FRAMEWORK TO MONITOR SOUND EVENTS IN THE CITY SUPPORTED BY THE FIWARE PLATFORM

Joel Preto Paulo[1,2,4], Pedro Fazenda[1,3,4], Tiago Oliveira[1,3,4], Carlos Carvalho[1,4], Miguel Félix[3]

[1]Instituto Superior de Engenharia de Lisboa – ISEL, Lisboa, Portugal,
jpaulo@deetc.isel.ipl.pt
pedro.fazenda@etconcept.com
tiago.oliveira@etconcept.com
cfc@cedet.isel.ipl.pt
miguel.felix@fi-sonic.com
[2]CAPS - Instituto Superior Técnico, TULisbon, Lisboa, Portugal,
[3]ETConcept Engenharia de Sistemas, Lda., Loures, Portugal,
[4]CEDET – ISEL, Lisboa, Portugal,

Abstract
Acoustic maps represent an essential element of acoustic expertise and acoustic studies, which involve: planning of urban areas, spatial development plans, vehicle speed limits, rational management of noise emissions, etc. The emerging Internet-of-Things applied to Smart Cities concept can be utilized effectively for managing urban environments since it offers ubiquitous sensing and actuation. The acoustic sensors network is a mean of delivering in real time specific sound features capable of determining the location of sound sources for security and surveillance purposes or just giving tools and mechanisms for control the road traffic in order to maintain the maximum noise allowed in urban areas. The FI-sonic project consists of developing the necessary low-cost technology to capture sound, using both pressure microphones and accelerometers devices, to integrate existing frameworks in city environments. The project also includes processing this information with intelligent audio analytics, useful to update noise maps and identify and localize a set of different sound events. Localizing sound sources is of vital importance, and applications range from localizing sniper fire in urban scenarios and people in-distress. All this information is used to create quasi-real-time dynamic noise and event maps. The EU is endorsing a set of tools for the IT called FIWARE. The adoption of FIWARE technologies by several stakeholders, such as cites, security forces, will facilitate the integration of FI-Sonic products with the available IT infrastructures.

Keywords: Noise maps, Sound analytics, Internet-of-Things, sound source localization and event identification, FIWARE platform.

1 Introduction

Exposure of populations to excessive and long-term noise levels is known as having a negative impact on the quality of life of citizens. The European Union issued a directive (2002/49/EC), transposed to Portugal law by the DL 146/2006, to establish a method for dealing with environmental noise, called the Environmental Noise Directive, END. This directive states that the major city councils and road, rail, and aviation management offices need to construct noise maps periodically and should make available information about environmental noise and noise planning actions to the public. These maps must describe the likely environmental noise levels: (i) in cities with more than 100,000 inhabitants and 2500/km2, (ii) near all major roads with more than 3 million vehicle passages per year, (iii) major railways with more than 30,000 train passages per year and (iv) airports with more than 50,000 take-off and landing movements per year. In fact, the END only applies to the environmental noise, in particular in built-up areas, in public parks or other quiet areas in an agglomeration, in quiet areas in open country, near schools, hospitals and other noise sensitive buildings and areas.

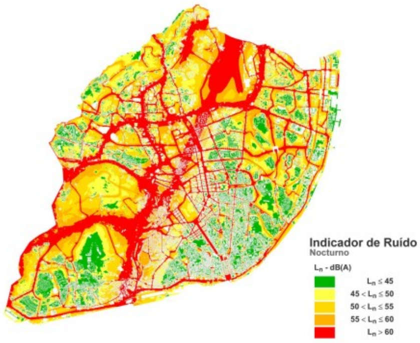

Figure1: Strategic noise map of Lisbon city for the night period, given by the Ln level.

Member states are obligated to develop strategies to improve the city environmental noise management. These strategies will create the necessity of continuous ambient noise monitoring and will have an impact on the corrective measures implemented by municipalities.

Nevertheless, the noise perspective not considered that is caused by: (a) the exposed person himself/herself, (b) noise from domestic activities, (c) noise created by neighbors, (d) noise at work places or noise inside means of transport, (e) due to military activities in military zones, etc..

The directive says that the public should be made aware of any noise assessment data, be consulted during the drafting of action plans and informed of any decisions taken.

Noise maps provide an insight into how environmental noise may be generated from a particular sound source such as traffic or rail transport. This information is not intended to be a precise calculation of the total noise exposure at any given point.

Community noise complaints are dealt with in a different way than environmental noise. Dealing with environmental noise involves looking at long term sources like roads and rail lines and needs longer term planning to manage the effects.

Municipalities should adopt a number of new policies to help manage environmental noise exposure throughout the city. This approach includes measurement of ambient sound quality to learn more about what citizens experience, plans for reducing noise exposure and ways of making all this information available to the public.

The most effective approach to manage noise involves building up a complete picture of the type of noise that is present in the community. This means learning about the levels, patterns and sources of noise so that a more effective plan can be made to reduce exposure. The best way to do this is by measuring the sound levels over a long period of time at places of interest implementing a network of environmental sound level monitors placed in locations that are of value to the community or give a good indication of sound levels from major noise sources.

Sound information acquired from the monitoring network can be used to generate plans for reducing noise exposure and for making better environmental policy decisions.

By other hand, noise is a community issue and the more useful information that is provided to the community, the more effective each citizen can be at helping to protect the environmental quality of their surroundings by helping to develop and then helping to implement better noise management policy. Thus, making available the information about environmental noise and planning to the general public using channels and media are both suitable and effective.

These initiatives are most effective when the information is shared with the community in a way that is accessible and easy to understand. Therefore, an Information Platform (website) is created to show all the sound level measurements and analysis gathered by the monitoring network, along with some useful information that is taken from the sound measurements. In short, these statements raise the awareness in the municipalities to the necessity of noise monitoring in the city environment.

Often, the council police forces face the problem of lack of information about community noise issues and exact location of noisy sources and there is the interest by the private security companies on the detection of trespassing of restricted areas due to the related installation and maintenance costs.

A problem that is presented to the police authorities relates to the night surveillance in bars and discos areas during and after the closure of these leisure places. For this reason, and to minimize complaints from neighborhood residents, the security features perform frequent rounds with these facilities, allocating human and material resources, which results in high costs for the public budget. Thus, a system with these features, by providing continuous information on the sound in those areas, can be an asset as it would save money to law enforcement agencies and security.

Frequently, there is a need for perimeter control to restricted areas, such as inside buildings or in outdoor open spaces. This supervision is usually performed by private security companies. Although there are many security systems used in this context (CCTV systems, microwave, infrared, etc.), the sound in FI-Sonic concept is not used. In fact, the system proposed here uses sound to extract relevant information about what is occurring *in situ*, identify types of events and its location.

Therefore, in the environmental context, there is a growing need to perform continuous acoustic measurements in the cities in order to help the authorities dealing with noise problems in deciding on the best mitigation and containment of environmental noise and community and involve populations in noise pollution issues.

Despite the fact of similar products are available on the market, they are much more expensive. In fact, in order to cover a broader area with accurate results, hundreds or even thousands of microphones should be placed in the urban area. Thus, the cost must be reduced for each unit.

2 Description

FI-Sonic is a network of sound acquisition nodes/stations, distributed geographically to cover a wide area in the city or private property. The system comprises several physical and logical modules for sound capturing and processing, and interconnectivity with the FIWARE platform. The FIWARE is an open initiative aiming to create a sustainable ecosystem to grasp the opportunities that will emerge with the new wave of digitalization caused by the integration of recent internet technologies. A multichannel microphone is the center piece of the acquisition system. Therefore, sound waves propagation direction can be estimated more easily and sound events detected, located and identified more accurately in the 3D sound field, as shown in Figure 2. The main goal of this project is to build a scalable and upgradable low-cost monitoring system to analyze the city sound environment in order to identify and detect events based on sound, like gun firing and people in distress, or to assist municipalities departments of environment on the decision process of action plans to improve the quality of life of the populations. We also intend to provide advanced analytics taking advantage of FIWARE platform on a subscription basis.

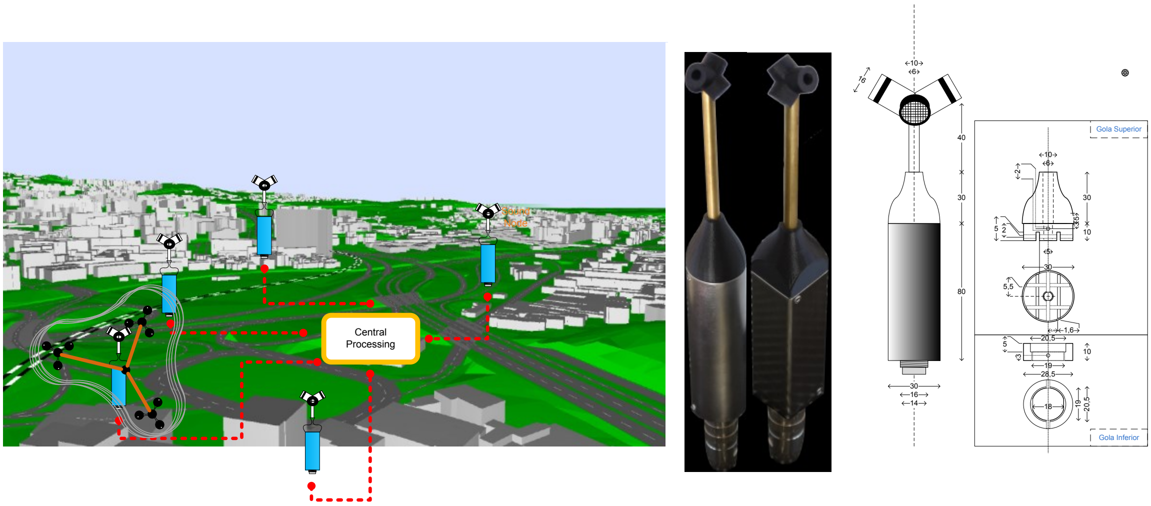

Figure2: FI-Sonic project design overview.

FI-Sonic is a low cost system and uses innovative techniques applied to event detection, localization and identification of sound sources in the city. By using recent hardware platforms and because all production is almost made by ETConcept startup the system cost is largely reduced. Our value proposition is: Dynamic sound and event maps, with event report; State-of-the-art audio analytics and data visualization; Remote configuration and calibration; scalability to include additional sensors types.

The FI-Sonic project has a Business to Business (B2B) perspective. The main market we are addressing is the European municipalities referred in the Environmental Noise Directive (2002/49/EC). According to our research at least 445 cities in 16 countries are covered by the Directive. Nevertheless, the Americas is also a promised market, Mexico, Brazil, Chile, Canada and US joint to the FIWARE platform already. With increasing implementation of sound monitoring systems and the sharing of experiences between municipalities, the awareness about the importance of the sound quality in cities will cause an increase on the necessity of such a system even for the municipalities that are not obliged by European Union directive. European governments with focus on their public safety forces and civil protection are also a potential market for our solution taking advantage of our sound event localization and identification. Multinational private security companies with engineering services to provide access and perimeter control are also a market we are considering. We identify as key customers Portuguese municipalities of major metropolitan areas in order to attest the advantage of our solution and create showcases to support the internationalization to other countries.

The FIWARE platform complements the FI-Sonic project by providing a cost-effective, mainly open source, and community supported set of tools and resources that are valuable for FI-Sonic to deliver and monetize the services associated with audio information by the network of Sound Acquisition Modules.

The adoption of FIWARE technologies by several stakeholders (cites, security forces, etc.) will facilitate the integration of FI-Sonic products with available IT infrastructures [1]. A possible methodology consists of using the Kurento WebRTC media server to group and process the audio content delivered by the different distributed FI-Sonic audio acquisition modules. WebRTC is an open source technology that enables web browsers with Real-Time Communications (RTC) capabilities via JavaScript APIs. It has been conceived as a peer-to-peer technology where browsers can directly communicate without the mediation of any kind of infrastructure. This model is enough for creating basic applications but features such as group communications, media stream recording, media broadcasting or media transcoding are difficult to implement on top of it. Each module acts like a WebRTC client with live audio streaming of each channel. However, this approach is very dependent of the available network bandwidth. Therefore, at this stage of the project, we focused our work on developing each sound capture module with capabilities of performing all the necessary computations locally. However, we are aware that this approach can compromise its performance and more simplified algorithms, less time consuming, should be used in order to keep it in a predefined level. Thus, some heavy computing signal processing algorithms should be simplified or less treads could be running simultaneously. The module connects to the Orion Context Broker and produces continuous context updates with relevant audio information. These updates are then consumed by users that use this information to build/update dynamic noise maps or to collect information from other subsystems such as video surveillance or CCTV systems.

3 System Architecture

The architecture of the FI-Sonic sound acquisition module, represented in Figure 3, is composed by a set of microphones, a multichannel acquisition card, the main processing unit, and the network interface.

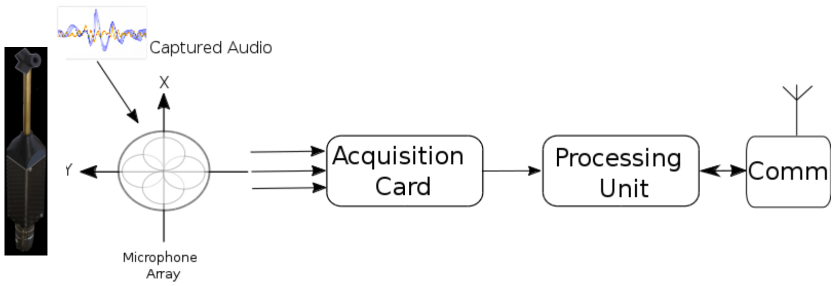

Figure3: FI-Sonic Sound Acquisition Module (Architecture).

Processing Unit and Operating System

Although the architecture is planned to not have any component tied to a specific vendor or card, we are currently using Edison (from Intel), for the main processing unit. Edison, illustrated in Figure 2, includes two Atom Silvermont cores running at 500MHz and one Intel Quark core at 100MHz. It has 1GB of integrated RAM, and 4 GB eMMC Flash on board, Wi-Fi, Bluetooth and 4 USB controllers. Edison runs Yocto Linux with development support for Eclipse (C, C++, Python).

Figure 4: Intel Edison (Top board), as the main Processing Unit.

Communications

The financial costs associated with choosing the appropriate processing platform for computation according to processing capacity, RAM, Flash and other features, is marginal in the final price of our product. On the other hand, communications are directly associated with running costs which can scale to unsustainable values when using a network with a significant number of acquisition nodes spread throughout the city. These costs depends on the infrastructure that our clients have access to and subscription contracts to communication services (such as LTE(4G) or Fiber). Costs can be associated with the amount of data traffic, and therefore the communication module must be specifically pondered considering this fact.

Depending on the amount of available bandwidth (we are assuming that bandwidth and amount of transmitted data is limited), we have different firmware solutions. In the most limited scenario we assume that the acquisition module only acquires meta-information about the audio such as sound levels and related measures, such as, max, min, rms, std values, spectrum information (wideband and by frequency bands), onsets, and other features, such as, semantic information that can be transmitted as a single vector to a centralized data base to be used by various maker stakeholders.

Firmware

Most of our current developments of the FI-Sonic Sound Acquisition Module are currently focused at the low level firmware level. The firmware is being programmed using C++ (for performance in signal processing) with the application running on the Yocto Linux operation system. For the sake of system modularity, audio acquisition uses the Advanced Linux Sound Architecture (ALSA) to be able to support different audio interfaces. We have been working on setting-up the development environment, configuring the operation system, and configuring ALSA to use a custom-built (still under development) acquisition card. Using the ALSA Library API we are acquiring several audio channels connected to the microphone array. Future work includes processing this audio signal for extracting features at a programmed sampling rate, determined depending on network connection limitations.

FIWARE Platform

The FIWARE Platform complements the FI-Sonic project by providing a cost-effective, mainly open source, set of tools and resources that are valuable for FI-Sonic to deliver and monetize the services associated with audio information by the network of Sound Acquisition Modules. The adoption of FIWARE technologies by several stakeholders (cites, security forces, etc.) will facilitate the integration of FI-Sonic products with available IT infrastructures.

Sound Bite Processing and Delivery - FIWARE architecture using Orion

We are considering two different design strategies. Considering the most probable scenario where FI-Sonic acquisition modules are deployed with constricted communication capacity, most of the audio processing will be executed locally at the Processing Unit. Audio acquired by the microphones is analyzed and features are extracted for the audio into a vector containing several data element attributes (<name>, <type>, <value>). This vector corresponds to context information that is periodically produced and sent to a Orion Context Broker (CB) Generic Enabler (GE). This audio feature information is then distributed by the CB to different subscribing applications as shown in Figure 5.

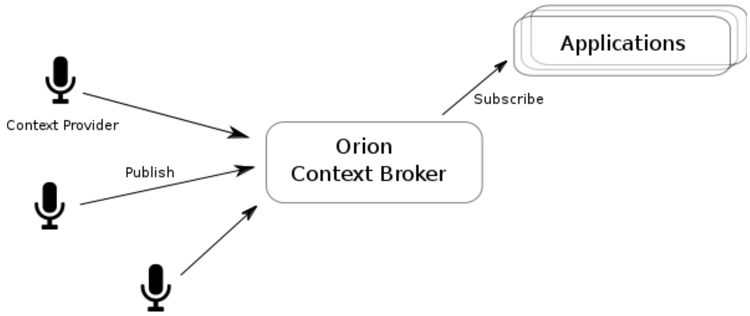

Figure 5: Audio Context Production and Application Subscription via Orion.

This basic architecture allows FI-Sonic to distribute context information according to a predefined sampling rate. This is used to distribute and store important audio features such as the amount of noise at a given location over time.

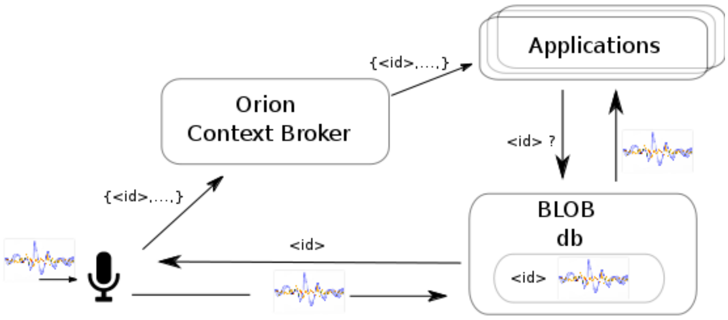

Figure 6: Processing events and saving sound bites to a binary large object (BLOB) storage database.

A refinement of this solution (that requires more bandwidth) includes having triggers that are fire based on features extracted from the audio stream (for example a loud sound). With the occurrence of these events it is important to save the raw audio sound bite and store it using a binary large object (BLOB) storage database (e.g., CRATE, https://crate.io/). Audio blobs are uploaded to the BLOB database and the id of the blob is send as a context element attribute to the CB, as shown in Figure 4.

Audio Processing Pipeline (FIWARE architecture using Kurento)

Considering a network with less constricted communications, an alternative architecture can be considered for sound acquisition and processing using Kurento. With the advantage of having less expensive (requiring less computational resources for the Processing Unit) audio acquisition modules, most of the computation can be delegated to a centralized computer running the Kurento Media Server. Kurento is a WebRTC media server with advanced media processing capabilities. Its modular architecture makes simple the integration of different media processing algorithms. A chain (media pipeline) of computational blocks (media elements) can be used to process the audio stream as shown in Figure 5.

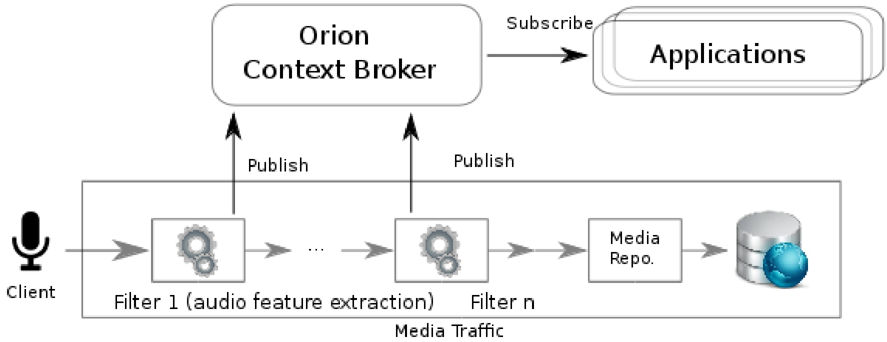

Figure 7: Media Pipeline using Kurento Media Server.

Each computational block extract a set of audio features and updates the corresponding context with the CB. This solution presents a higher level of flexibility because the pipeline can be very easily managed. New blocks can be added/removed and updated at any time as opposed to the configuration where the audio processing is done locally at the acquisition node. The Kurento solution also presents the possibility e of having the entire audio saved on a media repository, which can be useful for future reference (for example analyzing an audio event for some security application).

WireCloud, mashups and the Data Market

The applications illustrated in Figures 3, 4 and 5 can be tailor made for different stakeholders (security forces, city control center, etc.). Information on acquired audio data can be made available and sold on online markets. Through the WireCloud Mashup Platform we can use audio information (connecting to the CB) to create mashups, take them to the marketplace and sell them on our store reaching a broader audience of consumers. The already available infrastructure WStore GE provides us with the ability to manage our offerings and sales, including the charging of the customer.

Current Work

Currently, our developments are being focused at a very low (Firmware) level of our Sound Acquisition Module. Future work includes connecting our firmware to the CB, which we have already installed on a server using AWS (amazon web services). We are planning to transition to FI-Lab as it becomes more stable.

4 Methodology

The research and development of this system is multidisciplinary and requires its development in several phases. Therefore, the project is split in three phases, essentially. The first phase consists of developing all the necessary hardware and software for the measurement of the noise levels (sound monitoring stations, FIWARE processing, communications). The second concerns about signal processing for location and identification of sound sources (developing algorithms for sound propagation, feature extraction/selection and statistical classification). The last phase is about installing a pilot system for validation the proof of concept.

Although different localization methods could be used in this approach the energy-based localization method is adopted which employs measurement of the sound intensity vector. The first-order ambisonic microphone allows the estimation of the direction of sound source and by using more than one unit an estimate of the exact position of the sound source is possible [2-6]. Accuracy of triangulation depends on the precision of the angle estimation. The estimation of localization of the sound source with the use of the ambisonic microphone is made at first stage in laboratory conditions for maximal accuracy and for calibration purposes and later in real environment condition at the time of the pilot system validation.

For the identification of sound events responsible for the relevant sound sources at each time a number of types of sounds are considered for the analysis in order to cover most of the situations in urban environments. These types of sounds are people screaming or talking loud, shot guns, animals (dogs, birds, etc.), different types of vehicles noise (light and heavy cars, motorcycles, airplanes, train compositions, etc.), horns and road accidents.

In order to identify sound sources, at this early stage of research, two different types of classifiers are used. The Bayesian classifier corresponds to a quadratic discriminant analysis (QDA) and the Neural Network classifier is a feed-forward network with tan-sigmoid transfer functions for both the hidden layer, which uses 9 neurons, and output layer. An N-dimensional density model (where N is the length of the feature vector) is then adjusted to each class, based on the training set. One first advantage of using the Neural Networks against Bayesian classifiers is that the assumption of normal distribution for the selected features is not required.

Four distinct feature extraction stages are to be compared to evaluate their relative performance while using the same classifier stage. The main feature sets are: (1) low level signal properties; (2) MFCC; (3) psychoacoustic features including roughness, loudness and sharpness; and (4) an auditory model representation of temporal envelope fluctuations [7-8].

The feature vectors were divided into two sets: a training set and a test set. A randomly chosen subset of 30% of the vectors is assigned to the training set and the remaining 70% correspond to the test set [9-11].

A pilot system, consisting of the measurement of noise levels, localization and identification of sound sources, will be installed in situ to allow the validation of the all concept.

5 Conclusions and FIWARE Recommendations

FIWARE provides FI-Sonic with an appropriate set of APIs useful to deliver audio-related information captured by a network of audio acquisition modules. It is being adopted for the information technology infrastructure in different cities (for Smart Cities), as an open solution to integrate different services from different vendors and providers. It gives FI-Sonic the ability to take its information to market and share it with other stakeholders.

We recommend that extra efforts should exist towards normalizing context messages used with Orion Context Broker for several different recurrent domains (audio, measures from diverse sensors, etc.).

We recommend taking advantage of the extensive work already being done with ontologies for knowledge representation and extend the terminology for messages exchanged with the context broker. See, for instance, the Audio Features Ontology Specification http://motools.sourceforge.net/doc/audio_features.html)

The core concept of FI-Sonic project is the Sound Acquisition Module interconnected to a network, using the FIWARE platform. Currently, our developments effort is being focused on the Acquisition Module at a very low level (firmware). In the near future the work includes connecting our firmware to the CB, which we have already installed on a server using AWS (amazon web services). The physical part of the module, the hardware, consisting of a set of microphone capsules and the structure where they are fixed to and the body and the signal conditioning electric circuits (mic pre-amplifiers and power supply) is almost done.

In order to make available the proposed project to the public, a website for the FI-Sonic was created, http://fi-sonic.com/, covering the more relevant topics such as Description, General Overview, Applications and Achieved Results to name a few.

This website serves multiple purposes and can have multiple functions for both us and potential clients and to follow the evolution of the project. In fact, this material works for us as guidelines to structure our ideas in a visual and contextual perspective. For the people outside the project, potential clients and partners, sponsors, investors, etc., certainly, it will be more comfortable in viewing the undergoing of the works. We believe it is a powerful tool to have more realistic feedback from people interested in the project.

Since the project was recently initiated the results are preliminary.

Acknowledgements

The FI-Sonic project, research and development, has been supported by Soul-Fi, Startups Optimizing Urban Life with Future Internet, (Soul-fi is a FIWARE accelerator funded by the European Commission, FP7, aiming to grow web-based, innovative solutions (open data and crowd sourced data) for smarter urban life of EU citizens (topics are quality of life, mobility, resource efficiency, tourism)), acting as a facilitator to deliver innovative and promising products to the market.

References

[1] https://www.fiware.org/

[2] Michael Gerzon. 1975. The design of precisely coincident microphone arrays for stereo and surround sound. 50th Audio Engineering Society Conference, (AES preprint 8083L20)

[3] Dimoulas C., Avdelidis K., Kalliris G. andPapanikolaou G., “Sound Source Localization and B-Format Enhancement Using Sound Field Microphone Sets”, in 122nd AES Convention,Preprint Number: 7091, May 2007

[4] J.H. Wang, C. Chang, “Identification of the number and locations of acoustic sources”, Journal of Sound and Vibration, Vol. 284, no. 1-2, pp. 393-420, June 2005.

[5] Dimoulas C. A. 1, Kalliris G. M. 2, Avdelidis K. A. 3 and Papanikolaou G. V.. “Improved localization of sound sources using multi-band processing of ambisonic components”. Proc. 126nd AES Convention, Audio Engineering Society, 2009

[6] J. Wierzbicki, P. Malecki and J. Wiciak. “Localization of the Sound Source with the Use of the First-order Ambisconic Microphone” ACTA PHYSICA POLONICA A. Vol. 123. nº6. 2013.

[7] Courtenay Cotton & Dan Ellis, “Spectral vs. Spectro-Temporal Features for Acoustic Event Classification,” submitted to IEEE WASPAA, 2011

[8] Andriy Temko & Climent Nadeu, “Classification of acoustic events using SVM-based clustering schemes,” Pattern Recognition 39(4): 682-694, 2006

[9] Bugalho, M., Portelo, J., Trancoso, I., Pellegrini, T., Abad, A.: Detecting audio events for semantic video search. In: Proc. Interspeech 2009

[10] Alan Dufaux. “Detection and Recognition of Impulsive Sound Signals”. Institute of Microtechnology, University of Neuchatel, Switzerland. PhD thesis 2001

[11] Li, D., Sethi, I. K., Dimitrova, N. and McGee, T. Classification of general audio data for content-based retrieval, Pattern Recognition Letters, 22, 533-544, (2001).