



## On the application of big data techniques to noise monitoring of smart cities

Navarro, J.M.<sup>1</sup>, Tomas-Gabarron, J.B.<sup>2</sup>, Escolano, J.<sup>3</sup>

<sup>1</sup>Polytechnic School, UCAM Catholic University of San Antonio, 30107 Murcia, SPAIN.  
jmnavarro@ucam.edu

<sup>2</sup>Peranoid SL, 30800 Lorca, Murcia, SPAIN.  
juanba.tomas@peranoid.com

<sup>3</sup>Sophandrey Research, 11224 Brooklyn NY, USA.  
joseescolanocarrasco@gmail.com

### Abstract

The proliferation of networks of acoustic sensors to monitor the sound pressure levels in cities are allowing the collection of vast amounts of information. As is done in other fields, it would be interesting to investigate the applicability of the different techniques of big data for environmental acoustics. This paper presents an overview of potential applications where big data can be used for monitoring noise in smart cities. Before that, development phases and tools for big data analysis are described. Finally, a preliminary study with data obtained from a real smart city is presented, where equivalent sound pressure level to all sensors network in different time intervals are evaluated.

**Keywords:** Smart City, noise monitoring, big data, open data, acoustic sensor network.

**PACS no.** 43.50.+y, 43.58.+z, 43.60.+d

## 1 Introduction

During last years, the concept of Smart City, and the application of the so called Internet of Things (IoT) in it, has been followed with great interest by the scientific community. A Smart City is a holistic vision of a city who applies the information and communication technologies (ICT) to improve the life quality of its inhabitants and to ensure an economic, social and environmental sustainable development in continuous improvement. In a Smart City, it is necessary to collect data and to efficiently take decisions on the environment. To do this, it is making use of technologies such as wireless sensor networks that allow easily deploy of a large number of devices with the aim of measuring parameters and intelligently perform tasks; from traffic control systems, through irrigation of parks and gardens, to the evaluation of noise pollution.

For the development of the Internet of Things [1,2], and the Smart Cities [3], wireless sensor networks [4] are used. These sensor networks enable real-time monitoring of many parameters that can be used to facilitate sustainable lifestyle, save costs and improve the quality of life of people. However, this vast amount of information gathered by the sensors, together with the acquired data from other sources, such as mobile applications, social networking and administration management software, has

to be properly stored and processed in order to obtain useful information. Analysing large volumes of data is a great challenge, therefore, big data techniques[5,6] for collecting, storing and analysing big data sets should be applied to perform them efficiently and accurately.

People of today's cities are exposed to excessive noise levels which cause annoyance[7]. This noise disturbance is the result of repeated and prolonged exposure to these high levels of noise, leading to a decrease in quality of life, interfering with their daily work, increasing stress and fatigue, decreasing the concentration and rest, and getting to produce health disease[8]. Wireless sensor networks have been used to perform acoustic measurements, focusing on obtaining basic parameters such as sound pressure level [9,10] and annoyance[11,12] to mainly design and update noise maps in cities. Also, research has been conducted where the citizen collaborate and his mobile phone was used as a mobile sensor, both general parameters[13] and noise[14,15,16], showing the latter great dependency on the type of device used [17].

In acoustics, big data techniques have been firstly applied to analyse sound in captured audio files due to its large size and complexity; e.g. an acoustic sensor can generate several gigabytes of audio data per day. Zhang et. al [18] presented different techniques to store and process environmental audio data from monitoring. In [19], Mugagga et. al showed the use of big data for improving sound source localisation. Big data techniques had been also applied with audio data for speech analysis [20].

In this paper, big data techniques for noise monitoring in big data are proposed. Several challenges using huge noise data acquired by Smart Cities can be undertaken. For example, the analysis of large volumes of long term noise data could be used to predict future noise levels in areas and to improve traffic noise predictions models[21]. Moreover, big data techniques combined with machine learning algorithm allow to detect early warning of noise events. Dynamic updating of strategic noise map [22] with interpolation methods, environmental noise monitoring [23] and sound propagation outdoors are some examples that can also be improved with big data.

This paper presents an overview of potential applications where big data can be used for monitoring noise in smart cities and a preliminary experience in a real setup. The manuscript is divided in the following sections; after this introduction, Section 2 outlines big data basics and describes different phases of the regular procedure. Section 3 presents an experimental application of big data for a data set captured from Smart Santander platform and results are discussed. Finally, conclusions are enunciated.

## **2 Big data basics**

The term Big Data is used to define data set that are so large or complex that traditional data processing applications are inadequate. Big data requires several techniques and technologies to capture, store, curate, process and presents insights from data sets that are diverse and complex. In this section, some basics concepts about big data stages and techniques are described, together with the needed framework and tools for a smart data analysis using big data.

### **2.1 Obtaining and storing data**

In general, noise pollution data can be massively collected by three approaches: acoustic sensor network, mobile phones, and open data platform.

An acoustic sensor network is created by nodes that capture sound in near realtime, and pre-process it to send data; e.g. sound pressure levels, to a gateway or concentrator. These gateways collect data from nodes and transmit it to be stored in the central server; i.e. smart city platform. In Smart Cities wireless communication protocol, like 3G, Wifi or Zigbee [24, 12], are mostly deployed to obtain data from this kind of nodes. A sensor node consists in a hardware board that interface with an electret external microphone, pre-amplifier and external power supply; solar panel and batteries mostly, and a communication module. The system should be autonomous, and it should be able to operate continuously for months with minimal maintenance.

Instead of monitoring a fixed location, mobile phones allow both fixed and moving measurements. It is important to note that, it is recommended to install an external microphone to increase precision and accuracy. However, mobile phone high battery consumptions could be a restriction to large period monitoring.

Several already working Smart Cities have deployed a variety of sensor networks that publish open data sets through their cloud platform. In this research, Smart Santander open data system, called FiWare [25] was used. FiWare consists of a public-private cloud platform, initially supported by the European Commission, whose goal is to develop and implement a network of applications and services for the Future Internet. Differently from its main competitors like Microsoft Azure or Amazon Web Services, FiWare provides an open-source architecture that allows developers, service providers, SMEs and any type of organization the development of products and services with a very high value while minimizing costs of implementation and deployment.

One of the most important services that FiWare supports is Orion Context Broker, a so-called Generic Enabler that allows us to extract data captured by noise sensors resorting to a NGSII9/10 REST API interface [26, 27]. Orion can manage the entire lifecycle that regards the update, query, registration and subscription of data from sensors. Likewise, all data is stored in a non-relational database called MongoDB. In this regard, no-SQL storage engines are characterized by its higher simplicity with respect to SQL approaches, as well as its better horizontal scalability and resilience. That is the reason why no-SQL databases are increasingly popular nowadays, especially regarding the emerging fields of IoT, and the support given by giants like Amazon, Facebook and Google.

On the other hand, one of the most interesting advantages of an open cloud platform like FiWare is that there are several public/private entities that share large data-traces they gather with their architecture of sensors. That is the case of SmartSantander [28], the platform of the city of Santander (Spain) that continuously captures the value of several sound sensor nodes located throughout this city, as it is shown in Figure 1. Thanks to the availability of this data is how the present study is performed.

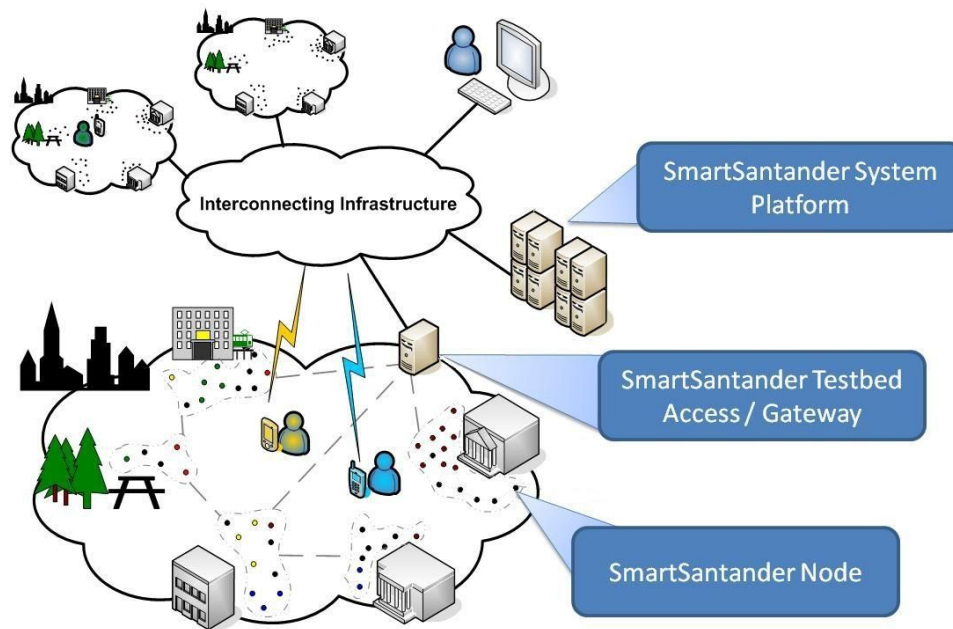


Figure 1 – Smart Santander architecture.

## 2.2 Analysing data

Noise sensor network provides environmental professionals with the capability to massively scale acoustical observations, both temporally and spatially. Due to the nature of the data and its analysis, batch processing seems to be a natural solution: data is stored as it is streamed but the analysis is performed every certain period of time, e.g. on daily basis. For that reason an Apache MapReduce Framework is used. This is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster [29]. The main advantage of this framework is to allow distributed computation following certain structures. Therefore, MapReduce facilitates developers to focus on the logic and not in the inherent difficulties of implement distributed/parallel software.

A MapReduce program is composed of a Map() procedure that performs filtering and a Reduce() method that performs a summary operation or rollup logic. A MapReduce-based framework orchestrates the processing by marshalling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and providing for redundancy and fault tolerance.

The key contributions of the MapReduce framework are not the actual map and reduce functions, but the scalability and fault-tolerance achieved for a variety of applications by optimizing the execution engine once. As such, a single-threaded implementation of MapReduce will usually not be faster than a traditional (non-MapReduce) implementation; any gains are usually only seen with multi-threaded implementations. The use of this model is beneficial only when the optimized distributed shuffle operation (which reduces network communication cost) and fault tolerance features of the MapReduce framework come into play. Optimizing the communication cost is essential to a good MapReduce algorithm.

One of the most essential steps relies on how the map phase generates one (or more) key-value pairs for each record. Every key-value pair that contains the same key will be processed by the same reduce process, independent and parallel to other similar reducer processes.

In this paper, as a preliminary study, every collected data from noise nodes are processed in order to calculate the Community Noise Equivalent Level (CNEL) [30], or day-evening-night equivalent level (Lden). This parameter, Lden, is an A-weight equivalent noise level measured over the 24 hour period with a penalty in evening and night periods, as defined in next equation:

$$L_{den} = 10 \cdot 10 \log \frac{12 \cdot 10^{\frac{L_{day}}{10}} + 4 \cdot 10^{\frac{L_{evening}+5}{10}} + 8 \cdot 10^{\frac{L_{night}+10}{10}}}{24} \quad (1)$$

where, Lday is the A-weight equivalent noise level measured between 7 and 19 hours, Levening is the A-weight equivalent noise level measured between 19 and 23 hours, and Lnight is the A-weight equivalent noise level measured between 23 and 7 hours. Therefore, it is needed to calculate, for each node, the equivalent noise level in different periods; day, evening, and night, and right after that to compute Lden for nodes, zones and the whole city. Lden is defined as a long term measurement because it must be determined along all periods day-evening-night of one year. The importance of using Big Data appears when dealing with data coming from many sensors, with high refresh rate and a period of measurement over a full year.

Focusing on the nature of the problem of this paper, a single MapReduce job has been defined, following the next steps:

- Map phase, which has two main purposes:
  - To clean/filter records that does not contain substantial information (missing values) will be dismissed. E.g. sound pressure level value is missing.
  - Create the right key to submit each record to its corresponding reducer. Since the purpose of parallelizing the analysis of the contribution of each sensor for a given time period, the key is defined as a 3-tuple that contains the sensor identifier, period of time if belongs, e.g. day/evening/night and date.
- Shuffle phase: For single value keys, the MapReduce framework sort the keys before send the key-value pairs to their corresponding reducers. For complex keys (n-tuples), it is necessary to indicate the sorting hierarchy. This technique is known as secondary sorting. In this particular case, each reducer will receive all the key with the same sensor id and time period, and sorted by time of the given date.
- Reducer phase: In this phase, an aggregation to calculate the desired averaged sound pressure level it is performed. Thanks to the fact every record is sorted by time, it is possible to calculate the time difference to perform some of the analytics that this paper aims.

### 2.3 Presenting data

Once the mapreduce task is finished, every reducer generates a single record that contains the following information: sensor id, date, time period (day/evening/night) and the equivalent sound pressure level. Moreover, different analytics are calculated to describe results classified by node, different time interval (period, day, week, month, year,...), and different areas (zones and whole city).

## 3 Experiment with Smart Santander data

In this section a preliminary experiment to calculate  $L_{den}$  from noise node data of Smart Santander is presented.

### 3.1 Environment

Thanks to the availability of acoustic information from the Smart Santander platform, enough data can be collected to carry out Big Data analysis of the behavior of ambient noise in different parts of the city, and for long intervals of time. For this particular experiment, it has been analysed the data of all sensors from the Smart Santander platform during a 10-day interval, at a very fine-grained accuracy (every time the measurement of the sensor changes, i.e. 1 minute at most). After a mapping phase, where malfunctional nodes data were deleted, e.g. nodes that give always 100 dB or 50 dB maximum and minimum threshold respectively, data from 30 nodes in total were analyzed.

Each node streams approximately 1.5 MB of data per day. The information captured during these 10 days of experiment was approximately 450 MB, that extrapolated to a year period is about 160 GB. The exponential grow of acoustic sensor deployment in city, e.g. in Santander more than 1,000 could be installed, and the amount of streamed data implies a big data situation.

Diagram in Figure 2 shows the infrastructure involved in this experiment. As a broker to gather the data and store it in persistent database (No-SQL), Orion Context Broker was resorted, from FiWare, as described in Section 2.1. Orion would then feed our Elastic MapReduce infrastructure in Amazon Web Services (AWS), where all the corresponding calculations were carried out applying the method presented in Section 2.2. Finally,  $L_{eq}$  and  $L_{den}$  results are obtained with several analytics, as enunciated in Section 2.3.

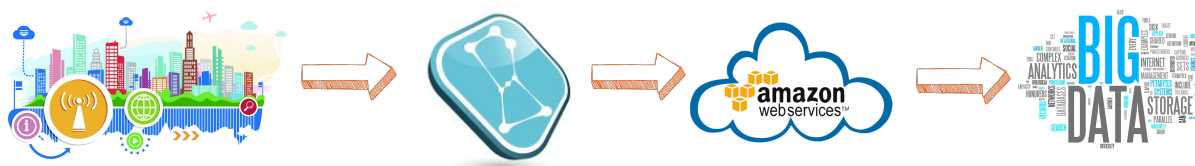


Figure 2 – Experiment infrastructure: 1) Smart Santander platform audio capture. 2) Orion Context Broker data storage. 3) Amazon Web Services, Elastic MapReduce infrastructure for calculation. 4) Results

### 3.2 Results and discussion

This section briefly presents some examples of results that can be obtained applying big data techniques to the infrastructure described in above section. After the map phase is finished, the shuffle phase classifies data by sensor and time period in order to carry out the reducer phase, so it can calculate and output analysed results.

Firstly, a summary of daily computed equivalent noise level of a particular node, in this case number 29 node, for day, evening and night intervals is shown in Figure 3. It should be notice that, sunday 10th of april shows the highest value for  $L_{eq,day}$ , and 12th of april has top value for night interval. Global  $L_{eq}$ , i.e. equivalent noise level integrated over ten days, for each interval is also presented in

the Figure 3, respectively, Leq, day is 77.56 dB(A), Leq, evening is 75.89, and Leq, night is 74.11 dB(A), resulting in a Lden of 81.22 dB(A).

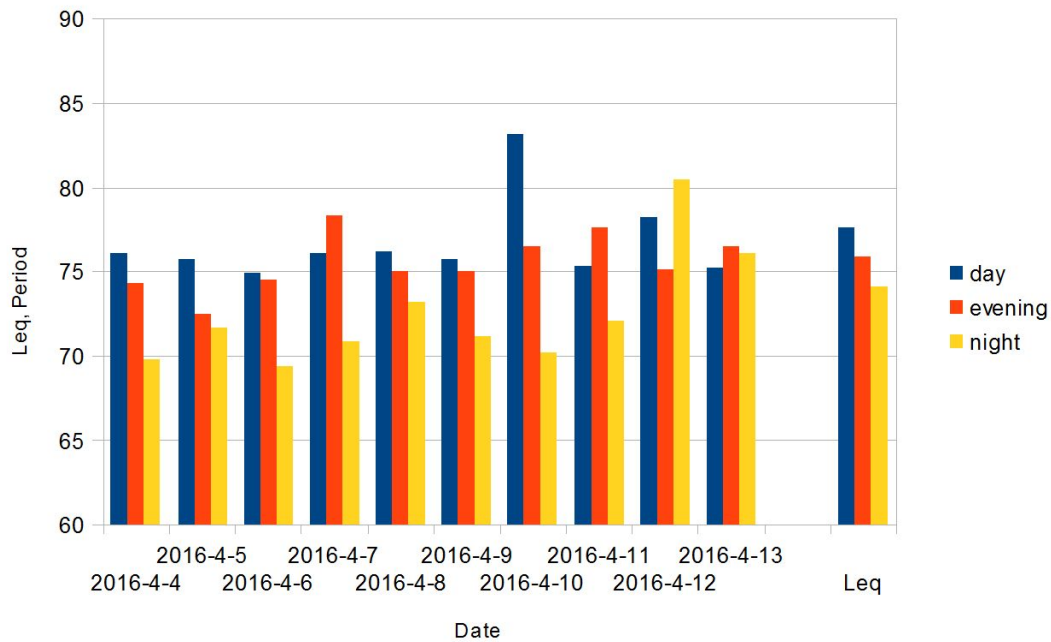


Figure 3 – Leq noise levels of node number 29 in each day of the 10 days experiment for three time intervals; day, evening and night.

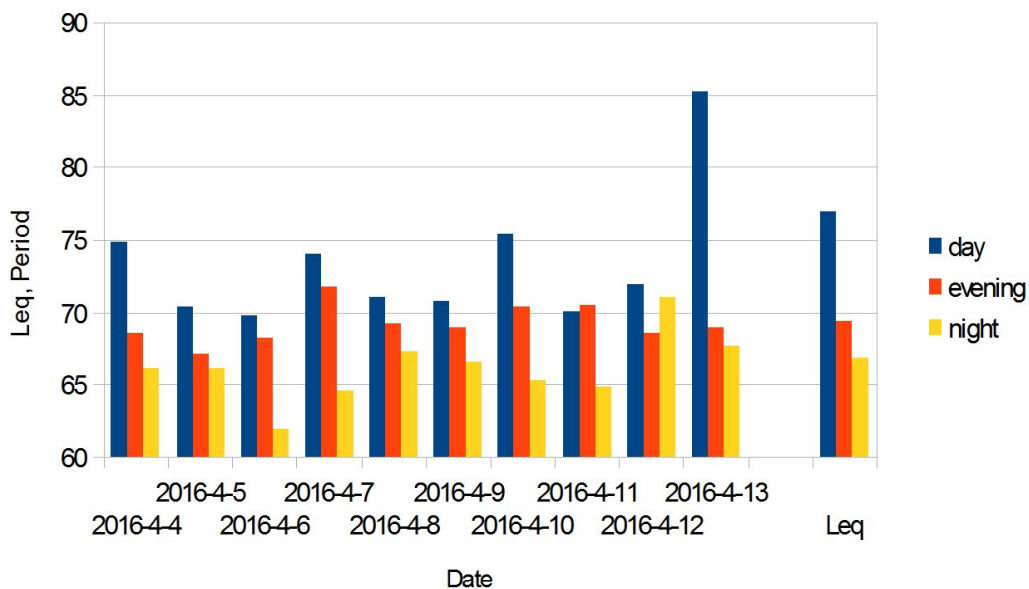


Figure 4 – Leq noise levels using data of all nodes in each day of the 10 days experiment for three time intervals; day, evening and night.

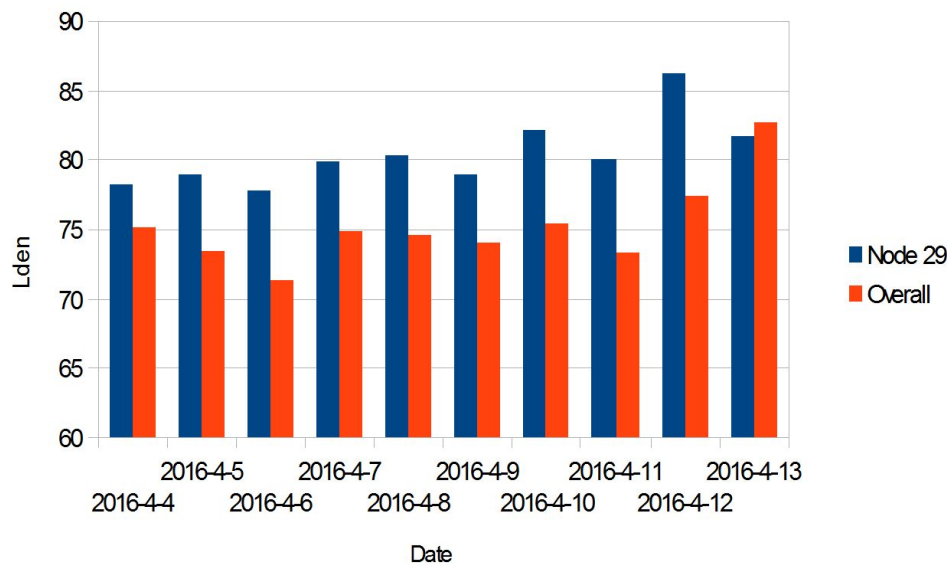


Figure 5 – Lden using data of node 29 compared with data of all nodes in each day of the 10 days experiment.

Then, applying the implemented MapReduce procedure, Leq noise level using data of all nodes can also be obtained, and results are graph in Figure 4. Date 13th of april shows the highest value for Leq,day, and 12th of april has again top value for night interval. Global Leq, is also presented in the Figure 4, resulting in levels lower than node 29 values, respectively, Leq, day is 76.94 dB(A), Leq, evening is 69.40, and Leq, night is 66.80 dB(A), resulting in a Lden of 76.56 dB(A).

Moreover, Lden can be calculated for each node and also for all nodes data together. Figure 5 shows comparison of computed Lden for each day between node 29 data and all nodes values. In general, node 29 presents higher values than the integration of all nodes, except for date 13th of april. The quietest day is wednesday 6th. Moreover, both situations reflect values of Lden over 65 dB(A), the recommended threshold for a city.

## 4 Conclusions

The high increase in the amount of collected noise level data by acoustic sensor network, both fixed and mobile, in smart cities is leading to the need to apply new analysis approaches. Big Data has techniques to analyse these data, extract useful information to help understanding the data and make better decisions. In this paper, an overview of various approaches of collecting, analysing and presenting large volumes of noise level data for environmental monitoring is presented. Three main approaches can be applied to capture acoustical data; sensor networks, mobile phones and open data from third parties platforms. This data is then stored and managed in a diversity of private-public smart cities platforms using a non-relational database, mainly, with no-SQL storage engines, that allow a better horizontal scalability and resilience. Finally, in this work, a MapReduce procedure is proposed to efficiently calculate Community Noise Equivalent Level, or Lden, from big data.





A real infrastructure is implemented to calculate Lden from noise node data of Smart Santander. Some examples of presenting results from a preliminary experiment with data from a 10-day period are described. The research findings so far are promising, but there are still many challenging research questions to be addressed. In the future, more research will be conducted in the following areas: increasing number of nodes and time interval, capturing data from other open data city platform, calculation other statistical acoustics indexes and representing and visualising output data providing insights through graphical means.

### Acknowledgements

This research is supported by the UCAM Catholic University of San Antonio under Project Ref. PMAFI-02-14.

### References

- [1] Conner, M. Sensors empower the "Internet of Things". *EDN (Electrical Design News)* 55.10 (2010): 32.
- [2] Atzori, L.; Iera, A.; Morabito, G. The internet of things: A survey. *Computer networks* 54.15 (2010): 2787-2805.
- [3] Su, K.; Jie, L.; Hongbo, F. Smart city and the applications. *Electronics, Communications and Control (ICECC), 2011 International Conference on* 9 Sep. 2011: 1028-1031.
- [4] Akyildiz, I. F.; Can Vuran, M. *Wireless sensor networks*. John Wiley & Sons, 2010.
- [5] Lynch, C. Big data: How do your data grow?. *Nature* 455.7209 (2008): 28-29.
- [6] Marx, V. The big challenge of big data. *Nature* 498.7453 (2013): 255-260.
- [7] Berglund, B.; Lindvall, T.; Dietrich, H. *Guidelines for community noise* (1999).
- [8] Vida, J. Valoración de la molestia por contaminación acústica mediante relaciones dosis-efecto. *CONAMA 8, comunicado técnico* (2006).
- [9] Prabahar, A. Development of high performance wireless sensor node for acoustic applications. *Green High Performance Computing (ICGHPC), 2013 IEEE International Conference on* 14 Mar. 2013: 1-5.
- [10] You, Y.; Jaehyun, Y.; Hojung, C. Event region for effective distributed acoustic source localization in wireless sensor networks. *Wireless Communications and Networking Conference, 2007. WCNC 2007. IEEE* 11 Mar. 2007: 2762-2767.
- [11] Segura, J.; Felici, S.; Cobos, M.; Torres, A.; Navarro, J. Psychoacoustic Annoyance Monitoring with WASN for Assessment in Urban Areas. *Audio Engineering Society Convention 138* 6 May. 2015.
- [12] Segura, J.; Felici, S.; Perez-Solano, J.; Cobos, M.; Navarro, J. Low-cost alternatives for urban noise nuisance monitoring using wireless sensor networks. *Sensors Journal, IEEE* 15.2 (2015): 836-844.
- [13] Jaimes, L.; Idalides, J.; Raij, A. A Survey of Incentive Techniques for Mobile Crowd Sensing. *Internet of Things Journal, IEEE* 2.5 (2015): 370-380.
- [14] Maisonneuve, N., Stevens, M., Niessen, M., and Steels, L. NoiseTube: Measuring and mapping noise pollution with mobile phones. *Information Technologies in Environmental Engineering* (2009): 215-228.
- [15] Rana R; Chou CT; Bulusu N; Kanhere S; Hu W. Ear-Phone: A context-aware noise mapping using smart phones. *Pervasive and Mobile Computing* 17 (2015): 1-22.

- [16] Kanjo, E. Noiseply: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications* 15.4 (2010): 562-574.
- [17] Kardous, C. A.; Shaw, P. B. Evaluation of smartphone sound measurement applications. *The Journal of the Acoustical Society of America* 135.4 (2014): 186-192.
- [18] Zhang, J.; Huang, K.; Cottman-Fields, M.; Truskinger, A.; Roe, P.; Duan, S.; Dong, X.; Towsey, M.; Wimme, J. Managing and analysing big audio data for environmental monitoring. *Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on* 3 Dec. 2013: 997-1004.
- [19] Mugagga, P.; Basajjabaka, K.; Winberg, S. Sound source localisation on Android smartphones: A first step to using smartphones as auditory sensors for training AI systems with Big Data. *AFRICON, 2015* 14 Sep. 2015: 1-5.
- [20] Schuller, B. W. Speech Analysis in the Big Data Era. *Text, Speech, and Dialogue* 14 Sep. 2015: 3-11.
- [21] Steele, C. A critical review of some traffic noise prediction models. *Applied acoustics* 62.3 (2001): 271-287.
- [22] Weigang, W.; Van Renterghem, T; De Coensel, B; Botteldooren, D. Dynamic noise mapping: A map-based interpolation between noise measurements with high temporal resolution, *Applied Acoustics*, Volume 101, 1 January 2016, Pages 127-140.
- [23] Can, A.; Dekoninck, L.; Botteldooren, D. Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation approaches, *Applied Acoustics*, Volume 83, September 2014, Pages 32-39.
- [24] Barontib, P.; Pillai, P.; Chooka, V.; Chessab, S.; Gottab, A.; Hua, F. Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and ZigBee standards. *Computer communications* 30.7 (2007): 1655-1695.
- [25] FIWARE NGSI9/10 RESTful binding, 2016
- [26] Galan, F. Orion Context Broker: introduction to context management FIWARE. 2015. <<https://www.fiware.org/2015/02/19/orion-context-broker-introduction-to-context-management-i/>>
- [27] NGSI-9/NGSI-10 information model - FIWARE Forge Wiki. 2015. <[https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/NGSI-9/NGSI-10\\_information\\_model](https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/NGSI-9/NGSI-10_information_model)>
- [28] SMARTSANTANDER MAPS. 2013. <<http://maps.smartsantander.eu/>>
- [29] Dean, J.; Ghemawat, S. MapReduce: simplified data processing on large clusters." *Communications of the ACM* 51.1 (2008): 107-113.
- [30] ISO: Acoustics -- Description, measurement and assessment of environmental noise -- Part 2: Determination of environmental noise levels. ISO 1996- 2:2007. Geneva, Switzerland, 2007.