Leveraging Urban Sounds: A Commodity Multi-Microphone Hardware Approach for Sound Recognition [†]

Fabrizio Tappero *,‡ [ID]**, Rosa Maria Alsina-Pagès**[ID]**, Leticia Duboc**[ID] **and Francesc Alías**[ID]

GTM: Grup de recerca en Tecnologies Mèdia, La Salle, University Ramon Llull, Carrer Quatre Camins 30, 08022 Barcelona, Spain; rosamaria.alsina@salle.url.edu (R.M.A.-P.); l.duboc@salle.url.edu (L.D.); francesc.alias@salle.url.edu (F.A.)

* Correspondence: fabrizio.tappero@salle.url.edu; Tel.: +34-932902455
† Presented at the 5th International Electronic Conference on Sensors and Applications, 15–30 November 2018; Available online: https://ecsa-5.sciforum.net.
‡ Current address: C/ Quatre Camins, 30, 08022 Barcelona, Spain

Abstract: City noise and sound are measured and processed with the purpose of drawing appropriate government legislation and regulations, ultimately aimed at contributing to a healthier environment for humans. The primary use of urban noise analysis is carried out with the main purpose of reporting or denouncing, to the appropriate authorities, a misconduct or correct a misuse of council resources. We believe that urban sounds carry more information than what it is extracted to date. In this paper we present a cloud-based urban sound analysis system for the capturing, processing and trading of urban sound-based information. By leveraging modern artificial intelligence algorithms running on a FOG computing city infrastructure, we will show how the presented solution can offer a valuable solution for exploiting urban sound information. A specific focus is given to the hardware implementation of the sound sensor and its multimicrophone architecture. We discuss how the presented architecture is designed to allow the trading of sound information between independent parties, transparently, using cloud-based sound processing APIs running on an inexpensive consumer-grade microphone.

Keywords: urban noise; sound identification; soundscape; acoustic sensor network; acoustic impact; multimicrophone; Amazon Echo

1. Introduction

Traditionally, urban noise is measured and analyzed with the purpose of drawing appropriate regulations [1,2] for a healthier urban environment [3]. As such, most urban noise monitoring activities are carried out with the sole purpose of reporting a misconduct (e.g., street noise, noisy bars) to the appropriate authorities or correcting a misuse of council resources (e.g., bus re-routing, city planning) [4]. Furthermore, noise data is often, if not always, collected and processed with dedicated expensive devices including high-end microphones [5,6]. The employment of costly hardware represents a limiting factor for the businesses that can be built around urban sound analysis [7]. As a result, the trading of sound-related information is almost invariably a measurement of absolute noise levels that sound experts provide to government-like institutions [8].

However, from urban sounds, it is possible to extract several other types of valuable information related to human activities (e.g., vehicle traffic, garbage collection, delivery of goods, gunshots, leisure activities, etc). If this information could be made available, in a easy-to-use manner and at a low cost, it could enable the development of several intelligent services and products in the urban environment.

We propose an urban sound classification system that is designed to enable the large-scale exploitation of urban sound. Our system intends to achieve this goal in the following ways:

1.　Instead of using proprietary and costly acoustic sensors, we propose to explore the use of a consumer-grade commodity hardware for noise capturing and measurement, as a means to drastically reduce the costs of sound exploitation systems.

　　With this respect, our idea is to follow a similar strategy to the successful driver-less car manufacturers [9]. They realised that if they could design an artificial intelligence solution that could run with inexpensive hardware (e.g., vision cameras), a driver-less car could become affordable. Before that, the driver-less car industry was very much heading towards vehicles equipped with expensive light detection and ranging (LIDAR) technology, heavily limiting the widespread and commercialisation of those vehicles.

　　In the personal voice assistant market, Amazon took a similar approach. They developed ALEXA, an AI solution that runs in fairly inexpensive US$100 multimicrophone-based device capable of talking and ,eventually, answer any question [10].

2.　Our system will offer sophisticate, yet easy-to-use, sound processing and artificial intelligent algorithms (AI) for extracting non-straightforward information from urban sounds. Such algorithms will only marginally depend on the capabilities of the sound capturing device. For that, the system will be heavily cloud-based (actually FOG computing), and the services will be available through web-based APIs. By doing so, the data collected by the consumer-grade sound capturing devices can be exploited by third-party applications, possibly through a pay-per-use business model.

　　The overarching goal of the envisioned idea, hopefully reflected in this paper, is to show that this approach can enable new socially responsible business opportunities in the area of urban sound analysis by enabling the creation of intelligent products and services for the city.

2. Urban Acoustic Event Detection Applications

　　There is a significant body of work focusing on Acoustic Event Detection (AED) in urban environments. In general terms, AED has been mainly driven by classification-based approaches. Among them, we should differentiate between one-class novelty detection oriented to the identification of events of interest from a majority class (e.g., background noise), and the multiclass classification, trained to detect a closed set of previously defined acoustic classes of interest (e.g., gunshots, screams).

　　Within the one-class approach, Ntalampiras et al. [11] assume the existence of an unbounded acoustic class to detect hazardous situations due to the local, occasional, diverse and unpredictable nature of this kind of sounds in a real scenario, instead of the classical bounded multiclass scenario. In [12], Nakagima et al. tackle a similar problem using a deep neural network sound recognition algorithm, using both real-life sounds and mixed artificially created sounds. Some researchers have concluded that, assuming that the unbounded class is a complex problem to solve by any machine-learning algorithm, the multiclass AED could take into account the signal to noise ratio of the events [12,13].

　　On the other hand, the multiclass AED has also provided good results in different domains of application for outdoor monitoring. For instance, in [14], a reliable detector of audio events in very noisy environments is described. And more recently, in [15] a two-class classification approach has proved to overcome one-class counterpart to detect anomalous events to tailor reliable road traffic noise maps within the DYNAMAP project [16]. In this context, valuable literature resources can be found in the proceedings of the DCASE annual event (Detection and Classification of Acoustic Scenes and Events), a competition that presents in each call different challenges, including several real-life ones. These challenges cover from acoustic scene classification [17] through the tagging of online contents from free sound to bird audio detection [18], among others. Among the different works presented to DCASE, it is worth mentioning the approach introduced in [19], based on the use of a multimicrophone architecture for acoustic event detection in indoor environments. This solution

takes advantage of the fact that the device is already installed at home. The work also argues the firm possibility that the same approach could give good performance in outdoor environments too.

Finally, let us note that specific challenges have to be addressed when employing microphone arrays for AED purposes [20]. As reported in [21], a grid-based method can be successfully leveraged for the localization of multiple sound sources within a Wireless Acoustic Sensor Networks (WASN) setup. Where, each sensor node, containing an array of microphones, can process the direction-of-arrival of a sound sample inside a processing node. For instance, in [22], the authors use a far-field random array to localise and separate several noise sources, by means of a two-stage noise source algorithm, mainly solved by the Tikhonov regularisation and the maximum likelihood estimation. The FISONIC is a FIWARE-based project focused on the continuous analysis of the urban environment with multiple microphones in a single sensor [6]. They are the only two projects that employ multiple microphones where signal processing is carried out in the cloud [23]. It is found that a proposal including an affordable platform with multi-microphone enables a substantial improvement in the accuracy of the AED algorithms deployed in the sensors of a WASN.

3. Proposed Urban Sound Measurement System

Sound processing is proposed as mean to allow the mining of urban sound for the generation of valuable derived information. For a successful exploitation, this information is then traded for a profit between parties. Seen in its whole, a successful urban sound processing system must be design maximizing the trading element of sound-derived information. Figure 1 gives an overview of the presented system.
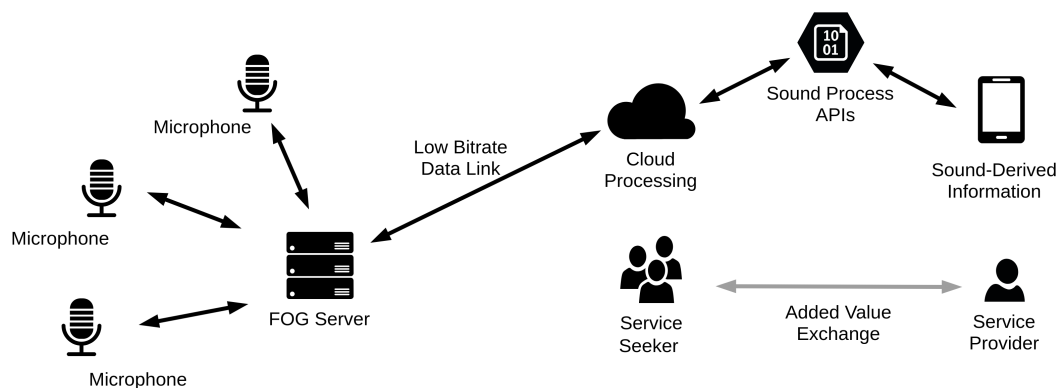

Figure 1. System Overview of the proposed solution.

The proposed solution is composed of a variable number of microphones all wirelessly connected to a local server located in the vicinity of the microphone network (FOG computing). The same server, responsible for part of the processing, is connected to a cloud service similar to an Amazon Web Service (AWS) where a second processing stage would take place. Here the actual sound processing algorithms are remotely employed via an API scheme. The existence of a sound processing API structure would imply the possibility of having multiple APIs providers, possibly competing among them for the best performance. Once the sound information is extracted, the trading of the latter and the establishing of the actual business service takes place on a more traditional and customizable environment (e.g., web page, smart phone). Here the client and service provider value added trading takes place.

It should be noted that the existence of an FOG processing unit allows for less stringent data rate requirement on the city network installation. Furthermore, the important issue of data privacy, typical of sound recording applications, is partially addressed by the use of certification exchange security on the edge computing platform.

4. Multimicrophone Sensor and Commodity Hardware

In 2015, Amazon commercialised their first intelligent personal assistant named Amazon Echo (https://en.wikipedia.org/wiki/Amazon$_$Echo). The challenge for Amazon was to design an artificial intelligence hosted in the cloud, later on named Alexa, that would allow an inexpensive multimicrophone unit to ideally understand and answer any voice question. As Echo sales show, Amazon greatly succeeded in this task.

As Amazon did with Echo, we propose a similar system structure where we host most of the sophisticated sound processing algorithm in a FOG computing platform (https://www.cisco.com/c/dam/en$_$us/solutions/trends/iot/docs/computing-overview.pdf) (refer to Figure 1). Freeing the microphones from the need of performing any heavy sound processing allows for the use of an inexpensive network of commercial-grade microphones. The possibility of employing inexpensive microphones would not only mean a lower overall cost for the system but would also mean the commercialization of urban sound sensors by non sound experts. A scenario much more convenient than the one where urban microphones are expensive professional grade devices, built and sold by fewer companies.

The Amazon Echo (docs.aws.amazon.com/a4b/latest/ag/a4b-ag.pdf) itself does represents a great example of commodity hardware that can be purchased or rented inexpensively. Prior a customization for outdoor use, the Amazon Echo is certainly one possible candidate for the sound sensor that we intend to explore in this research. Other types of commercial-grade microphones will be considered in the future. Amazon Echo is not only economical convenient, its multimicrophone architecture is certainly a promising feature that, if exploited well, could offer great additional advantages.

5. Business Opportunities

Frost and Sullivan's research estimates that the global Smart City market will be valued at US$2T by 2025, being AI a key driver and Europe the region with the largest number of smart city project investments globally [24]. The mining of urban sound is a rich source of information from which to develop services in the context of Smart City. Imagine, for instance, a high-end real-estate agency that wishes to offer a characterization of the soundscape of the properties in its catalogue as an added value to its customers.

In this case, the need for characterizing the soundscape is temporary and should include measures from several locations and over time (from days to a few weeks). To do so, the real-state agency needs to either install a network of dedicated high-end acoustic sensors or to hire a sound professional to make several manual measurements. Both options can be prohibitively expensive if the need for the measurements is not permanent.

We aim at allowing new businesses to exploit urban sounds within the scope of Smart Cities. The proposed technology allows organizations to leverage their services by taking advantage of noise information collected from a network of low-cost commodity sensors. Its main differentiating factor is that it won't depend on any existing infrastructure. Instead, since it relies on inexpensive commodity hardware, it allows the solution to be deployed, even temporarily, on different locations and conveniently mine noise/sound information and build services. Let us exemplify two scenarios that can illustrate the envisioned use of the proposed solution.

Noise City Management in Barcelona: Throughout the year, the Barcelona City Council receives many complaints that are difficult to deal with because they require the identification of the origin of the problem. This is often done by sending a sound technician to measure a specific location, which generally reveals a static and limited description of the problem. In such a scenario, the city council could install the proposed solution in the area of interest. For instance, a small sound network installed around the Sagrada Familia church could be used to determine what percentage of noise is coming from traffic, leisure activities, or urban services (e.g., garbage collection). Once the study is completed,

the sensor network would be taken down. With this information, the city council could take the necessary actions or create noise-limiting plans to protect the neighbours.

Security at Exhibitions or Festivals: Barcelona, as many major cities, hosts several international festivals (e.g., Mobile World Congress) that attracts yearly millions of people. Security is a paramount priority for the responsible authorities. State police and local law enforcement authorities do spend a great deal of energy and money to guarantee the maximum level of security for the attendees. Within this scenario, the organisations behind such a large exhibition would rent out the presented solution and, via its cloud, share the real-time audio information with the local authorities, who could be immediately alerted in case of a rise of the possible threats (e.g., gunshot sound or screams).

6. Conclusions

In this paper we have presented a urban sound processing system that employs inexpensive multimicrophone sensors together with a FOG computing architecture. The proposed system does represent a valuable solution for the data collection, processing and trading of urban sound. In a way, similarly to what Amazon did with their cloud product Alexa, we proposed to partially decentralize the sound processing task to a more agile FOG server. This will naturally free the consumer-grade sound sensors from the difficult task of processing, allowing noise-based services and products to be developed inexpensively. Additionally, the proposed system will allow all sound processing algorithms to be used remotely by any third party via convenient APIs. Lastly, the cloud structure of the proposed solution would facilitate the trading of sound-derived information between parties who are not actually expert in the field of sound signal processing. Directly fostering the business aspect behind urban sound exploitation.

References

1. Organization, W.H.; the Joint Research Centre of the European Commission. *Burden of Disease from Environmental Noise: Quantification of Healthy Life Years Lost in Europe*; Technical Report; World Health Organization, Regional Office for Europe: København, Denmark, 2011.
2. Union, E. Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise. *Off. J. Eur. Communities L* **2002**, *189*, 2002.
3. Goines, L.; Hagler, L. Noise Pollution: A Modern Plague. *South. Med. J. Birm. Alab.* **2007**, *100*, 287–294.
4. Seto, E.Y.W.; Holt, A.; Rivard, T.; Bhatia, R. Spatial distribution of traffic induced noise exposures in a US city: an analytic tool for assessing the health impacts of urban planning decisions. *Int. J. Health Geogr.* **2007**, *6*, 24.
5. Bartalucci, C.; Borchi, F.; Carfagni, M.; Furferi, R.; Governi, L. Design of a prototype of a smart noise monitoring system. In Proceedings of the 24th International Congress on Sound and Vibration (ICSV24), London, UK, 23–27 July 2017.
6. Paulo, J.; Fazenda, P.; Oliveira, T.; Carvalho, C.; Felix, 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.
7. Nencini, L. DYNAMAP monitoring network hardware development. In Proceedings of the 22nd International Congress on Sound and Vibration (ICSV22); The International Institute of Acoustics and Vibration (IIAV), Florence, Italy, 12–16 July 2015; pp. 1–4.
8. Tsai, K.T.; Lin, M.D.; Chen, Y.H. Noise mapping in urban environments: A Taiwan study. *Appl. Acoust.* **2009**, *70*, 964–972.
9. Villagra, J.; Acosta, L.; Artuñedo, A.; Blanco, R.; Clavijo, M.; Fernández, C.; Godoy, J.; Haber, R.; Jiménez, F.; Martínez, C.; et al. Chapter 8—Automated Driving. In *Intelligent Vehicles*; Jiménez, F., Ed.; Butterworth-Heinemann: Oxford, UK, 2018; pp. 275–342, doi:10.1016/B978-0-12-812800-8.00008-4.

10. Chung, H.; Park, J.; Lee, S. Digital forensic approaches for Amazon Alexa ecosystem. *Digit. Investig.* **2017**, *22*, S15–S25.

11. Ntalampiras, S.; Potamitis, I.; Fakotakis, N. On acoustic surveillance of hazardous situations. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 165–168, doi:10.1109/ICASSP.2009.4959546.

12. Nakajima, Y.; Sunohara, M.; Naito, T.; Sunago, N.; Ohshima, T.; Ono, N. DNN-based Environmental Sound Recognition with Real-recorded and Artificially-mixed Training Data. In Proceedings of the 45th International Congress and Exposition on Noise Control Engineering (InterNoise 2016), German Acoustical Society (DEGA), Hamburg, Germany, 21–24 August 2016; pp. 3164–3173.

13. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Audio Surveillance of Roads: A System for Detecting Anomalous Sounds. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 279–288.

14. Foggia, P.; Petkov, N.; Saggese, A.; Strisciuglio, N.; Vento, M. Reliable detection of audio events in highly noisy environments. *Pattern Recognit. Lett.* **2015**, *65*, 22–28, doi:10.1016/j.patrec.2015.06.026.

15. Socoró, J.C.; Alías, F.; Alsina-Pagès, R.M. An Anomalous Noise Events Detector for Dynamic Road Traffic Noise Mapping in Real-Life Urban and Suburban Environments. *Sensors* **2017**, *17*, 2323, doi:10.3390/s17102323.

16. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP – Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189.

17. 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. (TASLP)* **2018**, *26*, 379–393.

18. Stowell, D.; Stylianou, Y.; Wood, M.; Pamuła, H.; Glotin, H. Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *arXiv* **2018**, arXiv:1807.05812.

19. Dekkers, G.; Lauwereins, S.; Thoen, B.; Adhana, M.W.; Brouckxon, H.; Van den Bergh, B.; van Waterschoot, T.; Vanrumste, B.; Verhelst, M.; Karsmakers, P. The SINS database for detection of daily activities in a home environment using an Acoustic Sensor Network. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), Munich, Germany, 16–17 November 2017.

20. Bertrand, A.; Doclo, S.; Gannot, S.; Ono, N.; van Waterschoot, T. Guest editorial: Special issue on wireless acoustic sensor networks and ad hoc microphone arrays. *Signal Process.* **2015**, *107*, 1–3.

21. Griffin, A.; Alexandridis, A.; Pavlidiand, D.; Mastorakis, Y.; Mouchtaris, A. Localizing multiple audio sources in a wireless acoustic sensor network. *Signal Process.* **2015**, *107*, 54–67.

22. Bai, M.R.; Lo, Y.Y. A two-stage noise source identification technique using a far-field random array. In Proceedings of the INTER-NOISE and NOISE-CON Congress and Conference Proceedings, Hamburg, Germany, 21–24 August 2016; Volume 253, pp. 62–71.

23. 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; pp. 1–10.

24. Sullivan, F. Frost & Sullivan Experts Announce Global Smart Cities to Raise a Market of Over \$2 Trillion by 2025. 2018. Available online: https://ww2.frost.com/news/press-releases/frost-sullivan-experts-announce-global-smart-cities-raise-market-over-2-trillion-2025/ (accessed on 7 October 2018).