaluating the results in Figure 10, we can conclude that the three audio files presented in the dataset have an IR ranging from 0.24 (minute 90 from the Audio File #3) to 0.69 (minute 110 from Audio File #1). The relatively low values of IR (comparing them, for example, with the ones obtained from measurements in a local street in the work of Brambilla et al. in [19]), together with the impact values and number of events analyzed in previous sections (Figures 3 and 8), suggest that the street where the recordings took place is pretty noisy in terms of background noise, and the energy contribution is balanced between background noise and events. Nevertheless, there are several IR evaluations over 50%, so, at certain moments, mainly passes of *rtm* and other events related to leisure (e.g., *peop* shouting and others) can have a relevant contribution to the value of L_{Aeq} total of the acoustic file.

7. Materials

The labeled dataset can be downloaded from <https://doi.org/10.5281/zenodo.3956503>. The dataset is structured in six files: three audio files (.wav) and three label files (.txt). The two audio files recorded in the first campaign have been named *File-1.wav* and *File-2.wav*, and the audio file recorded in the second campaign has been named *File-3.wav*. The names of the label files belonging to each of the audio files follow the same naming scheme, adding a *_labels* at the end of the name (e.g., *File-1_labels.txt*).

8. Conclusions

This work has presented the creation and the analysis of a real-life environmental audio dataset in the district of Eixample, Barcelona. The dataset is composed of six hours of audio, was recorded on two Saturdays between 22:00 and 03:00, and contains 14 types of events, with a total of 6076 event occurrences. These events were classified into two categories: leisure and traffic, with the exception of the *rare* events, whose sources were not possible to determine or were a mix of sounds. The most common type of noise event is *people*, followed by road traffic noise, *brakes*, *doors*, and *rare* events. The fact that people and door events are among the most common indicates that the area and time chosen for the recording campaigns are suitable to measure the impact of leisure activities.

An SNR analysis comparing traffic noise with leisure noises revealed that traffic events have, on average, a higher value of SNR; *siren*, *horn*, and *road traffic noise* have the highest ones. An impact analysis suggested that the events with longer durations also had the highest impact. Both traffic and leisure had similar values of impact, but only a few had an increase in L_{Aeq} of greater than 0.01, meaning that they, in fact, contribute in a similar way to the noisiness of the environment.

The time event distribution indicates that the noises of people and traffic are constant during the recording, with the traffic noises being longer and confirming the greater SNR values observed previously. Finally, an analysis of the intermittency ratio shows that the recordings present low values of IR compared to other studies, which, together with the impact values and the number of events, indicates a noisy street with a balanced contribution of energy background noise and events over 3 dB, with some punctual exceptions, including loud *peop* or *rtn* street noise.

The dataset described in this paper is open and freely available to the community and may be used for different purposes. It can also be extended by means of further recordings or data augmentation, and also combined and compared with other datasets. A greater understanding of leisure and traffic events at night could help policymakers to regulate the noise produced in leisure locales, such as restaurants, bars, or discotheques. In particular, if an automatic sensor-based system is implemented to reliably distinguish and measure leisure activities, city councils would be able to continuously measure such noises, both in specific places in the city and during local festivities. Such information could inform urban planners and provide evidence to change the design of certain places in the city to improve the soundscape perceived by the neighbors.

Our future work is centered on validating the completeness of the dataset published. For this purpose, we plan to record another pair of days close to the locations of the other sensors in Figure 2, as pointed out by our colleagues from the Barcelona City Council. If more leisure events are detected in the new recordings, we would complete this corpus before proceeding to the event detection. Having more data points would also allow us to correlate the types of noises identified with the numbers of complaints in each area, which might give us some pointers for the types of noises that are most annoying to neighbors. In addition, given that the analysis focused on commonly used metrics in well-being- and health-related studies, we might be able to compare the characteristics of the noises observed with similar studies in other cities [8–11,20].

Furthermore, a deeper analysis of the impact of the labeled sounds will be conducted with a wider comparison between events belonging to the same category in order to determine differences between the impacts of each event, with special focus on the surrounding environmental noise, which is a key issue for the evaluation of SNR and impact. Once these analyses are conducted, the *rare* category has

to be vertically analyzed in order to determine which types of events usually correspond to that fuzzy label; they cannot be classified inside any of the other categories, but maybe we can make more acoustic information about all the *rare* events available. Finally, we plan to train an ML system (similar to [15]) in order to automatically classify noise events at night, taking into account at least leisure and traffic.

Author Contributions: All authors have significantly contributed to this work. E.V.-V. contributed to the tagging of the recordings and writing and carried out the data analyses. L.D. was involved in the project conceptualization and coordination, as well as writing. R.M.A.-P. participated in the tagging and writing, and also offered conceptual and technical support. F.P. and H.V. carried out the recording campaigns, participated in the tagging, and carried out some preliminary data analyses. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 712949 (TECNIOspring PLUS) and from the Agency for Business Competitiveness of the Government of Catalonia. The authors thank the Secretaria d’Universitats i Recerca del Departament d’Economia i Coneixement (Generalitat de Catalunya) under grant 2017 SGR 966 and the Ramon Llull University (ref. 2020.URL-Proj-053).

Acknowledgments: The authors would like to thank our colleagues from the Environmental Department of Barcelona City Council for the guidance in this work, especially to Julia Camps.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANEs	Anomalous Noise Events
ANED	Anomalous Noise Event Detector
AI	Artificial Intelligence
CNOSSOS-EU	Common Noise Assessment Methods in Europe
END	Environmental Noise Directive 2002/49/EC
IR	Intermittency Ratio
SNR	Signal-to-Noise Ratio
WASN	Wireless Acoustic Sensor Networks

References

- Ritchie, H.; Roser, M. Urbanization. Our World in Data. 2020. Available online: <https://ourworldindata.org/urbanization> (accessed on 25 June 2020).
- Observatori del Turisme a Barcelona: Ciutat i Regió. Informe de l’Actividad Turística 2019—Capsula 1. 2019. Available online: <https://www.observatoriturisme.barcelona/> (accessed on 25 June 2020).
- European Environment Agency. The European Environment—State and Outlook 2020. Available online: <https://www.eea.europa.eu/soer/2020> (accessed on 1 April 2020).
- Cik, M.; Lienhart, M.; Lercher, P. Analysis of Psychoacoustic and Vibration-Related Parameters to Track the Reasons for Health Complaints after the Introduction of New Tramways. *Appl. Sci.* **2016**, *6*, 398. [CrossRef]
- Clark, W.W. Noise exposure from leisure activities: A review. *J. Acoust. Soc. Am.* **1991**. [CrossRef] [PubMed]
- Jarup, L.; Babisch, W.; Houthuijs, D.; Pershagen, G.; Katsouyanni, K.; Cadum, E.; Dudley, M.L.; Savigny, P.; Seiffert, I.; Swart, W.; et al. Hypertension and Exposure to Noise Near Airports: The HYENA Study. *Environ. Health Perspect.* **2008**, *116*, 329–333. [CrossRef] [PubMed]
- Basner, M.; Müller, U.; Elmenhorst, E.M. Single and Combined Effects of Air, Road, and Rail Traffic Noise on Sleep and Recuperation. *Sleep* **2011**, *34*, 11–23. [CrossRef]
- Dratva, J.; Phuleria, H.C.; Foraster, M.; Gaspoz, J.M.; Keidel, D.; Künzli, N.; Liu, L.J.S.; Pons, M.; Zemp, E.; Gerbase, M.W.; et al. Transportation Noise and Blood Pressure in a Population-Based Sample of Adults. *Environ. Health Perspect.* **2012**, *120*, 50–55. [CrossRef]
- Hofman, W.; Kumar, A.; Tulen, J. Cardiac reactivity to traffic noise during sleep in man. *J. Sound Vib.* **1995**, *179*, 577–589. [CrossRef]
- Ottoz, E.; Rizzi, L.; Nastasi, F. Recreational noise: Impact and costs for annoyed residents in Milan and Turin. *Appl. Acoust.* **2018**, *133*, 173–181. [CrossRef]

11. Easteal, M.; Bannister, S.; Kang, J.; Aletta, F.; Lavia, L.; Witchel, H. Urban Sound Planning in Brighton and Hove. *Forum Acusticum Conf.* **2014**. [CrossRef]
12. World Health Organization. Environmental Noise Guidelines for the European Region. 2018. Available online: <https://www.euro.who.int/en/publications/abstracts/environmental-noise-guidelines-for-the-european-region-2018> (accessed on 31 July 2020).
13. Cox, P.; Palou, J. Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise—Declaration by the Commission in the Conciliation Committee on the Directive Relating to the Assessment and Management of Environmental Noise (END). *Annex I OJ* **2002**, 189, 2002.
14. European Commission, Joint Research Centre—Institute for Health and Consumer Protection. *Common Noise Assessment Methods in Europe (CNOSSOS-EU) for Strategic Noise Mapping Following Environmental Noise Directive 2002/49/EC*; European Commission Joint Research Centre: Ispra, Italy, 2012.
15. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323. [CrossRef]
16. Alías, F.; Alsina-Pagès, R.M. Review of Wireless Acoustic Sensor Networks for Environmental Noise Monitoring in Smart Cities. *J. Sens.* **2019**, *2019*, 7634860. [CrossRef]
17. Nakajima, Y.; Sunohara, M.; Naito, T.; Sunago, N.; Ohshima, T.; Ono, N. DNN-based Environmental Sound Recognition with Real-recorded and Artificially-mixed Training Data. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (InterNoise 2016), Hamburg, Germany, 21–24 August 2016; German Acoustical Society (DEGA): Hamburg, Germany, 2016; pp. 3164–3173.
18. Salamon, J.; Bello, J.P. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [CrossRef]
19. Brambilla, G.; Confalonieri, C.; Benocci, R. Application of the intermittency ratio metric for the classification of urban sites based on road traffic noise events. *Sensors* **2019**, *19*, 5136. [CrossRef] [PubMed]
20. Wunderli, J.M.; Pieren, R.; Habermacher, M.; Vienneau, D.; Cajochen, C.; Probst-Hensch, N.; Rössli, M.; Brink, M. Intermittency ratio: A metric reflecting short-term temporal variations of transportation noise exposure. *Environ. Epidemiol.* **2016**, *26*, 575–585. [CrossRef] [PubMed]
21. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146. [CrossRef]
22. Valero, X.; Alías, F. Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys. *Arch. Acoust.* **2012**, *37*, 423–434. [CrossRef]
23. Heittola, T.; Mesaros, A.; Eronen, A.; Virtanen, T. Context-dependent sound event detection. *EURASIP J. Audio Speech Music. Process.* **2013**, *2013*, 1–13. [CrossRef]
24. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recognit. Lett.* **2015**, *65*, 22–28. [CrossRef]
25. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Audio Surveillance of Roads: A System for Detecting Anomalous Sounds. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 279–288. [CrossRef]
26. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189. [CrossRef]
27. Alsina-Pagès, R.M.; Alías, F.; Socoró, J.C.; Orga, F. Detection of Anomalous Noise Events on Low-Capacity Acoustic Nodes for Dynamic Road Traffic Noise Mapping within an Hybrid WASN. *Sensors* **2018**, *18*, 1272. [CrossRef] [PubMed]
28. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E.; Bellucci, P. The LIFE DYNAMAP project: Towards a procedure for dynamic noise mapping in urban areas. *Appl. Acoust.* **2017**, *124*, 52–60. [CrossRef]
29. Bellucci, P.; Peruzzi, L.; Zambon, G. LIFE DYNAMAP project: The case study of Rome. *Appl. Acoust.* **2017**, *117*, 193–206. [CrossRef]
30. Alsina-Pagès, R.M.; Orga, F.; Alías, F.; Socoró, J.C. A WASN-Based Suburban Dataset for Anomalous Noise Event Detection on Dynamic Road-Traffic Noise Mapping. *Sensors* **2019**, *19*, 2480. [CrossRef]
31. Alías, F.; Orga, F.; Alsina-Pagès, R.M.; Socoró, J.C. Aggregate Impact of Anomalous Noise Events on the WASN-Based Computation of Road Traffic Noise Levels in Urban and Suburban Environments. *Sensors* **2020**, *20*, 609. [CrossRef]

32. Bello, J.P.; Silva, C.; Nov, O.; Dubois, R.L.; Arora, A.; Salamon, J.; Mydlarz, C.; Doraiswamy, H. SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution. *Commun. ACM* **2019**, *62*, 68–77. [[CrossRef](#)]
33. Cartwright, M.; Mendez, A.E.M.; Cramer, J.; Lostanlen, V.; Dove, G.; Wu, H.H.; Salamon, J.; Nov, O.; Bello, J. Sonyc urban sound tagging (sonyc-ust): A multilabel dataset from an urban acoustic sensor network. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York, NY, USA, 25–26 October 2019; New York University: New York, NY, USA, 2019; pp. 35–39.
34. Salamon, J.; Jacoby, C.; Bello, J.P. A dataset and taxonomy for urban sound research. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; ACM: Orlando, FL, USA, 2014; pp. 1041–1044.
35. Mesaros, A.; Heittola, T.; Virtanen, T. Acoustic scene classification in DCASE 2019 challenge: Closed and open set classification and data mismatch setups. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York, NY, USA, 25–26 October 2014; New York University: New York, NY, USA, 2019; pp. 164–168.
36. Mesaros, A.; Heittola, T.; Virtanen, T. TUT database for acoustic scene classification and sound event detection. In Proceedings of the 24th European Signal Processing Conference (EUSIPCO 2016), Budapest, Hungary, 29 August–2 September 2016; IEEE: Budapest, Hungary, 2016; Volume 2016, pp. 1128–1132. [[CrossRef](#)]
37. Zoom Corporation. *H5 Handy Recorder—Operation Manual*; Zoom Corporation: Tokyo, Japan, 2014.
38. Jarup, L.; Dudley, M.L.; Babisch, W.; Houthuijs, D.; Swart, W.; Pershagen, G.; Bluhm, G.; Katsouyanni, K.; Velonakis, M.; Cadum, E.; et al. Hypertension and exposure to noise near airports (HYENA): Study design and noise exposure assessment. *Environ. Health Perspect.* **2005**, *113*, 1473–1478. [[CrossRef](#)]
39. Orga, F.; Alías, F.; Alsina-Pagès, R. On the Impact of Anomalous Noise Events on Road Traffic Noise Mapping in Urban and Suburban Environments. *Int. J. Environ. Res. Public Health* **2017**, *15*, 13. [[CrossRef](#)]
40. St Pierre, R.L., Jr.; Maguire, D.J. The impact of A-weighting sound pressure level measurements during the evaluation of noise exposure. In Proceedings of the Noise-Con 04: The 2004 National Conference on Noise Control Engineering, Baltimore, MD, USA, 12–14 July 2004; pp. 12–14.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).