electronics

MDPI

Article

Low-Cost Distributed Acoustic Sensor Network for Real-Time Urban Sound Monitoring

Ester Vidaña-Vila [1,*,†], Joan Navarro [2,†], Cristina Borda-Fortuny [2,†] and Dan Stowell [3] and Rosa Ma Alsina-Pagès [1,†]

1 GTM—Grup de Recerca en Tecnologies Mèdia, 08022 Barcelona, Spain; rosamaria.alsina@salle.url.edu
2 GRITS—Grup de Recerca en Internet Techologies and Storage, 08022 Barcelona, Spain; jnavarro@salleurl.edu (J.N.); cristina.borda@salleurl.edu (C.B.-F.)
3 Machine Listening Lab, Centre for Digital Music, Queen Mary University of London, London E1 4NS, UK
* Correspondence: ester.vidana@salle.url.edu; Tel.: +34-932902400
† La Salle, Universitat Ramon Llull. c/Quatre Camins, 30, 08022 Barcelona, Spain.

Received: 2 November 2020; Accepted: 7 December 2020; Published: 11 December 2020

Abstract: Continuous exposure to urban noise has been found to be one of the major threats to citizens' health. In this regard, several organizations are devoting huge efforts to designing new in-field systems to identify the acoustic sources of these threats to protect those citizens at risk. Typically, these prototype systems are composed of expensive components that limit their large-scale deployment and thus reduce the scope of their measurements. This paper aims to present a highly scalable low-cost distributed infrastructure that features a ubiquitous acoustic sensor network to monitor urban sounds. It takes advantage of (1) low-cost microphones deployed in a redundant topology to improve their individual performance when identifying the sound source, (2) a deep-learning algorithm for sound recognition, (3) a distributed data-processing middleware to reach consensus on the sound identification, and (4) a custom planar antenna with an almost isotropic radiation pattern for the proper node communication. This enables practitioners to acoustically populate urban spaces and provide a reliable view of noises occurring in real time. The city of Barcelona (Spain) and the UrbanSound8K dataset have been selected to analytically validate the proposed approach. Results obtained in laboratory tests endorse the feasibility of this proposal.

Keywords: ubiquitous acoustic sensor network; distributed consensus; acoustic propagation; noise management; isotropic radiation pattern planar antenna; acoustic event detection; signal processing

1. Introduction

Research dating back to the last century [1] has acknowledged that continuous exposure to high levels of noise is harmful for human beings, as recently highlighted by the World Health Organization (WHO) [2]. For instance, noise can negatively affect sleep quality [3], induce chronic effects on the nervous sympathetic system [4], or even cause psycho-physiological effects such as annoyance, reduced performance or aggressive behavior [5]. In this context, noise is often defined as a type of unwanted and/or harmful sound that disturbs communication between individuals [5,6], i.e., the overall acoustic energy measured in Sound Pressure Levels (SPLs) exceeds a predefined limit [7].

Accordingly, several agencies and public departments (e.g., NSW Environment Protection Agency, NYC Department of Environmental Protection, European Commission) have defined regulations [7] to limit the amount of noise (i.e., equivalent averaged level *LAeq*) that the population can be exposed to. For instance, the WHO recommends that noise must be below 35 dBA in classrooms to enable good teaching and learning conditions, or below 30 dBA in bedrooms to enable good quality sleep [8]. Most of these regulations define the maximum level of noise allowed in a specific

scenario (e.g., home buildings, factories, schools) and a specific acoustic source (e.g., motor vehicles, air conditioners, machinery, water heaters, etc.) [9]. However, such a standard way of defining and regulating noise faces two important challenges when applied and enforced in the real world [7]: acoustic source isolation and identification and practical on-field noise measurement for automatic acoustic surveillance:

1. It is very difficult to isolate and identify a specific noise source from the overall acoustic landscape since the aforementioned SPL measurements aggregate the energy level from all the acoustic sources at the same time [6]. Indeed, in a real-world environment, several different acoustic sources may emerge over time and, thus, the definition of a fixed SPL threshold for a given area is not always appropriate [10], i.e., the acoustic threshold should be dynamic according to the sound (noise) that is currently occurring. Unfortunately, the SPL value *per se* does not provide enough practical information to facilitate the identification of the sound (noise) source(s) [10], which complicates the task of verifying whether an acoustic landscape meets the local regulations or not.
2. Also, effectively measuring the amount of noise in large-scale environments (e.g., urban areas) requires a considerable number of resources in terms of highly qualified professionals—it has been reported that the NYC Department of Environmental Protection has up to 50 professionals designated to dealing with noise complaints in the city of New York (despite this, their average response time is still about 5 days) [6]—and expensive equipment [6]. Indeed, this equipment can range from $1500 up to $20,000 depending on the type, measurement range, and capability of the microphone to produce noise spectral data [11]. Therefore, conducting scalable, long-term (i.e., conducting measurements 24 h a day 365 days a year) noise surveillance tasks in wide span areas has emerged as a hot research topic in recent years.

Over the last decade, Ubiquitous Sensor Networks (USNs) [12] have emerged as a powerful alternative to address the challenges of scalable and cost-effective [13] sensing in large-scale areas [14]. The benefits of USNs have been exploited in several domains, ranging from water pollution monitoring [15] to smart agriculture [13], including Wireless Acoustic Sensor Networks (WASNs) for Ambient Assisted Living [16]. Indeed, USNs provide a design reference to conceive versatile architectures able to interconnect a high number of devices—typically with limited capabilities in terms of storage, computation and communications—while providing fault tolerance and robustness with the aim of increasing the performance of individual sensors [15,17,18]. Indeed, the idea of using an interconnected set of inexpensive commodity hardware to beat the performance of individual high-end devices is well known in the literature of distributed systems and has been massively exploited—the Google File System [19] being one of its most representative examples.

This work aims to extrapolate this idea to the field of urban sound monitoring, i.e., the use of a set of low-cost microphones deployed in a redundant topology—being the sensing layer [12] of an ubiquitous sensor network that will later provide them with additional storage and computing features—to *listen* to events from large-scale areas in a cost-effective way while obtaining a reasonable accuracy. Hence, the modest performance of the low-cost microphones can be compensated by the robustness of the computing algorithms running on top of the ubiquitous sensor network [20]. Therefore, the purpose of this paper is to propose a low-cost cost distributed acoustic sensor network for real-time urban sound monitoring in large-scale scenarios. More specifically, the proposed approach aims to present a network composed of inexpensive hardware (i.e., Raspberry Pi Model 2B (RPi) [21]) in which each node is conceived to (1) process a real-time audio stream from a directly connected low-cost microphone, (2) locally identify the occurring events in this audio stream by means of a deep neural network, (3) communicate the identified events to the neighboring nodes of the network by means of a custom planar antenna with almost isotropic radiation, and (4) globally validate these locally discovered events by means of a distributed consensus protocol.

To sum up, the main contributions of this work are:

- A deep-learning algorithm for urban sound identification in real time to be deployed in low-cost devices with modest computing and storage capabilities.
- A custom planar antenna with almost isotropic radiation pattern for robust and low-energy consumption communications between nodes.
- A distributed consensus protocol to compare the detection results of each individual node with its neighboring nodes.

To further validate the proposed approach, we evaluated automatic recognition against the UrbanSound8K dataset [22] as a source of typical urban audio events and selected the city of Barcelona (Spain) as a reference model to deploy the proposed system. Indeed, Barcelona was designed following a particular square block grid (see Figure 1) that makes it an ideal scenario to deploy urban ubiquitous sensor networks. However, current noise surveillance initiatives in Barcelona only focus on sound pressure levels and span an average area of 1 square kilometer per sensor. This work aims to enrich the measurements by identifying the sound source and providing a fine-grained analysis of their location. The evaluation of the proposed system has been done as follows: (1) the communication antenna has been validated by means of simulation, and (2) the acoustic recognition together with the distributed consensus protocol have been validated by means of laboratory testing rather than full real-world deployment, planned for future work. The regularly defined urban grid of Barcelona greatly facilitates this aspect of spatial modelling.

Figure 1. Aerial view of the urban grid structure of the city of Barcelona [23].

The remainder of this paper is organized as follows. Section 2 reviews the related work on acoustic sensor networks for environmental noise monitoring. Section 3 details the proposed system architecture and details its three layers: data processing, distributed consensus, and communications. Section 4 presents the conducted experimental evaluation. Section 5 discusses the obtained results. Finally, Section 6 concludes the paper and proposes some future work directions.

2. Related Work

In this section, we describe related works to the main WASN-based approaches developed in recent years to monitor environmental noise. The main goal of these networks is to collect the L_{Aeq} levels alone or together with extra information obtained in each node. In some situations, this extra information gathered in each node corresponds to data about the sound source measured in each sensor.

2.1. WASNs to Monitor the Noise Levels

Most of the WASNs in this first category use commercial sound level meters as sensor nodes. These devices are usually connected to a central server of the WASN, which collects all the L_{Aeq} information collected by the nodes. Projects such as Telos [24], which correspond to one of the first experiences in this WASN design by means of an ultra-low power wireless sensor module designed by the University of California (Berkeley). Some years after that experience, Santini et al. in [25,26] showed how a WASN can be used in a wide variety of environmental monitoring applications, with a special focus on urban noise.

More recent projects include the deployment of a network to monitor the traffic noise in Xiamen City (China) for environmental purposes [27]. The project covers 35 roads in 9 green spaces in the city, and the scientists use the data from the monitoring stations to model the traffic of 100 other roads in the city. The deployment included noise level meters, with ZigBee and GPRS communications.

The FI-Sonic Project is focused on continuous noise monitoring surveillance [28]; the main goal is to develop the technology required to process urban sounds by means of artificial intelligence, enabling the generation of noise maps but also the identification and location of groups of sound events [29]. It is based on a FIWARE platform (https://www.fiware.org/). Finally, the RUMEUR project (Urban Network of Measurement of the sound Environment of Regional Use) is based on a hybrid wireless sensor network deployed by BruitParif [30] in Paris and its surroundings. The network has high accuracy on monitoring critical places (for example, airports) but also uses other less precise measuring equipment, whose final goal is to evaluate the equivalent noise level of the environment. The RUMEUR project has evolved to Medusa [31], a system that combines four microphones and two optical systems so that noise levels can be represented on a 360° image of the environment, by means of the identification of the source location. Its computational load is high, and it cannot be resolved by most of the low-cost acoustic sensor systems.

The Barcelona Noise Monitoring Network (NMN) was described in [32] and reviewed in [33]. The network is designed to reduce the impact of urban infrastructures on the environment in the city of Barcelona. The results of the analysis carried out in [33] suggest that both the costs and the number of manual tasks carried out by technicians should be reduced. In Barcelona, several other initiatives to empower the citizens of critical urban areas, such as Plaça del Sol [34], have also been developed, but so far they have been only able to complement the measurements conducted by the calibrated sensors deployed by the City Council.

2.2. WASNs Based on Ad-Hoc Designed Nodes

To satisfy the increasing demand of an automatic monitoring of noise levels in urban areas, as described in [35], several WASN-based projects are being developed in different countries and designed, then deployed ad-hoc for their application; some of these projects include other environmental measurements used to determine other aspects of citizens' quality of life besides noise pollution.

The CENSE project (Characterization of urban sound environments) focuses on the design of noise maps in France [36]. It integrates both simulated and measured data by means of a wide network of low-cost sensors. The project includes environmental acoustics, statistics, Graphical Information System (GIS) to plot the results, as well as network sensor design, signal processing and the proposal of the production of perceptive noise maps. The IDEA project (Intelligent Distributed Environmental Assessment) [37] focuses on noise and air quality pollution in several urban areas of Belgium. It integrates a sensor network based on a cloud platform, and it measures noise and air quality [38]. The MESSAGE project (Mobile Environmental Sensing System Across Grid Environments) [39] not only monitors noise, carbon monoxide, nitrogen dioxide, temperature, but also humidity and traffic occupancy, and it gives real-time noise data information in the United Kingdom. The UrbanSense project [40] and the MONZA project [41] follow both the idea of monitoring urban noise real-time together with other air pollutants; UrbanSense in Canada and MONZA in the Italian city of Monza.

The urban acoustic environment of New York City is monitored using a low-cost static acoustic sensor network in the framework of a project named SONYC project (Sounds of New York City) [11]. The goal of this project is to describe the acoustic environment while monitoring noise pollution. It collects longitudinal urban acoustic data, to process them and have generous sampling to work with acoustic event detection [6].

Another interesting approach of the monitoring network projects is the hybrid approach of crossing the acoustic information with subjective perception surveys, to consider the typology of the events in relationship with sleep quality [42]. A sound recognition system is applied to provide information about the detected sounds and establish a relationship between the perception surveys and the identified events related to road traffic noise [43]. However, this project is only aimed at the identification of the acoustic events and their perception, it has no impact on any noise maps generation.

The DYNAMAP project [44] achieves a good trade-off between cost and accuracy in the design of a WASN. The project deployed two pilot areas in Italy, located in Rome [45] and Milan [46], so as to evaluate the noise impact of road infrastructures in suburban and urban areas, respectively. The two WASNs monitor road traffic noise reliably collecting data at 44.1kHz to remove specific audio events, which are unrelated to road traffic [47,48] for the noise map computation [49]. Based on their experience in this project and particularly the time and effort spent transforming the original prototyping code into real operable language, the team developed a low-cost flexible acoustic sensor for rapid real-time algorithm development and testing [50].

3. System Architecture

This section details the proposed system architecture and further elaborates on the rationales to implement the ubiquitous acoustic sensor network for urban sound identification. The main constraints [14] and design guidelines that have driven the conception of this distributed system are the following:

Cost affordability. The system must be conceived to cover large-scale areas (i.e., hundreds of km^2) in a redundant topology [20] (i.e., at least 4 nodes per city block (As a matter of reference, in Barcelona city (Spain) the sides of the blocks measure around 110 m on average). Therefore, the individual cost of each of the nodes that articulate the ubiquitous sensor network must be kept as low as possible. This prevents us from using expensive high performance computing devices (e.g., GPUs [16]) and leads us to consider alternative solutions with more modest computing and storage features.

Physical distance between neighboring nodes. As individual nodes must be composed of inexpensive hardware—in terms of both acoustics and computing— they need to take advantage of each other to provide robust answers and good performance [20]. In this regard, each node will need to constantly communicate with its neighbors to check, compare, and validate the identified acoustic events. Therefore, there is a trade-off on the physical distance between nodes: one the one hand it must be kept low so that an event can be *heard* by more than one node, and on the other hand, the larger distance, the more area will be covered.

Real-world deployment. The system must be deployed in urban spaces, which makes it vulnerable to extreme weather conditions (e.g., heat, cold, wind, rain), vandalism, or theft [14]. Therefore, the nodes that compose the proposed USN must be as small as possible so they can be installed in existing street furniture (e.g., traffic lights [51]). Also, the power consumption of each node must be low to facilitate its integration. This means that the proposed approach will need to be efficient both in terms of communications (i.e., exchanging little data among nodes) and of computing (i.e., using as low computing resources as possible to obtain the maximum event identification accuracy).

Fault tolerance and recovery. Since the nodes of the system will be exposed to harsh environmental conditions and given the difficulty of physically accessing them to conduct maintenance and

reparation duties (e.g., reboot), the nodes must be self-managed, i.e., a node must keep operating even in case of failure of their neighboring nodes.

Acoustic quality. The nodes must be capable of acquiring and processing data at a minimum sample rate of 22,050 samples per second (to be able to analyze frequency information ranging from 0 to 11,025 Hz) and a depth of 16 bits per sample. Before deployment, the microphones must be calibrated, and the gain must be adjusted so all the microphones of the USN capture similar signal levels when exposed to the same sounds.

To meet all these requirements, we propose the use of a Raspberry Pi device augmented with acoustic and communications capability. We select the Raspberry Pi Model 2B [21] that has a 900 MHz quad-core ARM Cortex-A7 processor, 1 GiB RAM, and an average power consumption of 200 mA. An external USB microphone for acoustic data processing and a custom communications antenna for data exchanging among nodes must be plugged to the RPi (see Figure 2). The remainder of this section (1) describes the proposed acoustic data-processing framework to locally identify acoustic events in urban areas, (2) details and justifies the design of a communications module (modem and custom antenna) to enable data communications among nodes, and (3) presents a distributed consensus protocol aimed at comparing the locally identified acoustic events to obtain robust and global-scope acoustic findings.

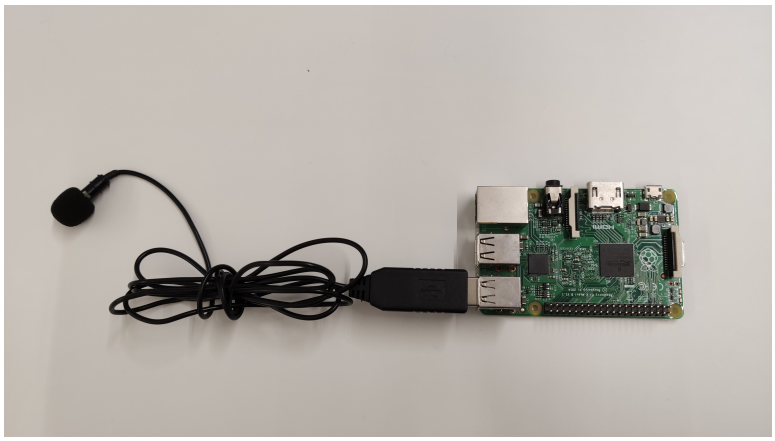


Figure 2. Raspberry Pi Model 2B with USB microphone.

3.1. Data Processing

For data acquisition and processing, a low-cost omnidirectional electret USB microphone is used. The reference number for the microphone is OUT-AMLO-0872 and it is manufactured by Seacue. The frequency response is almost flat for the frequency range where the events are taking place (50 Hz–10 KHz), meaning that it does not generate *colorations* (i.e., alterations or distortions) on that frequencies. The price is as low as 12 EUR and it is plug-and-play, meaning that there is no need for an external Analog-to-Digital converter (ADC). Once one window (audio fragment of a certain duration) of audio is captured, the spectrogram is calculated and fed to the neural network. Using spectrograms as input features for the network is a technique that has been proven to be effective for sound classification tasks [52], since they provide information about acoustic energy in both frequency and time. The selected network architecture is a Convolutional Neural Network (CNN). The reason behind using a CNN lies in the fact that they typically require storing fewer parameters than traditional deep neural networks, which reduces the model size [53]. Moreover, CNNs have been extensively validated for sound event detection [54].

The output result of this data-processing layer is an events vector with as many components as acoustic event types (i.e., classes), where each component is a value between 0 and 1 representing the probability of the event belonging to that class. For instance, in the case of UrbanSound8K dataset being used in this work, there are 10 event types.

This resulting vector will be sent to the neighboring nodes using the communications antenna and the distributed consensus protocol that is detailed in the following sections.

3.2. Communications

For inter-communication between neighboring nodes, a custom bespoke antenna has been designed to achieve higher specifications with the limited physical space available on the RPi. The performance of the whole communications system is calculated using the Friis Transmission Equation [55]. This equation states that the signal received by the communications module (i.e., antenna plus transceiver) is calculated and compared to the noise level, giving a Signal-to-Noise Ratio (SNR). If the SNR is high enough, the signal will be successfully decoded. The Friis Transmission Equation shows that losses increase with frequency, and higher power is lost at higher frequencies. Therefore, the very first design constraint to be addressed is the operating frequency.

The 2.4 GHz band (UN-51) [56–58] (i.e., Wi-Fi) is a convenient choice for communications in USNs [14]. However, this band is often absorbed by structural elements, such as walls and floors or ceilings and it also coincides with the resonant frequency of water, which makes it inappropriate for urban spaces. In addition, such a high frequency limits the communication range of each node [55]. Alternatively, the 433 MHz band (UN-30) is slightly better than the 2.4 GHz band for the range—as it operates at a lower frequency—but it does not guarantee a secure data transmission, due to a lot of interference in this specific frequency band, such as remote controls and parking remote controls which can produce high levels of interference. The frequency band of 868 MHz is an ISM band designated by the UN-39 in Spain [56–58] that offers a better range than the 2.4 GHz band, increased by 2–3 times, and is less populated than the 433 MHz band. This frequency band is used by LoRa, Zigbee and Sigfox in ITU region 1 (Europe) [59]. In case of the system being used in other regions, such as the US, the ITU-RR-5.150 specifies a band in 915 MHz in the ITU region 2. In this case, the communication system would need to be accordingly updated and fine-tuned with the new requirements to radiate at the specific frequency band for the new region.

To summarize, transmission in the 868 MHz band is (1) able to penetrate obstructions in the line-of-sight and (2) suitable for connecting medium and long-distance remote monitoring systems. However, it presents limited maximum data rates compared to other bands. As in the proposed large-scale urban sound monitoring use-case, low data rates (a few kbps) for medium range are enough to transmit the vector with the classification results (see Section 3.1), there is no need to use a higher frequency band.

After selecting the 868 MHz operating band, a transceiver for this frequency to be attached to the RPi is required. There are many off-the-shelf communication modules available in the market for RPi—which is in fact one of the advantages of using this device. For a very low price there are numerous modules to radiate at the frequency band of 868 MHz, some with a short range and others with medium range. For example, the ENOCEAN PI 868, RTX-868-FSK and SX1272 [60,61]. The SX1272 module for RPi operates at 868 MHz but uses a simple monopole antenna. The monopole antenna is troublesome as it could lead to null communication in certain directions. An isotropic antenna would best fit the requirements for this project. The design of a custom bespoke antenna to be connected to the SX1272 module is presented below.

As shown in Figure 3, the antenna is designed with two planar crossed dipoles in a low-cost FR-4 substrate to present an isotropic pattern. Consequently, the same signal will be received at the receiver due to its isotropic properties, regardless of the orientation of the antenna. Therefore, there is no risk that communication will be lost when the sensor is in certain orientations. The selected design is based on an isotropic Crossed Dipoles designed for a higher frequency band [62]. The proposed design must be optimized to operate at the 868 MHz frequency band and to fit the RPi case. The planar crossed dipoles placed on top of a low-cost FR-4 substrate with 1 mm thickness and relative permittivity of 4.4, are a low-cost solution to fit in the RPi Shield and provide isotropic radiation. The crossed dipoles are fed at a 90° phase shift to achieve isotropic radiation [62].

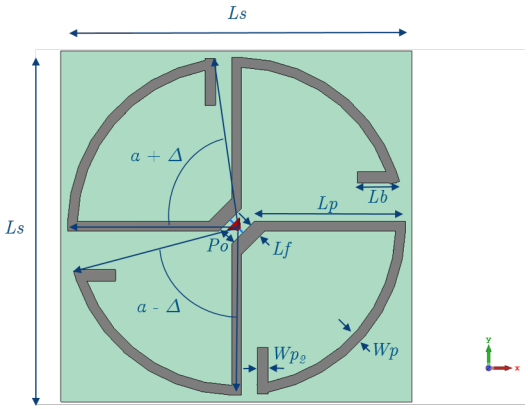


Figure 3. Proposed Planar Crossed Dipoles for isotropic radiation.

To optimize the antenna parameters and achieve the aforementioned requirements of the communication system (operating frequency, isotropic radiation and size constraints), CST Microwave Studio has been used to conduct the parametric studies depicted in Figure 4:

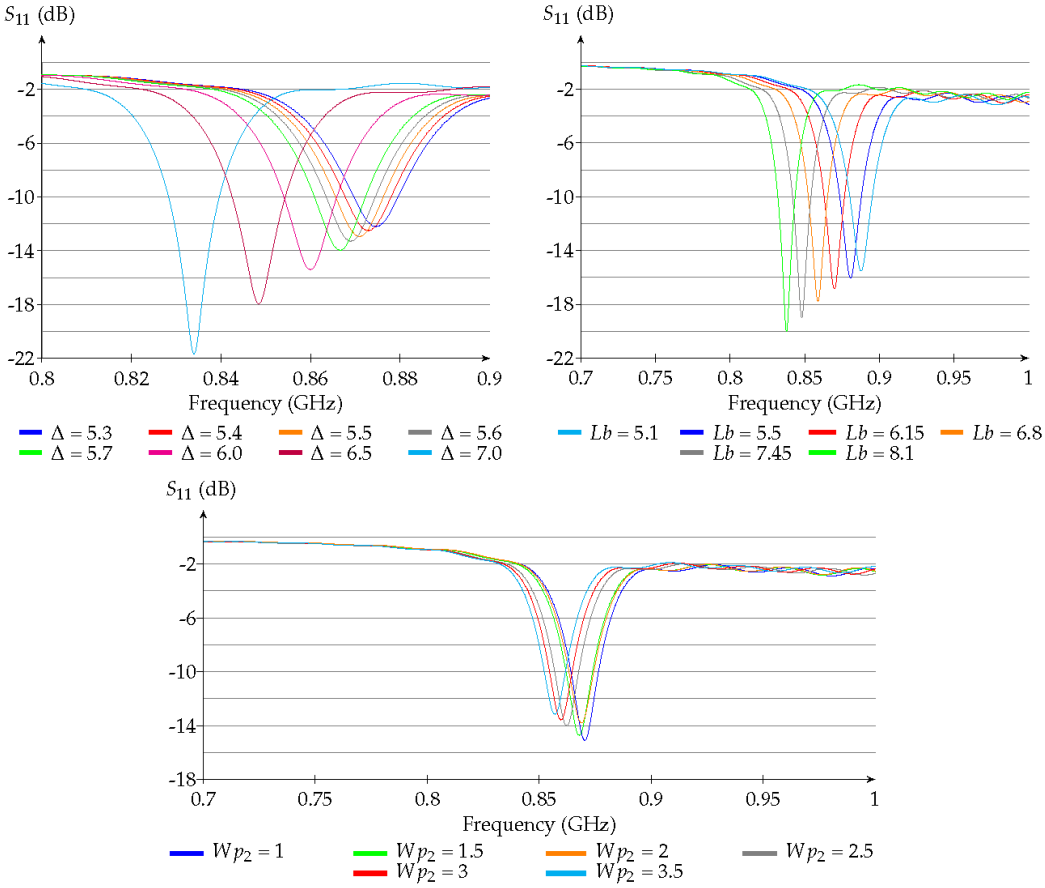


Figure 4. Parametric study of the reflection coefficient S_{11}. Top left: changing the Δ. Top right: changing the branch length (*Lb*). Bottom: changing the branch width (Wp_2).

1. Delta (Δ) is the variation between the length of the horizontal dipole and the length of the vertical dipole. By varying the parameter Δ, the current of the two crossed dipoles can be excited at the same magnitude and 90 degrees phase shift, which is only in this conditions that isotropic radiation can be achieved [62]. The top left plot in Figure 4 shows that when the 90 degrees phase shift between dipoles is achieved the operating frequency is better matched

and, thus, S_{11} parameter becomes lower. A Δ of 7.0° is selected because it better matches the input impedance of the antenna and it presents isotropic radiation. Once isotropic radiation is achieved, the input impedance can be matched at the operating frequency by varying the dipole length and width as shown in the next steps.

2. The length of each dipole branch is a crucial parameter to match the antenna at 868 MHz frequency band. The branch length (Lb) is the length at the end which will be longer for lower frequencies or shorter if the antenna needs to operate at higher frequencies. Lb is added to the design to match the input impedance of the antenna at the desired operating frequency of 868 MHz, considering the size of the RPi cannot accommodate long-enough dipoles to be resonant at this frequency. Otherwise, antenna efficiency would be reduced at the required operating frequency band. By making the dipole branches longer a lower frequency can be matched. As expected, the resonant frequency decreases as the length of the dipole increases. Therefore, the desired operating frequency can be achieved by varying this parameter.

   Lb is the same for both dipoles, as the only difference in length comes from the parameter Δ, mentioned before. Δ is used to achieve isotropic radiation, Lb adjusts the matched input impedance so that the antenna operates at the required operating band.

   The top right plot in Figure 4 shows the effect of varying the Lb parameter in the reflection coefficient of the antenna (S_{11} parameter).

3. Finally, the length of the 2 dipole branches is determined, although the width will also impact the operating frequency of the antenna, as shown in Figure 4 (bottom). A parametric study of the $Wp2$ is used to fine-tune the value of this parameter and match the antenna at exactly 868 MHz with a value of 1.5 mm.

Combining the results of these studies, the optimized parameters of the proposed antenna are presented in Table 1:

Table 1. Values of the Optimized design parameters for the antenna geometry.

Ls (mm)	Lp (mm)	Lb (mm)	Lf (mm)	Po (mm)	Wp (mm)	Wp_2 (mm)	α	Δ
60	25.8	5.1	3.18	3	1.7	1.5	78.6°	7°

With this configuration, a matching of -22 dB can be achieved at the 868 MHz band. As a result, Figure 5 presents the simulated radiation pattern obtained from exciting each dipole at a time. It can be observed that the combination of the planar crossed dipoles is essentially isotropic.

As shown in Figure 6, the antenna is matched at the 868 MHz frequency band and has a good rejection rate of other bands. Also, it presents the higher gain at the frequency that is matched (868 MHz). Outside this band the gain severely declines. The gain obtained at the frequency of interest is 1.41 dBi which is close to the isotropic gain radiation expected for the proposed antenna. Recall that the antenna gain needs to be low to produce isotropic radiation, so that the physical orientation of the node will not affect the communication link in deployment.

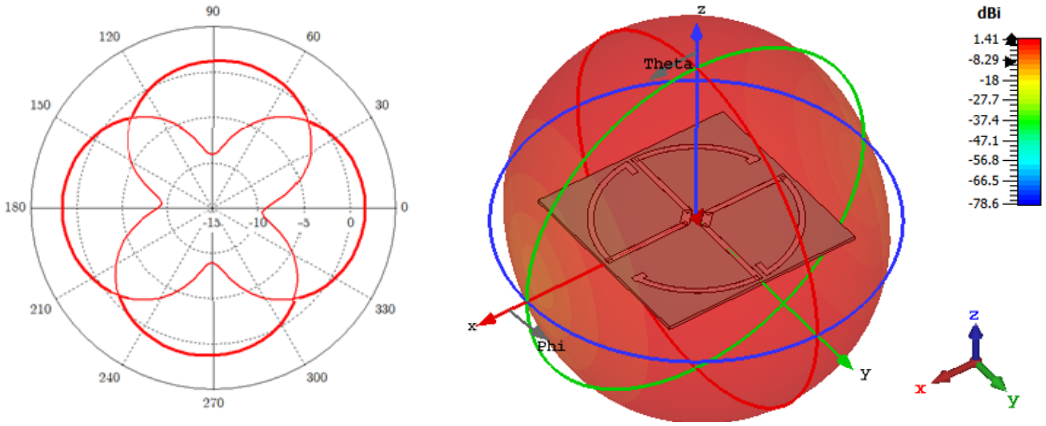


Figure 5. Radiation patterns of the Planar Crossed Dipoles for isotropic radiation in 3D (**right**) and the combination for Theta=0 exciting each dipole at a time (**left**).

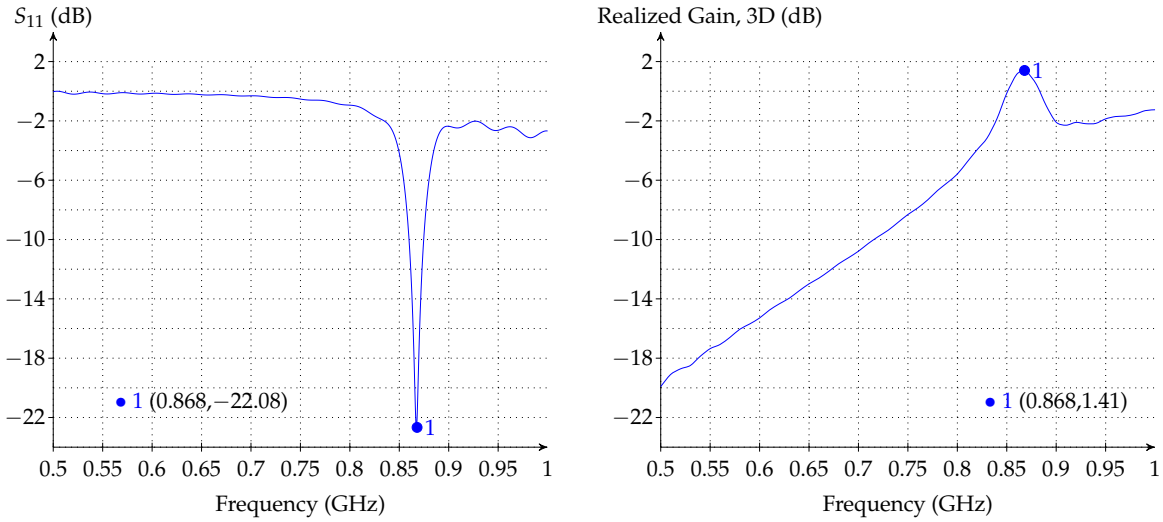


Figure 6. Performance results of the proposed Planar Crossed Dipoles. Reflection coefficient for isotropic radiation on the left. Realized Gain over frequency on the right.

3.3. Distributed Consensus

Designers of USNs typically select cloud or edge computing architectures [63] to outsource the heavy computation tasks associated with data streams processing [16]. This alleviates the requirements in terms of storage and computing of USN nodes but requires a reliable communications infrastructure able to transfer a large amount of data traffic to (and from) the cloud. However, in the specific scenario of large-scale urban sound monitoring, streaming the sensed acoustic data to a central entity (or cloud) would increase the complexity (in terms of codecs and connectivity to the Internet), the delay, the power consumption [64] and the overall cost of the nodes [65]. Therefore, the proposed USN has been designed to be autonomous (i.e., it can reliably identify acoustic events without the aid of powerful cloud devices) and self-managed. In this regard, a custom distributed consensus layer that enables synchronous communications among nodes has been implemented.

This layer is committed to increasing the robustness of the local acoustic event identification by comparing the identified local events at a single node with the events detected by neighboring nodes with the aim to emulate an ensemble decision system [66]. For instance, if a node detects a car horn but none of the surrounding nodes have detected this event, the system may decide to discard such event.

As shown in Figure 7, nodes are organized following a token ring topology. To keep the size of the ring small—recall that the purpose of the proposed USN is to take advantage of physical redundancy to enable more than one node listen the same event—and minimize the delay, all the nodes that keep a

close physical distance are assigned to the same ring. Therefore, to cover a large-scale physical area the same node can belong to more than one ring, which results in a multiring topology [67].

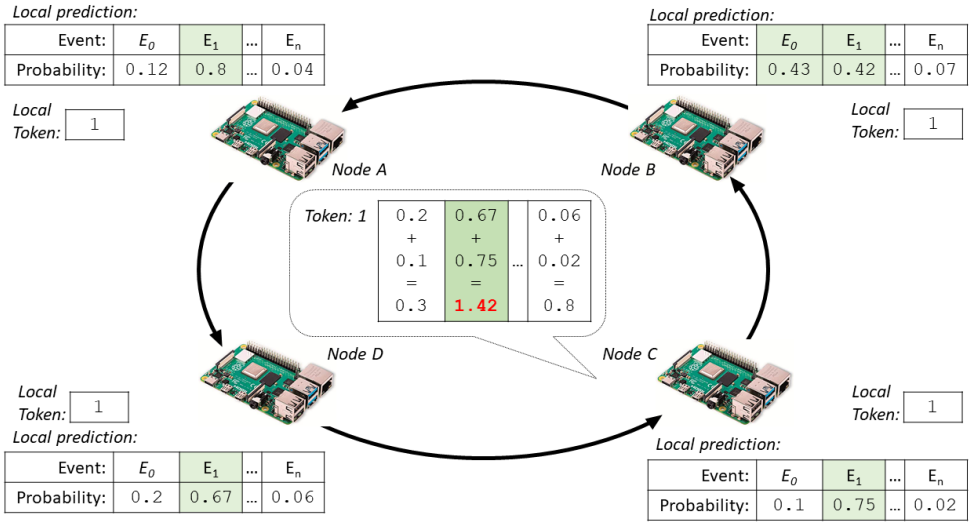


Figure 7. Logical organization of nodes.

The behavior of each node from the ring is as follows:

1. Run the local data processing (i.e., CNN classification algorithm) and wait for the classifier to populate the events vector (see Section 3.1).
2. Next, increment a local token number, which will give a notion of virtual synchrony among the nodes, i.e., a natural number indicating the logical time in which the classification has been conducted.
3. If the node has the lowest identifier among all the nodes of the ring, it sends a message to the next neighbor in the ring containing the local token number and the obtained events vector. The remaining nodes will wait to receive their associated message.
4. When a node receives the vector with a token number matching its local token number for the first time, it will make a component-wise addition between its own events vector and the vector contained in the message. Next, it will forward the resulting vector and the token number to the subsequent node.
5. When a node receives the vector with a token number matching its local token number for the second time, it means that all the nodes of the ring have contributed to the events vector contained in the message. At that moment, the node will apply a set of heuristic rules to determine the final label of the event. Specifically:
    - If the node locally classifies an event whose *Leq* is typically low (i.e., *air conditioner, children playing, dog bark and engine idling*) with a probability of more than 90%, the results obtained from the rest of the nodes (i.e., events vector from the message) of the ring will be ignored. In this case, the local events vector will be examined and the component with the highest value will be considered the winning label. The rationale behind this decision is that it is not likely than in a noisy street these sounds can be heard by different sensors of the ring, as background noise will probably mask them.
    - However, for the rest of the events (that typically have higher *Leq* such as horns or sirens), or if the network was not completely sure of whether one of the other events had actually occurred, the events vector from the message will be examined and the component with the highest value will be considered the winning label.

6. Increment the local token and go back to the first step. Please note that thanks to the local token, the system can associate the events vector with a logical time frame, which would be very useful in the case of faults (e.g., node crash, or communication fading).

Figure 7 shows an example of the proposed approach with a ring of 4 nodes. It can be seen that the most probable event detected at *Nodes* 0, 1, and 3 is E_1; however, *Node* 2 believes that the most probable event is E_0. However, after sharing the events vector with all the nodes of the ring, it will correct the local classification and agree with its neighboring nodes that the most probable event is E_1.

4. Experimental Evaluation

This section aims to validate the feasibility of the proposed approach by means of two experiments. In the first experiment, authors evaluate several deep network architectures. Using the original UrbanSound8K dataset [22], audio files are tested in a RPi to find out which classification algorithm offers the best trade-off between classification accuracy and memory/computing requirements. The aim of this experiment is to find an algorithm capable of classifying acoustic data in real-time with the resources provided by a single low-cost device. The results obtained in this first experiment will give a best-case scenario accuracy values that will be used as a baseline to compare the results of the second experiment.

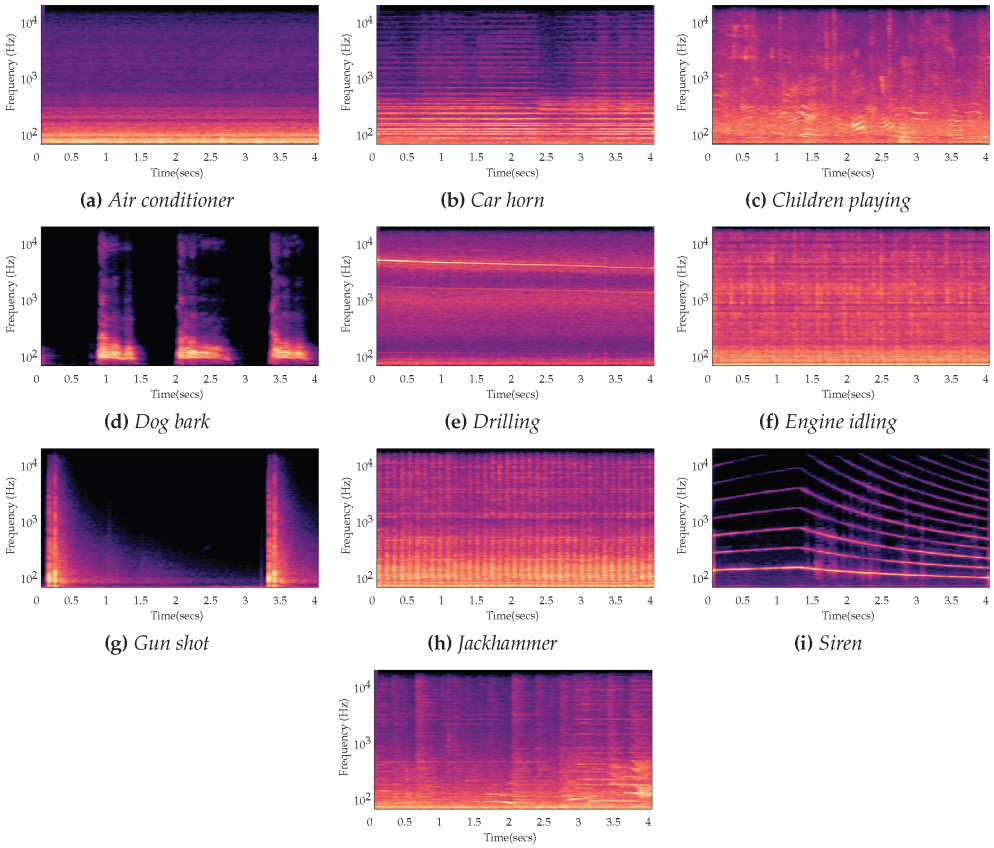
The second experiment aims to evaluate how different neighboring nodes connected as described in Section 3 would behave in an emulated (i.e., laboratory) real-world operation. For this purpose, the audio files from the dataset are modified emulating the air channel and physical topology of the streets of Barcelona. Moreover, road traffic noise recorded in the city of Barcelona [68] has been randomly added to each of the audio files so each sensor perceives the event partially masked by traffic noise. This experiment shows how a ubiquitous sensor network can improve the classification results over individual sensors when perceiving the same acoustic data from different locations and masked with traffic noise.

4.1. Experiment 1: Event Detection in Each Individual Sensor

UrbanSound8K is an online free dataset containing 8732 labelled sound events of 4 s or less from 10 different urban categories: *air conditioner, car horn, children playing, dog bark, drilling, engine idling, gun shot, jackhammer, siren,* and *street music*. The dataset has a total duration of 31,500 s and is preorganized in 10 different folds that must not be mixed according to [22]. For this work, we have used folds 1, 2, 3, 4, 6, 7 and 8 as training folds, fold 5 as validation fold and fold 10 as testing fold. Figure 8 shows a spectrogram example of each of the classes of the dataset.

Each sensor will be constantly running a deep-learning pre-trained network to be able to classify events in real time. For the experiment, we have obtained a 4-s window spectrogram of each of the audio files of the dataset. For those audio files on the UrbanSound8K dataset with a duration shorter than 4 s, we have applied the same methodology as Singh et al. in [69], which consists of replicating the same audio file until it reaches the uniform length of 4 s.

Table 2 details the different deep networks that were evaluated to find which classifier offers the best trade-off between accuracy, the number of floating-point operations (FLOPs) [70–74], and the size of the model after training and storing it into disk. For this experiment, all the networks were first trained using ImageNet [75], and we then applied transfer learning to fine-tune them so they could classify the spectrograms of the selected dataset. As the expected inputs of the network are RGB images such as the ones contained in ImageNet, each 4-s gray-scale spectrogram has been normalized in the range of [0, 1], then replicated three times (one per each RGB channel), and then normalized again with the mean and standard deviation of the ImageNet dataset.



(a) *Air conditioner* **(b)** *Car horn* **(c)** *Children playing*

(d) *Dog bark* **(e)** *Drilling* **(f)** *Engine idling*

(g) *Gun shot* **(h)** *Jackhammer* **(i)** *Siren*

(j) *Street music*

Figure 8. Spectrograms of the ten types of sounds of the UrbanSound8K dataset.

Table 2. Number of FLOPs, model size and accuracy on the testing fold for different network architectures.

Network Architecture	FLOPs	Accuracy	Model Size
ResNet 152	11.3×10^9	79.71%	223 MB
DenseNet 121	6×10^9	77.31%	28 MB
AlexNet	0.725×10^9	77.31%	218 MB
MobileNet v2	$\mathbf{0.3 \times 10^9}$	78.75%	8.8 MB
ShuffleNet v2	0.591×10^9	51.74%	5 MB
ResNet 18	1.8×10^9	77.19%	43 MB
VGG 16	15.3×10^9	77.91%	513 MB
SqueezeNet 18	0.833×10^9	**80.19%**	**2.9 MB**

The training of the network was carried out following standard good practice in deep learning, as follows. We used a batch size of 16 spectrograms per batch. The learning rate was initially set to 0.01, and a scheduler was programmed to decrease it by a factor of 0.1 with a patience of 3, using an SGD optimizer. Moreover, an early stopping criterion was used to obtain the optimal network configuration by using the validation fold.

As shown in Table 2, several network architectures obtain considerable high accuracy values compared to the baseline system that Salamon and Bello presented in [76], which obtained an accuracy of 68%. Concretely, the network architecture that provides the best results is SqueezeNet 18 [77], followed by Resnet 152 [70]. Comparing the size of both networks in terms of numbers of operations, Squeezenet 18 is clearly a smaller network, with only 0.833 GFLOPs compared to the 11.3 GFLOPs of ResNet 152. As SqueezeNet 18 fits into the purposed architecture (i.e., RPi) and can perform the classification in less than one window time (i.e., 4 s), it has been selected—adapting the last layer with

a 2D Convolutional layer with 10 outputs—to be the network installed in every node of the USN. Figure 9 depicts a diagram of the final architecture of the classifier system, and Figure 10 depicts the accuracy and loss when training and validating the selected model.

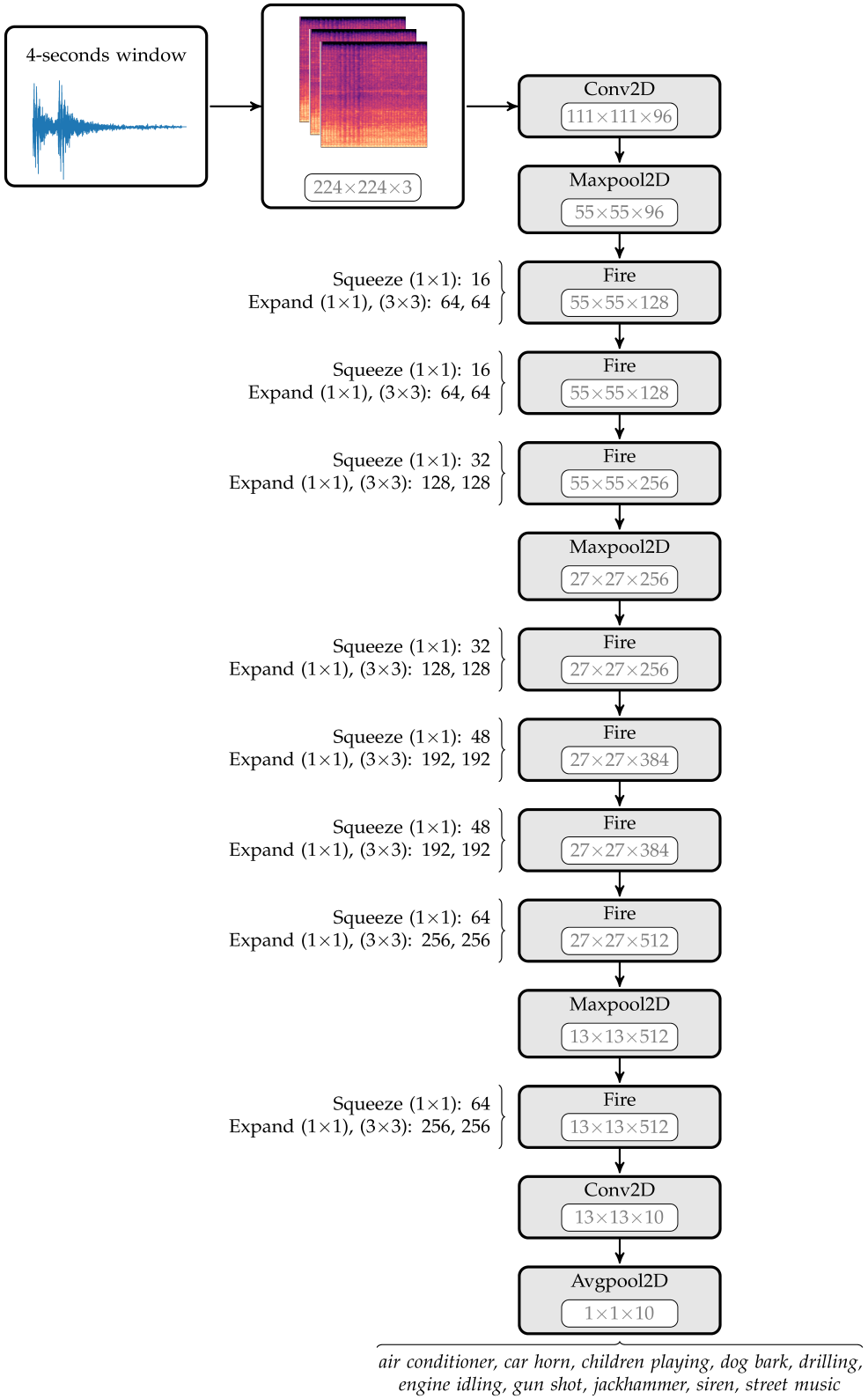


Figure 9. Deep network architecture for the local data processing.

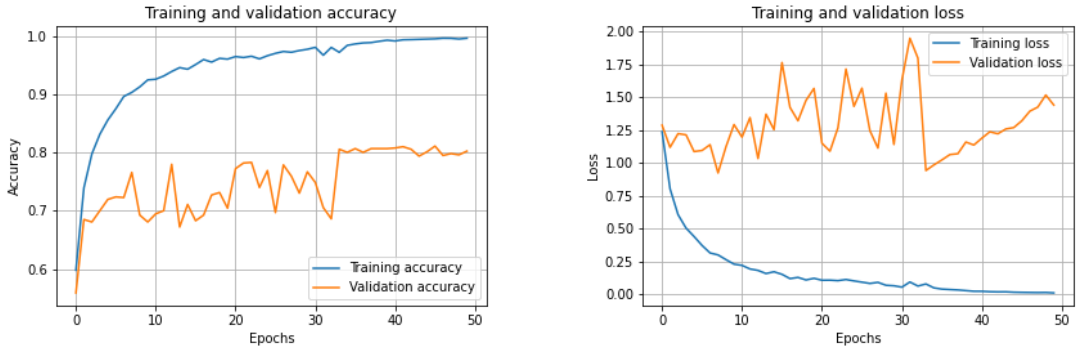


Figure 10. Training and validation accuracy and loss of the selected model.

The acoustic processing and classification of a single 4 s window in the RPi (Quad-Core Cortex A7 at 900 MHz and 1 GiB of RAM), using the aforementioned deep network, was carried out in 2 s. This time includes (1) taking the 4 s audio data acquired by the microphone, (2) calculating the spectrogram of a 4 s audio file of the UrbanSound dataset, (3) processing the spectrograms as explained above, and (4) passing the audio file through the deep net to obtain a local classification result. Hence, we can conclude that the proposed hardware platform is able to classify in real time—considering real-time as getting a classification result in less time than one window—in a 4 s basis, which is considerably faster than existing solutions that provide minute by minute data [39].

4.2. Experiment 2: Network of Sensors

The second experiment is aimed to assess how physical redundancy can increase the accuracy of the proposed system. To emulate the acoustic characteristics of a real-world scenario, the attributes of the Barcelona city center have been taken as a reference. According to *Pla Cerdà* [78], the physical sizes of the building blocks and streets from Barcelona are depicted in Figure 11: blocks size of 113.3 m $\times$ 113.3 m with a separation of 20 m between each block (horizontally and vertically). Please note that in Figure 11, white and green icons represent the acoustic sensors (microphone, antenna, and RPi), and the red dot represents an acoustic event that would be detected on nodes A, B, C and D. To deploy the proposed system, the following laboratory environment has been configured:

1. The trained model from the deep neural network used in Experiment 1 has been deployed in four different nodes (i.e., RPis). This aims to emulate the placement of the sensors in a street intersection (see Figure 11).
2. The distributed consensus protocol has been installed at each node.
3. A sound source has been placed in the scenario (i.e., red dot in Figure 11). To assess the advantages of the physical redundancy, a position located at different, yet reachable, distances from the four nodes was considered to be of interest (otherwise, more than one node could *hear* the same identical signal and, thus, the effect of physical redundancy could not be perceived). For this experiment, the selected distances between the sound source and nodes A, B, C, and D are 23.91 m, 57.95 m, 55.76 m, and 33.50 m respectively. This aims to emulate the location of the sound source just before the street crosswalk.
4. A synthetic acoustic test set has been generated for each node to later perform the event classification emulating how each sound would be perceived by each node. To obtain comparable results with the previous experiment, the test set has been derived from the same test fold as Experiment 1 and modified as follows:

    - The amplitude and phase of all the audio files from the testing fold have been changed according to the distance between the sound source (i.e., red dot in Figure 11) and the sensors A, B, C and D, respectively. Hence, the same audio file has been modified as many times as

neighbor nodes have been considered (four, in this case). The modifications of phase and amplitude of the samples have been carried out following the work of Bergadà et al. [79] in which, essentially, they propose the following equation:

$$y(n) = x(n - \tau)ae^{i\tau\omega}, \quad (1)$$

where x is the original signal, τ is the delay on the direct path considering the speed of sound and a is the absorption coefficient times the distance between the audio source and the sensor. Please note that the absorption coefficient is dependent on the frequency, temperature and humidity. For this experiment, the average values in the Barcelona city center have been taken: temperature of 20 °C and 70% relative humidity.

- Last, but not least, urban recordings of *road traffic noise* recorded on the city center of Barcelona [68] have been added (i.e., weighted sum) to each audio sample to further emulate a real-world environment. This is aimed to assess what happens when the background noise partially masks the acoustic event of interest. After conducting a grid search on which realistic combinations of weights better explain the effects of physical redundancy to improve detection accuracy, the following configuration for the attenuation factors of *road traffic noise* has been selected: 0.9 for node A, 0.88 for node B, 0.7 for node C, 0.68 for node D. This configuration ensures that the events to be detected are not completely masked. Also, this makes an uneven distribution of *road traffic noise* over the nodes—note that if all the nodes were exposed to the same amount of *road traffic noise* the individual output would be the same at all of them and, thus, physical redundancy would not improve accuracy. This configuration can be best seen as emulating the presence of traffic going from South to North in the street between nodes C and D of Figure 11 being the traffic noise closer to sensors D and C.

With this configuration, each node will later classify, at the same time, the same root acoustic sample with differences on its amplitude and phase and slightly different values of background noise.

5. Each node takes the acoustic sample from its test set, runs the local deep neural network, shares the classification vector to the neighboring nodes, and applies the consensus protocol to obtain the final result.

Table 3 depicts the confusion matrix obtained by averaging the results of the modified audio files on the four nodes. The local accuracy of the classifier with the modified dataset is: 68.19% on node A, 62.67% on node B, 59.94% on node C and 60.42% on node D. Hence, adding the *road traffic noise* to the audio files has made accuracy decrease by a ∼20% on average (recall that in Section 4.1, the accuracy obtained using the unaltered audio files was 80.19%). The reason behind this phenomenon is that the network has not been retrained with the modified audios. We have chosen not to retrain it for this experiment because we aim to emulate a real-world generic deployment where the background noise would be, a priori, unknown. Therefore, when deploying the system in a real-world urban environment with background traffic noise, we would expect to have the same accuracy drop.

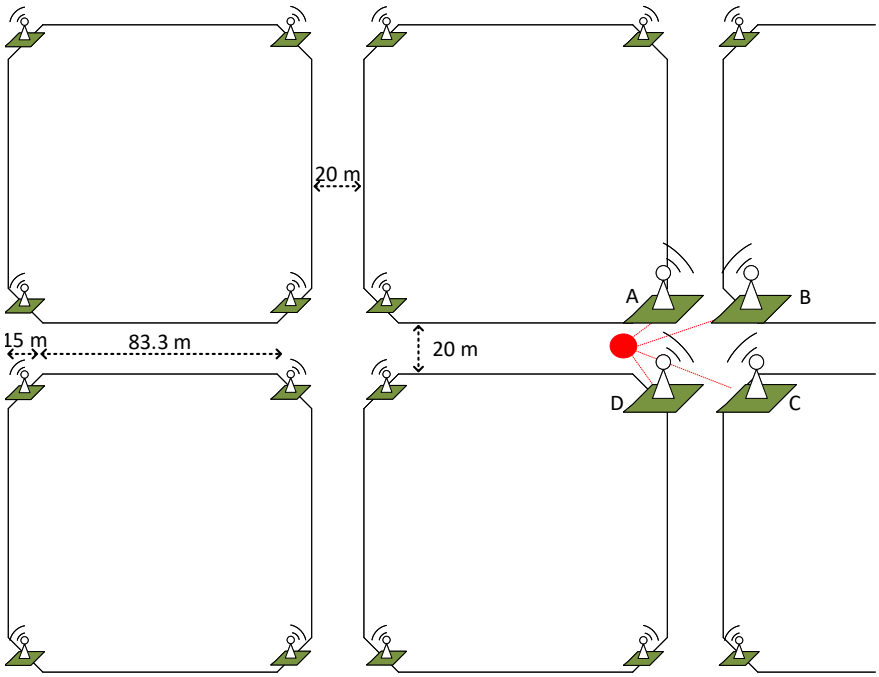


Figure 11. Diagram of the network of sensors (nodes) in the building blocks of the city of Barcelona. The green and white icons represent the sensor devices and the red dot represents an acoustic event.

However, after applying the distributed consensus protocol described in Section 3.3 and considering the neighboring event vectors obtained from nodes A, B, C and D; the accuracy values obtained in each of the four nodes are: 64.47% in node A, 64.35% in node B, 62.6% in node C and 64.42% in node D, which is, on average, higher than the accuracy obtained by a single node as shown below:

	Node A	**Node B**	**Node C**	**Node D**
Local accuracy	68.19%	62.67%	59.94%	60.42%
After consensus accuracy	64.47%	64.35%	62.60%	64.42%
Improvement	**−3.72%**	**+1.68%**	**+2.66%**	**+4.00%**

Table 3. Confusion matrix considering the classification of the modified audio files in a single sensor.

		Predicted Class									
		Air conditioner	*Car horn*	*Children playing*	*Dog bark*	*Drilling*	*Engine idling*	*Gunshot*	*Jackhammer*	*Siren*	*Street music*
Actual Class	*Air conditioner*	**30%**	0%	6%	0%	0%	50%	0%	0%	1%	13%
	Car horn	3%	**85%**	3%	0%	0%	3%	0%	0%	3%	3%
	Children playing	0%	0%	**84%**	6%	0%	3%	0%	0%	1%	6%
	Dog bark	0%	0%	8%	**82%**	2%	2%	0%	0%	0%	6%
	Drilling	5%	3%	3%	0%	**36%**	9%	1%	24%	2%	17%
	Engine idling	0%	0%	17%	0%	0%	**72%**	0%	0%	4%	7%
	Gunshot	6%	0%	15%	15%	0%	44%	**17%**	0%	0%	3%
	Jackhammer	0%	3%	0%	0%	2%	29%	0%	**56%**	0%	11%
	Siren	2%	0%	7%	29%	0%	1%	0%	0%	**60%**	1%
	Street music	0%	1%	23%	0%	0%	2%	0%	0%	0%	**74%**

Table 4 shows the average confusion matrix after applying the consensus algorithm. Observing both confusion matrices, we can see that the network tends to classify events from other categories as *Engine idling* (specially *Air conditioner*), probably because of the similarities between the spectral distribution of the *Road traffic noise* class from the BCNDataset and the *Air conditioner* and *Engine idling* classes of the UrbanSound8K dataset. The similarity between these two last events, which can be seen in Figure 8, is further emphasized when adding *Road traffic noise* to the audio files to be classified, as it contains noises from passing cars and motorbikes that can contain fragments of engines idling.

Table 4. Confusion matrix considering the classification of the modified audio files in a network of four nodes.

		Predicted Class									
		Air conditioner	*Car horn*	*Children playing*	*Dog bark*	*Drilling*	*Engine idling*	*Gunshot*	*Jackhammer*	*Siren*	*Street music*
Actual Class	*Air conditioner*	**31%**	0%	5%	0%	0%	46%	0%	0%	1%	17%
	Car horn	0%	**97%**	0%	0%	0%	3%	0%	0%	0%	0%
	Children playing	0%	0%	**82%**	7%	1%	3%	0%	0%	0%	7%
	Dog bark	0%	0%	6%	**84%**	2%	2%	0%	0%	0%	6%
	Drilling	3%	3%	2%	1%	**51%**	3%	1%	21%	3%	12%
	Engine idling	0%	0%	21%	0%	0%	**71%**	0%	0%	2%	6%
	Gunshot	3%	0%	22%	25%	0%	19%	**25%**	0%	0%	6%
	Jackhammer	0%	3%	0%	0%	1%	24%	0%	**58%**	0%	14%
	Siren	2%	0%	5%	33%	0%	0%	0%	0%	**59%**	1%
	Street music	0%	1%	17%	0%	0%	1%	0%	0%	0%	**81%**

Comparing the accuracy values obtained before and after applying the distributed consensus protocol, we can conclude that the multi-sensor approach has improved the classification results in all the sensors except for node A. As node A is the node that is closer to the acoustic event and it is the sensor where less background noise has been added, the consensus protocol has slightly decreased the accuracy of the classifier. However, this loss is compensated by the improvement in neighboring nodes.

5. Discussion

So far, we have shown the potential of taking advantage of the physical redundancy to increase the classification accuracy and robustness of an individual node. Alternative approaches such as the WASNs deployed in the cities of Rome and Milan on the DYNAMAP project [44] aim to deploy several sensors distributed in an area to enable noise map generation by interpolation, without taking into account physical redundancy. Our proposed system takes advantage of physical redundancy as the same physical space is *heard* by more than one sensor concurrently (four, in this case). According to Ref. [11], the main factors that drive scalability, accuracy, adaptability, and autonomy in urban sensor networks are the following:

Monitoring sound pressure levels accurately. In our case, the proposed approach is aimed at classifying acoustic events, but as the microphone is small enough to fit a standard acoustic calibrator, the modification of the RPi software to measure sound pressure levels would be relatively straightforward.

Providing intelligent, in situ signal processing, and wireless raw audio data transmission capabilities. In our case, although the raw audio data transmission would be feasible, we have reduced the amount of data to be transmitted by taking advantage of the edge computing paradigm. In this way, the amount of data (i.e., event labels) to be transmitted among nodes is

lower, which avoids bottlenecks in the communication network and, thus, shall improve the overall scalability.

As it is autonomous in its operation. In our case, the proposed system has been conceived considering fault tolerance by design. Therefore, if one node fails, the distributed protocol will be able to reconfigure itself to continue operating.

Having a price per node lower than or close to 100 USD. In our case, all the components of the proposed system have been conceived with low-cost devices to meet this requirement.

After demonstrating the feasibility of our proposal for urban sound monitoring in the proof-of-concept described in the previous section, we would like to share some lessons and experiences learnt during the design and development stages of the platform that might contribute to improving future versions.

5.1. Alternative Requirements for the Communications Antenna

The chosen operating band is 868 MHz (UN-39) because it presents advantages compared to other bands, such as robustness against absorption and higher data rates, and is not affected by a high number of other devices using this band (for example, remote controls). The antenna is designed for the UN-39 band in the ITU region 1. If the system were going to be used in another ITU region, the antenna design would need to be tuned to match the new frequency. The changes would affect the length of the crossed dipoles (*Lp* and *Lb*, in Table 1). Accordingly, the size of the support for the antenna may also increase to accommodate the longer arms.

The isotropic radiation pattern is a good option to eliminate arrangement issues with the sensors. The current bespoke antenna design is prepared to radiate in all directions so that the position of the sensor will not affect the communication link. This is an advantage of the current setup, with a medium range of 200 m maximum. For a longer range, the gain of the antenna may need to increase, losing the isotropic radiation property. In this case, by increasing the Δ in Table 1, the antenna will increase its gain, which will compensate for losses for the longer path–as shown in Friis Transmission Equation [55]. Still, this new configuration will lose the ability to be unaltered by the sensor positions.

5.2. Fault Tolerance

As discussed in Section 3, the proposed USN must tolerate a certain degree of faults. This means that the system must keep operating in case of failure in a node or communication link. The system is designed to support a *fail-stop* failures in a limited number of nodes or communication links.

For *fail-stop* failures in the nodes or the communication links, the system behaves as follows. When a node detects that the last time it received the token of the distributed consensus protocol is higher than a predefined threshold, it will try to reach the following node of the ring. If the communication is successful the ring will be reconfigured. If the communication is not successful, the node will change the token direction (i.e., clockwise or counterclockwise) and the system will adopt a token bus behavior instead of a token ring.

Additionally, each node should incorporate self-reboot policies (e.g., every 48 h, and/or when connection with neighbors is lost for more than 1 min) to avoid the nodes being frozen forever.

5.3. Real-World Deployment of the Proposed System

So far, the proposed system (i.e., deep network, communications antenna, and consensus protocol) has been assessed under laboratory conditions as shown in Section 4. Methodologically, this has enabled us to individually validate each component of the system and its end-to-end performance under a controlled environment. The lessons learnt during this process have let us consider the following points when deploying the system in a real-world scenario:

- The location of the nodes should be selected according to the architectonic profile of the scenario to enable physical redundancy. Please note that the proposed approach tolerates the addition or

the removal of nodes at will. Also, if larger communication distances—while ensuring that several nodes can *hear* the same high *Leq* acoustic event—were required due to the scenario characteristics, alternative strategies such as simultaneous wireless information and power transfer could be explored [51].

- In the case of microphones different from the OUT-AMLO-0872 being selected (e.g., MEMS), the most important requirements during deployment would be (1) omnidirectional pattern so they are able to pick up signal equally from all directions—to facilitate the installation of each node—(2) flat frequency response in the frequency range of, at least, 50 Hz–10 KHz (i.e., where acoustic events are taking place), and (3) 16 bits resolution per sample.
- As the same way as the microphone, the communications antenna has been designed to radiate following an isotropic pattern to facilitate its real-world deployment (i.e., no matter how the node is oriented). Although the electromagnetic interferences that may degrade the performance of the proposed antenna have not been considered in this work, it is worth mentioning that the Transport Control Protocol (TCP/IP) can be used to detect when the frames between nodes are lost or corrupted. If that happened, the consensus protocol would reconfigure the ring accordingly as described above.
- Experiments conducted over the RPi platform using the UrbanSound8K dataset suggest that the proposed system architecture would be capable of detecting acoustic events in real time using a deep convolutional neural network. However, when deploying the system in a real-world scenario, the classifier might struggle to distinguish anomalous acoustic events in noisy environments (i.e., locations with traffic background noise partially masking the acoustic events). Hence, the Squeezenet model should be retrained using data collected on the location where the nodes would be located. If, in the future, other types of events (not included in the UrbanSound8K) were to be detected, the following modifications should be made to the system:

  1. Add as many neurons as new event types to the last CONV2D layer of the deep network. In this case, the network should be retrained to be able to classify the new categories.
  2. Increase the size of the events vector sent to the neighbor nodes.
  3. Adapt the heuristic rules of the distributed consensus protocol to decide whether the new class contains noises that typically have a low value of L_{eq} or not.

  Therefore, the proposed acoustic USN could be easily adapted to potential classification of new event types.

6. Conclusions and Future Work

This research presents a low-cost acoustic sensor network to monitor urban sounds in large-scale areas. The proposed approach uses a pipeline composed of the following stages: (1) acoustic data acquisition and spectrogram computation, (2) local classification using a convolutional neural network (SqueezeNet architecture), (3) a custom bespoke antenna with isotropic radiation to share the local predictions with neighboring nodes, and (4) a distributed consensus protocol and a set of heuristic rules to unify the local predictions conducted at each node. To validate this proposal, the urban environment of the city of Barcelona has been selected. The proposed system detects the most probable events occurred on an acoustic sample taking advantage of the physical redundancy of the nodes. Regarding the physical redundancy, there are several reasons to consider four nodes per street intersection. The first of them is that the authors have chosen the Eixample district of Barcelona to conduct these experiments due to its symmetric structure. All the street intersections are of the same size and distance, which facilitates the design of a symmetric network. This leads us to the second reason. The number of nodes per street is probably too redundant, and possibly two nodes would have been sufficient to detect the noise events occurring around the intersections. However, the goal of the design is to have lots of low-cost nodes, collecting the same type of data at the same time for a large

number of locations simultaneously. This data redundancy is based on the concept that a low-cost node can appear as a commodity for the project, and it is the only way of gathering a huge amount of data to, not only by reliably detecting the acoustic events occurring, but also by having enough available information for other future applications such as drawing a precise map of the noise levels and their noise source.

A further potential application of this system is to automatically test whether a specific urban area meets certain acoustic regulations: for instance, when a specific event (e.g., air conditioner) is detected, it could be straightforward to decide whether the L_{Aeq} is below its associated threshold. Indeed, the obtained results of the proposed system encourage researchers to continue working on this direction, which in later stages will go through its implementation in a real-world and real-operation environment. This will enable practitioners to (1) evaluate the validity of the training carried out using UrbanSound8K and BCNDataset, and (2) verify the completeness of the model used for acoustic propagation, assuming that in a real-world situation the additive noise from the street will be more relevant.

Actually, in a real-world environment multiple events may occur simultaneously in the same acoustic sample. To maintain the excellent results in single-event-detection obtained in this validation test, the proposed deep network should move into a multi-label acoustic samples training, hence assuming that multiple events will occur simultaneously [80]. In this future stage, the consensus function of the distributed protocol should be adapted to tolerate the identification of multiple events.

Author Contributions: All authors have significantly contributed to this work. Conceptualization, E.V.-V., J.N., C.B.-F. and R.M.A.-P.; Data curation, E.V.-V.; Formal analysis, E.V.-V., J.N., C.B.-F., D.S. and R.M.A.-P.; Funding acquisition, R.M.A.-P.; Investigation, E.V.-V., J.N., C.B.-F. and D.S.; Methodology, J.N. and D.S.; Project administration, R.M.A.-P.; Resources, R.M.A.-P. and J.N.; Software, E.V.-V. and C.B.-F.; Supervision, J.N., D.S. and R.M.A.-P.; Validation, D.S. and R.M.A.-P.; Visualization, C.B.-F.; Writing—original draft, E.V.-V., J.N., C.B.-F. and R.M.A.-P.; Writing—review & editing, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Secretaria d'Universitats i Recerca of the Department of Business and Knowledge of the Generalitat de Catalunya under grants 2017-SGR-966 and 2017-SGR-977. Ester Vidaña-Vila and Rosa Ma Alsina-Pagès would like to thank La Salle Campus BCN - URL for partially funding the joint research with Queen Mary University (London) in the framework of Ms Vidaña-Vila PhD Thesis. Also, this work was partially funded by the Spanish Ministry of Science, Innovation and University, the Investigation State Agency and the European Regional Development Fund (ERDF) under grant RTI2018-097066-B-I00 for Joan Navarro and Cristina Borda-Fortuny.

Acknowledgments: The authors would like to thank Lisa Kinnear for her never-ending patience, support and thorough review of this work. Also, we gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

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

Abbreviations

The following abbreviations are used in this manuscript:

ADC	Analog-to-Digital Converter
CNN	Convolutional Neural Network
FLOP	Floating-Point Operation
GPU	Graphics Processing Unit
L_{Aeq}	Equivalent Level
MEMS	Micro Electrical Mechanical System
NMN	Noise Monitoring Network
USN	Ubiquitous Sensor Network
RGB	Red-Green-Blue
RPi	Raspberry Pi Model 2B
SGD	Stochastic Gradient Descent
SNR	Signal-to-Noise Ratio

SPLs Sound Pressure Levels
WHO World Health Organization
WASN Wireless Acoustic Sensor Network

References

1. Alexander, W. Some harmful effects of noise. *Can. Med Assoc. J.* **1968**, *99*, 27. [PubMed]
2. WHO/Europe | Noise-Data and Statistics. Available online: www.euro.who.int/en/health-topics/environment-and-health/noise/data-and-statistics (accessed on 6 September 2020).
3. Test, T.; Canfi, A.; Eyal, A.; Shoam-Vardi, I.; Sheiner, E.K. The influence of hearing impairment on sleep quality among workers exposed to harmful noise. *Sleep* **2011**, *34*, 25–30. [CrossRef] [PubMed]
4. Su-bei, M. Harm to human health from low frequency noise in city residential area. *China Med. Her.* **2007**, *4*, 17–19.
5. Moudon, A.V. Real noise from the urban environment: How ambient community noise affects health and what can be done about it. *Am. J. Prev. Med.* **2009**, *37*, 167–171. [CrossRef] [PubMed]
6. 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]
7. Flindell, I.; Walker, J. Environmental noise management. In *Advanced Applications in Acoustics, Noise and Vibration*; CRC Press: Boca Raton, FL, USA, 2004; p. 183.
8. Hurtley, C. *Night Noise Guidelines for Europe*; WHO Regional Office Europe: Bonn, Germany, 2009.
9. Office, P.C. Protection of the Environment Operations (Noise Control) Regulation 2017. *Legal Service Bull.* **2017**, *1*, 44.
10. Mun, S.; Geem, Z.W. Determination of individual sound power levels of noise sources using a harmony search algorithm. *Int. J. Ind. Ergon.* **2009**, *39*, 366–370. [CrossRef]
11. Mydlarz, C.; Salamon, J.; Bello, J.P. The implementation of low-cost urban acoustic monitoring devices. *Appl. Acoust.* **2017**, *117*, 207–218. [CrossRef]
12. ITU. Ubiquitous Sensor Networks (USN). In *Technical Report, ITU-T Technology Watch Briefing Report Series*; ITU, No. 4; ITU: East Lansing, MI, USA, 2008.
13. Ferrández-Pastor, F.J.; García-Chamizo, J.M.; Nieto-Hidalgo, M.; Mora-Pascual, J.; Mora-Martínez, J. Developing ubiquitous sensor network platform using internet of things: Application in precision agriculture. *Sensors* **2016**, *16*, 1141. [CrossRef]
14. Murty, R.N.; Mainland, G.; Rose, I.; Chowdhury, A.R.; Gosain, A.; Bers, J.; Welsh, M. Citysense: An urban-scale wireless sensor network and testbed. In Proceedings of the 2008 IEEE Conference on Technologies for Homeland Security, Waltham, MA, USA, 12–13 May 2008; pp. 583–588.
15. Shin, D.; Na, S.Y.; Kim, J.Y.; Baek, S.J. Fish robots for water pollution monitoring using ubiquitous sensor networks with sonar localization. In Proceedings of the 2007 International Conference on Convergence Information Technology (ICCIT 2007), Gyeongju, Korea, 21–23 November 2007; pp. 1298–1303.
16. Navarro, J.; Vidaña-Vila, E.; Alsina-Pagès, R.M.; Hervás, M. Real-time distributed architecture for remote acoustic elderly monitoring in residential-scale ambient assisted living scenarios. *Sensors* **2018**, *18*, 2492. [CrossRef]
17. Bagula, A.; Zennaro, M.; Inggs, G.; Scott, S.; Gascon, D. Ubiquitous sensor networking for development (usn4d): An application to pollution monitoring. *Sensors* **2012**, *12*, 391–414. [CrossRef] [PubMed]
18. Koucheryavy, A.; Vladyko, A.; Kirichek, R. State of the art and research challenges for public flying ubiquitous sensor networks. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; Springer: Cham, Switzerland, 2015; pp. 299–308.
19. Ghemawat, S.; Gobioff, H.; Leung, S.T. The Google file system. In Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, Bolton Landing, NY, USA, 19–22 October 2003; pp. 29–43.
20. Piper, B.; Barham, R.; Sheridan, S.; Sotirakopoulos, K. Exploring the "big acoustic data" generated by an acoustic sensor network deployed at a crossrail construction site. In Proceedings of the 24th International Congress on Sound and Vibration (ICSV), London, UK, 23–27 July 2017; pp. 23–27.
21. Raspberry Pi Official Web Site. Available online: https://www.raspberrypi.org (accessed on 1 October 2020).

22. 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 (ACM-MM'14), Orlando, FL, USA, 3–7 November 2014; pp. 1041–1044.
23. Wikipedia Contributors. Eixample—Wikipedia, The Free Encyclopedia, 2020. Available online: https://ca.wikipedia.org/wiki/Eixample_de_Barcelona#/media/Fitxer:Eixample_aire.jpg (accessed on 5 October 2020).
24. Polastre, J.; Szewczyk, R.; Culler, D. Telos: Enabling ultra-low power wireless research. In Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, Los Angeles, CA, USA, 25–27 April 2005; p. 48.
25. Santini, S.; Vitaletti, A. Wireless sensor networks for environmental noise monitoring. In *6. Fachgespräch Sensornetzwerke*; Technische Universität Hamburg: Hamburg, Germany, 2007; p. 98.
26. Santini, S.; Ostermaier, B.; Vitaletti, A. First experiences using wireless sensor networks for noise pollution monitoring. In Proceedings of the 2008 Workshop on Real-World Wireless Sensor Networks (REALWSN), Glasgow, Scotland, 1 April 2008; pp. 61–65.
27. Wang, C.; Chen, G.; Dong, R.; Wang, H. Traffic noise monitoring and simulation research in Xiamen City based on the Environmental Internet of Things. *Int. J. Sustain. Dev. World Ecol.* **2013**, *20*, 248–253. [CrossRef]
28. Paulo, J.; Fazenda, P.; Oliveira, T.; Carvalho, C.; Félix, M. Framework to monitor sound events in the city supported by the FIWARE platform. In Proceedings of the 46o Congreso Español de Acústica, Valencia, Spain, 21–23 October 2015; pp. 21–23.
29. Paulo, J.; Fazenda, P.; Oliveira, T.; Casaleiro, J. Continuos sound analysis in urban environments supported by FIWARE platform. In *Proceedings of the EuroRegio2016/TecniAcústica*, Porto, Portugal, 13–15 June 2016; Volume 16, pp. 1–10.
30. Mietlicki, F.; Mietlicki, C.; Sineau, M. An innovative approach for long-term environmental noise measurement: RUMEUR network. In Proceedings of the EuroNoise 2015, Maastrich, The Netherlands, 31 May–3 June 2015; pp. 2309–2314.
31. Mietlicki, C.; Mietlicki, F. Medusa: A new approach for noise management and control in urban environment. In Proceedings of the EuroNoise 2018, Heraklion, Crete, Greece, 27–31 May 2018; pp. 727–730.
32. Camps-Farrés, J. Barcelona noise monitoring network. In Proceedings of the Euronoise, Maastrich, The Netherlands, 31 May–3 June 2015; pp. 218–220.
33. Camps-Farrés, J.; Casado-Novas, J. Issues and challenges to improve the Barcelona Noise Monitoring Network. In Proceedings of the EuroNoise 2018, Heraklion, Crete, Greece, 27–31 May 2018; pp. 693–698.
34. Coulson, S.; Woods, M.; Scott, M.; Hemment, D.; Balestrini, M. Stop the noise! enhancing meaningfulness in participatory sensing with community level indicators. In Proceedings of the 2018 Designing Interactive Systems Conference, New York, NY, USA, 9–13 June 2018; pp. 1183–1192.
35. Basten, T.; Wessels, P. An overview of sensor networks for environmental noise monitoring. In Proceedings of the 21st International Congress on Sound and Vibration (ICSV21), Beijing, China, 13–17 July 2014; pp. 1–8.
36. Cense-Characterization of Urban Sound Environments. Available online: http://cense.ifsttar.fr/ (accessed on 1 December 2020).
37. Botteldooren, D.; De Coensel, B.; Oldoni, D.; Van Renterghem, T.; Dauwe, S. Sound monitoring networks new style. In *Acoustics 2011: Breaking New Ground: Proceedings of the Annual Conference of the Australian Acoustical Society*; Mee, D.J., Hillock, I.D., Eds.; Australian Acoustical Society: Brisbane, QLD, Australia, 2011; pp. 93:1–93:5.
38. Domínguez, F.; Dauwe, S.; Cuong, N.T.; Cariolaro, D.; Touhafi, A.; Dhoedt, B.; Botteldooren, D.; Steenhaut, K. Towards an environmental measurement cloud: Delivering pollution awareness to the public. *Int. J. Distrib. Sens. Netw.* **2014**, *10*, 541360. [CrossRef]
39. Bell, M.C.; Galatioto, F. Novel wireless pervasive sensor network to improve the understanding of noise in street canyons. *Appl. Acoust.* **2013**, *74*, 169–180. [CrossRef]
40. Rainham, D. A wireless sensor network for urban environmental health monitoring: UrbanSense. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2016; Volume 34, p. 012028.
41. Bartalucci, C.; Borchi, F.; Carfagni, M.; Furferi, R.; Governi, L.; Lapini, A.; Bellomini, R.; Luzzi, S.; Nencini, L. The smart noise monitoring system implemented in the frame of the Life MONZA project. In Proceedings of the EuroNoise 2018, Heraklion, Crete, Greece, 27–31 May 2018; pp. 783–788.

42. De Coensel, B.; Botteldooren, D. Smart sound monitoring for sound event detection and characterization. In Proceedings of the 43rd International Congress on Noise Control Engineering (Inter-Noise 2014), Melbourne, Australia, 16–19 November 2014; pp. 1–10.
43. Brown, A.; Coensel, B.D. A study of the performance of a generalized exceedance algorithm for detecting noise events caused by road traffic. *Appl. Acoust.* **2018**, *138*, 101–114. [CrossRef]
44. 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*, 1. [CrossRef]
45. Bellucci, P.; Peruzzi, L.; Zambon, G. Life Dynamap project: The case study of Rome. *Appl. Acoust.* **2017**, *117*, 193–206. [CrossRef]
46. 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]
47. 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]
48. 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]
49. Bellucci, P.; Cruciani, F.R. Implementing the Dynamap system in the suburban area of Rome. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Hamburg, Germany, 21–24 August 2016; pp. 5518–5529.
50. Alsina-Pagès, R.M.; Hervás, M.; Duboc, L.; Carbassa, J. Design of a Low-Cost Configurable Acoustic Sensor for the Rapid Development of Sound Recognition Applications. *Electronics* **2020**, *9*, 1155. [CrossRef]
51. Ji, B.; Chen, Z.; Mumtaz, S.; Liu, J.; Zhang, Y.; Zhu, J.; Li, C. SWIPT Enabled Intelligent Transportation Systems with Advanced Sensing Fusion. *IEEE Sensors J.* **2020**, *4*, 1. [CrossRef]
52. Huzaifah, M. Comparison of time-frequency representations for environmental sound classification using convolutional neural networks. *arXiv* **2017**, arXiv:1706.07156.
53. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press Cambridge: Cambridge, MA, USA, 2016; Volume 1.
54. Mesaros, A.; Heittola, T.; Benetos, E.; Foster, P.; Lagrange, M.; Virtanen, T.; Plumbley, M.D. Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *26*, 379–393. [CrossRef]
55. Pozar, D. *Microwave Engineering*, 4th ed.; Wiley: Hoboken, NJ, USA, 2011.
56. Ministerio de Energía, Turismo y Agenda Digital. Real Decreto 123/2017, de 24 de Febrero, por el que se Aprueba el Reglamento Sobre el uso del Dominio Público Radioeléctrico. Available online: https://www.boe.es/buscar/act.php?id=BOE-A-2017-2460&tn=1&p=20170308#ar-6 (accessed on 30 July 2020).
57. CNAF. Notas UN. Available online: https://avancedigital.gob.es/espectro/CNAF/notas-UN-2017.pdf (accessed on 30 July 2020).
58. CNAF. ATRIBUCIÓN A LOS SERVICIOS según el RR de la UIT. Available online: https://avancedigital.gob.es/espectro/CNAF/tablas_2017.pdf (accessed on 30 July 2020).
59. CNAF. Artículo 5 del Reglamento de Radiocomunicaciones. Available online: https://avancedigital.gob.es/espectro/CNAF/notasRR-2017.pdf (accessed on 30 July 2020).
60. González, L.P.; Jaedicke, C.; Schubert, J.; Stantchev, V. Fog computing architectures for healthcare. *J. Inf. Commun. Ethics Soc.* **2016**, *14*, 334–349. [CrossRef]
61. Links, E.R. LoRa 868/900 MHz SX1272 LoRa Module for Arduino Waspmote and Raspberry Pi. Available online: https://www.cooking-hacks.com/documentation/tutorials/ (accessed on 1 December 2020).
62. Pan, G.; Li, Y.; Zhang, Z.; Feng, Z. Isotropic Radiation From a Compact Planar Antenna Using Two Crossed Dipoles. *IEEE Antennas Wirel. Propag. Lett.* **2012**, *11*, 1338–1341.
63. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.; Konwinski, A.; Lee, G.; Patterson, D.; Rabkin, A.; Stoica, I.; et al. A view of cloud computing. *Commun. ACM* **2010**, *53*, 50–58. [CrossRef]
64. Ji, B.; Chen, Z.; Chen, S.; Zhou, B.; Li, C.; Wen, H. Joint optimization for ambient backscatter communication system with energy harvesting for IoT. *Mech. Syst. Signal Process.* **2020**, *135*, 106412. [CrossRef]

65. Pham, C.; Cousin, P. Streaming the sound of smart cities: Experimentations on the smartsantander test-bed. In Proceedings of the 2013 IEEE International Conference on Green Computing And Communications and IEEE Internet of Things and IEEE Cyber, Physical And Social Computing, Beijing, China, 20–23 August 2013; pp. 611–618.
66. Nanni, L.; Costa, Y.M.; Aguiar, R.L.; Mangolin, R.B.; Brahnam, S.; Silla, C.N. Ensemble of convolutional neural networks to improve animal audio classification. *EURASIP J. Audio Speech Music. Process.* **2020**, *2020*, 1–14. [CrossRef]
67. Aiello, W.; Bhatt, S.N.; Chung, F.R.; Rosenberg, A.L.; Sitaraman, R.K. Augmented ring networks. *IEEE Trans. Parallel Distrib. Syst.* **2001**, *12*, 598–609. [CrossRef]
68. Vidaña-Vila, E.; Duboc, L.; Alsina-Pagès, R.M.; Polls, F.; Vargas, H. BCNDataset: Description and Analysis of an Annotated Night Urban Leisure Sound Dataset. *Sustainability* **2020**, *12*, 8140. [CrossRef]
69. Singh, S.; Pankajakshan, A.; Benetos, E. Audio Tagging using Linear Noise Modelling Layer. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York, NY, USA, 25–26 October 2019; pp. 234–238.
70. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
71. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2016**, arXiv:1608.06993.
72. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *arXiv* **2017**, arXiv:1707.01083.
73. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L. Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation. *arXiv* **2018**, arXiv:1801.04381.
74. Ma, N.; Zhang, X.; Zheng, H.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. *arXiv* **2018**, arXiv:1807.11164.
75. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 22–24 June 2009; pp. 248–255.
76. Salamon, J.; Bello, J.P. Unsupervised feature learning for urban sound classification. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia, 19–24 April 2015; pp. 171–175.
77. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *arXiv* **2016**, arXiv:1602.07360.
78. Aibar, E.; Bijker, W.E. Constructing a city: The Cerdà plan for the extension of Barcelona. *Sci. Technol. Hum. Values* **1997**, *22*, 3–30. [CrossRef]
79. Bergadà, P.; Alsina-Pagès, R.M. An Approach to Frequency Selectivity in an Urban Environment by Means of Multi-Path Acoustic Channel Analysis. *Sensors* **2019**, *19*, 2793. [CrossRef]
80. Cartwright, M.; Mendez, A.E.M.; Cramer, J.; Lostanlen, V.; Dove, G.; Wu, H.H.; Salamon, J.; Nov, O.; Bello, J.P. Sonyc urban sound tagging (sonyc-ust): A multilabel dataset from an urban acoustic sensor network. In Proceedings of the Acoustic Scenes and Events 2019 Workshop (DCASE2019), New York, NY, USA, 25–26 October 2019; p. 35.