2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing

Enhancing Wireless Sensor Networks with Acoustic Sensing Technology: Use Cases, Applications & Experiments

Danilo Hollosi, György Nagy, René Rodigast, Stefan Goetze
Fraunhofer Institute for Digital Media Technology - IDMT
Project Group Hearing, Speech and Audio Technology
Department Acoustics
Oldenburg & Ilmenau, Germany
Email: Danilo.Hollosi@idmt.fraunhofer.de

Philippe Cousin
Easy Global Market
Sophia Antipolis, France
Email: Philippe.cousin@eglobalmark.com

***Abstract*—This paper presents research carried out within the EU FP7 EAR-IT project, which is working on the challenges of bringing acoustic sensing intelligence to large-scale indoor and outdoor wireless sensor networks, i.e. into two existing testbeds out of the EU FP7 FIRE projects SmartSantander and Hobnet. Besides the benefits by integrating machine-learning based acoustic sensing technology, the general deployment approach of the so-called Acoustic Processing Unit, an embedded device whose capabilities go beyond state-of-the-art IoT sensors, as well as the EAR-IT indoor and outdoor use cases are described. This includes hardware qualification for the applications such as energy efficiency of buildings, traffic monitoring and emergency vehicle detection and tracking outdoor as well as acoustic emergency detection for indoor environments and further, a detailed description of the individual Acoustic Processing Unit software components. Latest efforts and simulation results for acoustic source localization using audio sensing technology for wireless sensor networks are presented, indicating that a more intelligent usage of the audio modality enables a wide range of applications and services with high social and technological value.**

I. INTRODUCTION

In the past few years, large-scale wireless sensor networks gained a lot of attention due to their huge potential to enable innovative services and applications with high social, technological and economic impact. Long-term research roadmaps laid out in this context following the Internet-of-Things (IoT) concept investigate questions related to interoperability, sustainability, communication infrastructure, data handling and possibilities towards global and ubiquitous sensing by incorporating multimodal wireless sensors. This led to the deployment of large-scale wireless sensor networks, so-called test beds, throughout various cities or in buildings across the globe allowing for use case development, hands-on experiments and the integration and verification of real-life application scenarios. Two important and powerful test beds - besides others - have been established within the EU FP7 projects SmartSantader [26] in the city of Santander, Spain for outdoor environments and Hobnet [27] in the city of Geneva, Switzerland for indoor environments. Wireless sensors of various kind have been deployed already in the mentioned test beds, e.g. for measuring light, CO_2, humidity, etc., however, the potential benefit of the audio modality remains widely un-investigated although coming with obvious advantages.

Audio sensors are cheap, energy efficient and often easy to deploy, do not depend on a line-of-sight (NLOS), allow for omnidirectional sensing and are basically independent from weather conditions and lighting situations. The advantage of NLOS acoustic sensing compared to video cameras is its quasi-independence from the sensor position, while providing the possibility to see through obstacles and not being limited to a certain viewing angle. This is important, as many sensors may need to be added into the environment in order to have a full coverage of the area of interest whereas one could use much less sensors and less complex sensing solutions if deeper incorporating the audio modality. The acoustic sensors (i.e. microphones) are furthermore multipurpose by definition. They not only capture relevant environmental information (through the sound) and provide physical measures, e.g. loudness or direction of sound, but also allow an identification of specific events within the audio stream if equipped with a reasonable amount of processing power. Hence, once deployed, its intelligent sensing capability on a modular software level together with communication capabilities makes them very interesting devices also within the IoT context. With the increasing processing power and networking capabilities of IoT end devices nowadays, it is possible to exploit audio data for a broad range of applications. Although the untapped value of audio data is still to be revealed, research activities and projects already exist on, e.g. energy efficiency [25], traffic monitoring applications [6]- [7] and emergency detection for care environments and hospitals [28], thus implying great potential on intelligent audio based solutions to support a myriad set of applications.

The possibility and value to integrate intelligent audio based technology into existing test beds is investigated in the EU FP7 project EAR-IT [24]. There, two complementary areas of use cases are addressed, namely outdoor and indoor use cases, targeting applications like monitoring, security as well as traffic and environmental control (see Figure 1). All use cases make use of the audio modality on a low level (e.g. loudness measures), on an intermediate level (acoustic events) and on a high level (temporal modeling of events to get access to contextual information) as well as the notification of this information to the network. Over the course of the project, a direct interaction between this new type of sensor and already deployed IoT devices will be further investigated.

978-0-7695-5046-6/13 $26.00 © 2013 IEEE
DOI 10.1109/GreenCom-iThings-CPSCom.2013.75

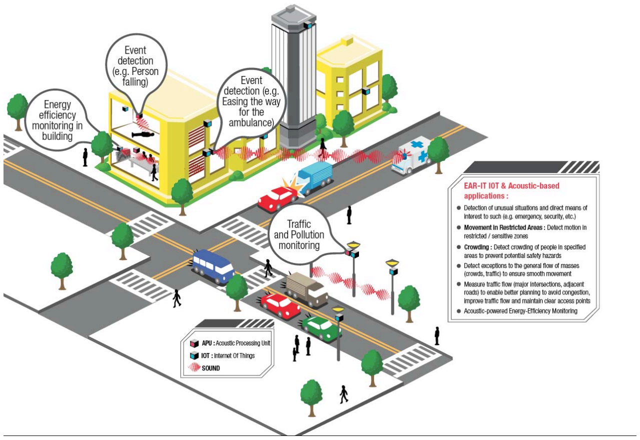

Fig. 1. EAR-IT application scenarios

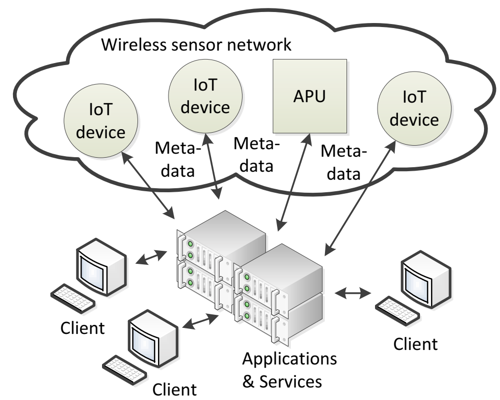

Fig. 2. System overview of a smart acoustic-enabled object within the EAR-IT project

This paper presents current research and activities to combine acoustic sensing technology with large scale wireless sensor networks aiming to provide situation aware innovative services potentially useful for indoor and outdoor environments. The paper is organized as follows. First, the general research approach, its challenges and the EAR-IT use cases are reviewed in Section II. In Section III, embedded processing and sensor hardware qualification for the envisioned use cases is presented, followed by a detailed description on the Acoustic Processing Unit (APU) as a new device introduced to the wireless sensor network in Section IV. This includes details on the overall system design, audio signal preprocessing, acoustic event detection, and statistical modeling and metadata generation using the APU. Afterwards, latest efforts and simulation results for acoustic source localization using audio sensing technology enhanced wireless sensor networks are presented in Section V. The paper concludes with first related project results and an outlook on future activities in Section VI.

II. Enhancing Wireless Sensor Networks with Acoustic Sensing Technology

A. General Approach

In order to enhance existing wireless sensor networks with acoustic sensing technology, we introduce a new set of devices that go beyond the capabilities of the already deployed sensors, the Acoustic Processing Units (APU). An APU is a computational powerful embedded device, which is still cheap and small and therefore, can be counted the IoT device family. In general, it consists of a microphone unit and an embedded processing board. Which components was chosen for the envisioned use cases is evaluated in Section III. On the APU, intelligent algorithms will analyze captured audio data, perform pre-processing such as filtering and de-noising where necessary and identify target acoustic events using machine-learning based acoustic event detection algorithms. Detailed information can be found in Section IV. The output of the APU is capsuled into a predefined metadata container which is sent over the network instead of raw audio data. A server that holds an application modeling environment, i.e. intelligent algorithms and processes, will allow to develop new innovative services combining the new information coming from e.g. APUs and lower level IoT devices (e.g. Libellium Wapsmotes, AdvanticSys motes) while offering the possibility to modify the APUs to work together due to the availability of bi-directional communication interfaces. Within EAR-IT, the server will be used for data visualization, database handling and controlling of the use case dependent peripherals. Although the APU is a dedicated device at the current state of the project (as indicated by Figure 2), it will become more integrated within the IoT (wireless) network later on.

B. EAR-IT Usecases

1) Indoor Usecases: Environment and Comfort: Speech command recognition functionality will be integrated to the Hobnet test-bed via Acoustic Processing Units (APU) to interact with the environment in a natural way. Here, a focus lies on the comfort of the user. The APUs will be installed, e.g. in the accommodation rooms or office environments and hold a predefined speech vocabulary to interact with the environment, i.e. to control lighting, air-conditioning and heating via speech commands. The recognized speech commands will be transcribed into metadata and will be transmitted to a central service platform. Unique identifiers of the sensor, time stamps, etc. will provide the necessary basis of data to actively change the environmental parameters in the desired areas. In a second phase, already existing IoT sensors and their data will be taken into consideration. The service platform will then be able to fuse information from multiple sources to create a contextual and also personalized user profile (define a modus and activate it through a keyword) of the environment (in terms of preferred temperature, humidity, brightness level, etc.). This could be used to either prepare an environment for an end-user prior to its arrival (if announced) to address for comfort, or to automatically select the cheapest energy provider for a given point in time.

Security & Monitoring: Acoustic event detection capabilities will be integrated in the Hobnet infrastructure to account for security and monitoring related applications. Research in the past already revealed a high demand for these kinds of application scenarios for care homes and hospitals [28], where improved safety of the patient and a higher efficiency in residential and non-residential care are important issues. Similarly, the acoustic event detection functionality to be implemented into an APU and to be integrated in Hobnet within the EAR-IT project will inform personnel and staff about occurring emergencies or at least provide them with additional contextual information about a security related incident. Since Quality of Service (QoS) in terms of data transmission delays, time synchronization of sensor data coming from distributed sensors are of high importance within this use case, the capabilities of the existing Hobnet infrastructure will be investigated. Furthermore, the tradeoff between end user privacy and benefit of additional security related applications and their potential will be investigated in this use case. In [12] a pre-processing mechanism for acoustic event detection that is capable to account for privacy concerns was already presented to technologically address this issue.

Energy Efficiency: In this use case, it will be investigated how acoustic sensing can be used to detect the presence of people and their approximate number within certain areas in the buildings and potentially, distinguish between different actions to create contextual information. Knowledge about related services and applications coming from the EU FP7 project S4ECoB [25] will be transferred to this use case. Detected acoustic events in an office scenario like keyboards, phone ringing, door open-close, speech, etc. indicate the presence of end-users in a room or around a certain area. The events will be forwarded to a centralized service platform and further fed to a building management system. If no presence is detected over a certain period of time, lighting, heating and air-conditioning will be switched of automatically. This functionality directly complements the Environment and Comfort use case, where the end-user is able to adjust the environmental settings to his needs via speech commands.

2) Outdoor Usecases: Audio Based Traffic Density Monitoring: The European Union makes a lot of effort to reduce the environmental noise. Therefore, the EU Directive 2002/49/EC [29] provides guidelines for the assessment and management of environmental noise introduced by railways, aircraft and traffic, which was found to be the main contributor to noise pollution [32]. Studies in the past revealed that approximately 30% of the people exposed to environmental noise are annoyed by aircraft noise, about 20% by road traffic noise and about 10% by rail traffic noise respectively, resulting in a decreased quality of life, health, mood and increased stress levels [30]. Even though the World Health Organization (WHO) proposes to limit the noise level to 55 dB(A) (serious annoyance), about 44% of the population of the EU25 (over 210 million people) were exposed to road traffic noise levels above this limits and more than 54million people were exposed to noise levels exceeding 65 dB(A) [31]. With the deployment of wireless sensor networks in Smart Cities, new possibilities towards traffic density monitoring and quantification became possible. Currently, video cameras, seismic sensors, ultrasonic detectors, inductive loops, magnetometers etc. are in use [26]. Unfortunately, these sensors do not take subjective factors in terms of noise perception by people into account. Within EAR-IT, the possibilities to incorporate the acoustic modality into traffic density monitoring applications are investigated in detail. Besides purely physical loudness level measures by making use of the distributed wireless sensors already available in the test bed, noise type classification, its quantification and assessment on a subjective level using the APU becomes possible. By incorporating this new kind of information into existing data management systems, the development of more reliable noise maps incl. their historical progression become possible and more robust parameters can be derived to intelligently and adaptively steer traffic management systems, e.g. to actively reduce noise pollution in a certain area and therefore, enable applications with high social, economic and ecological value.

Emergency Vehicle Detection and Tracking: Besides Traffic Density Monitoring, EAR-IT will further investigate the value to identify specific acoustic events in an outdoor environment. In particular, emergency vehicle sirens in cities are in the focus of this use case. By using the acoustic event detection functionality provided by the APUs to be deployed in the wireless sensor network at a suitable spot, sirens will be identified. Research conducted recently already showed that machine learning based siren detection is possible for various applications [6], [7]. Loudness measures provided by the already deployed acoustic sensors in the wireless sensor network will complement this new type of information and enables localization tracking of the emergency vehicle, i.e. its siren across the urban area. This data can then be fed to a traffic management system to actively steer traffic lights with the goal to reduce the overall reaction time of official authorities in case of an incident.

C. Challenges

Enhancing existing wireless sensor networks with acoustic sensing technology is not straight forward and comes with numerous challenges. In particular:

- analysis of the foreseen application scenario, identification of all stakeholders and their individual requirements
- selecting suitable hardware equipment
- robust preparation of the equipment for the various use cases, protecting them against environmental influences
- positioning of the sensors in the environment
- gathering of audio training material to develop models for acoustic event detection
- interfacing the APU to the wireless sensor network and establish a communication route with already existing IoT devices.
- creating situational awareness based on the sensor data on an application level.
- data handling and visualization in close conjunction with the envisioned use cases.

III. Hardware Qualification

A. Embedded Processing Platform

To select a suitable embedded CPU as the processing core for the APU, a benchmarking of different embedded platforms was performed. For choosing a suitable APU system architecture, three different benchmarks were used. As it is not known at the time of the benchmarking if the final audio processing can be distributed to different CPU cores, only single

TABLE I. EVALUATED SYSTEMS FOR CPU PERFORMANCE TESTS

Board	CPU	Core	#Cores	Core Clock (MHz)
APF51	i.MX51	Cortex-A8	1	800
BeagleBoard	AM3359	Cortex-A8	1	720
phyCore	OMAP4430	Cortex-A9	2	1008
PandaBoard ES	OMAP4460	Cortex-A9	2	1200
(Zotac ZD41)	D525	Atom	2	1800
(PC)	Q6600	Core 2	4	2400

core performance was measured in all benchmarks. Table I introduces the platforms considered within the benchmark.

1) Generic performance test: To estimate the overall performance of the selected embedded systems the generic benchmark *nbench* was chosen. *nbench* is designed to expose the capabilities of a system's CPU, FPU, and memory system. It provides a summarized result for the fix point (Int-Idx), floating point (FP-Idx) and memory bandwidth (Mem-Idx) performance of the system under test. The *nbench* benchmark was compiled using different tool chains and various compiler optimization flags for each of the embedded boards. The best result for each system was logged and is shown in Table II. To allow a better comparison, the index numbers are normalized to a CPU clock of 1 GHz.

2) Application Performance Test: In order to provide a more realistic estimation of the required processing platform performance, a test algorithm for acoustic localization [33] was used. Using pre-recorded six channel 48 kHz audio data as input, the position of a sound source was calculated. In order to get a meaningful parameter for performance comparison purposes of the different embedded platforms, the runtime of the algorithm analyzing a 28 second piece of audio data was measured. In Table II, the analyzed sample time divided by the runtime of the algorithm is shown. Hence, a value greater 1.0 means that on this system the algorithm is able to perform the localization at least in real time. It is worth stating that the localization algorithm consisted of un-optimized code. Furthermore, the test conditions were very strict to emulate the worst case computational scenario in EAR-IT.

3) IIR Filter and Level Normalization: As a third test, an algorithm calculating a IIR filter after normalizing the input signal was used. This benchmark also uses a pre-recorded audio signal as input to make the results on the different platforms comparable. In Table II, the number of algorithm runs per second is shown.

4) Selected Hardware: The generic benchmark indicates an approximately equal fixed point performance of the tested embedded systems whereas floating point operations of Cortex-A9 based CPUs show a significant performance gain over the Cortex-A8 architecture. Using the *Application Performance Test* and *IIR Filter and Level Normalization* as a reference, this difference becomes even more prominent, as the audio processing algorithms used herein heavily rely on floating point mathematics, too. The reason for this performance gain is the greatly improved floating-point unit of the Cortex-A9 CPUs (more registers, better cycles per instruction value). However, the performance gain comes with higher prices. Disregarding the significant performance advantage in floating point calculations of the ARM Cortex-A9 architecture, we consider the ARM Cortex A8 architecture as sufficient to address the needs of EAR-IT, i.e. the BeagleBoard.

TABLE II. EMBEDDED PROCESSING PLATFORM BENCHMARKING RESULTS

CPU	Generic Performance (Mem-Idx)	(Int-Idx)	(FP-Idx)	App. Perf.	IIR Filter & Norm.
i.MX51	5980	6125	980	0.3	12
AM3359	7050	6250	1815	0.5	28
MAP4430	6150	6250	7700	2.7	76
OMAP4460	7150	7000	8050	2.6	102
D525	5500	4050	4950	1.9	68
Q6600	10525	7250	13300	5.5	138

B. Microphones

In order to select suitable microphones for the purpose of EAR-IT, one has to take the parameters robustness, price, quality and the actual use cases into account. Within EAR-IT, both in- and outdoor use cases are investigated. Hence, the selected microphones should be robust against environmental influences such as high humidity, wind and a wide temperature range. Desirably, the selected microphones can be used in both use cases in order to keep the effort in developing APUs in large quantities low. At the same time, the microphones should come in a decent quality such that the project outcome is not limited due to poor hardware selection. Even if microphones of superior quality in terms of bandwidth, dynamic, noise floor, etc. are selected that go beyond the needs of the current project, they might still be of high value for upcoming or follow-up projects. Hence, investing into the test-beds from a technological point of view may pay out later on and will encourage the further use of the installed hardware and the overall network. Obviously, the selected microphones should be cheap, exchangeable and easily replaceable if necessary. For the purpose of EAR-IT, three possible microphones have been pre-selected.

1) tBone LC97: The tBone LC97 can be seen as a plug-&-play solution for the needs in EAR-IT. Its frequency response and dynamic is reasonably good for this price segment and is affordable in large quantities. Although robustness may be achieved by using a proper housing and a workaround to use the anti-wind-foam, the tBone LC97 is specifically made for indoor environments. Additionally, quality varies heavily among the microphones of the same type, which makes calibration difficult.

2) Sennheiser MKE23 Gold C: The Sennheiser MKE23 Gold C is a professional condensator microphone for live-stage-usage. Its frequency response is highly linear, with a large dynamic range and very low distortions. This microphone is naturally protected from sweat and offers high reliability and robustness also under harsh environmental conditions. However, it is by far the most expensive miniature microphone within this pre-selection and may not be a suitable choice for large-scale deployments.

3) Shure MX183: Similarly, the Shure MX183 omnidirectional condensator microphone is made for professional live-stage-usage. Hence, frequency response, dynamic range and distortions are excellent. The microphone package comes with a dedicated preamplifier, manufacturer support and all connectors necessary to integrate it into an APU with a reasonable price. Hence, we decided to choose the Shure MX183 due to the best tradeoff between cost and quality.

IV. Acoustic Processing Unit

As it was shown in the past, acoustic sensors in combination with appropriate signal processing strategies [1]- [5] are able to detect, analyze and track various information in in- and outdoor environments unobtrusively, such as sirens & traffic noise [6]- [7], possibly dangerous situations [8], [9] or the position of the user [10]. Hence, this research discipline can bring high social value in form of innovative services to the IoT community.

A. System overview

The EAR-IT approach towards acoustic event detection consists of three major processing stages: 1) a pre-processing stage to obtain low-level information about the input signal and to derive suitable signal representations; 2) an event detection stage to derive mid-level contextual information about the audio data, and, 3) statistical modeling stage to formulate short- and long-time high-level semantics for application and service development. Additionally, suitable metadata is defined to ensure resource efficient communication between the intelligent acoustic sensor and the service layer within the IoT network. A general overview of the proposed approach is given in Figure 3. The system is highly adaptable due to modular structure in each stage, fully automated and non-obtrusive. It respects privacy issues and does not store any contextual information at any point of time, which leads to higher end-user acceptance e.g. than video surveillance.

B. Preprocessing

An audio signal recorded by a microphone in real-world environment is considered to be composed of background sounds and target acoustic signals [11]. Usually, only the latter ones are of interest and should be focused on to keep computation time and consequently, energy consumption, of the IoT devices low - and therefore, have to be separated from the background sound in a pre-processing stage. In [12], multiple voice activity recognition (VAD) algorithms are investigated according to their suitability for various applications while taking parameters like computational complexity and performance into account. Within the EAR-IT use cases, a high temporal resolution to meet the event characteristics while having the possibility to respect for privacy related issues and low computational complexity of the VADs is desired. Furthermore, denoising functionality [15], foreground-background separation [12], filtering, channel selection and localization algorithms [16]- [18] may be part of the pre-processing stage with the goal to provide a high-quality audio signal representation to the consecutive acoustic event detection stage.

C. Acoustic Event Detection

The acoustic event detection is composed of three main steps, namely audio feature extraction, model development and training using a machine learning algorithm and the actual detection of acoustic events. Extracting acoustic event characteristics and features is the first step of most classification systems. Features may be related to the main dimensions of audio characteristics including time, spectral distribution, energy, modulation and other psychoacoustic measures. The audio files themselves cannot be used as direct input for machine-based classification, because they, besides their high dimensionality, are covered with redundancy which first needs to be removed. Therefore, extracted features should clearly reflect the mentioned characteristics of an acoustic event, should ideally have unique appearance for every event class and should be highly correlated with the same features extracted from other event class members. Additionally, the number of features representing one data item should be minimal in order to reduce the computational complexity. For audio feature extraction, we refer to [13] for a list of commonly used audio descriptors. Once significant features have been extracted, any classification scheme may be used to map features to a certain class of acoustic events. Nowadays, audio feature representations are used that are inspired by the human auditory system to account for e.g. non-linearities in frequency and loudness perception instead of relying on plain linear physical measures. It was shown in the past that acoustic event detectors can benefit from this transformation, as shown in [6], [8]- [9] and [14]. Various machine learning algorithms and approaches for - but not limited to - acoustic event detection and classification are available in the literature. In general, two different classes of classification approaches are distinguished:

1) Unsupervised Classification: The data is clustered in a non-supervised way. The classification scheme emerges from the data based on objective similarity measures. An audio file is represented by a set of features, and a similarity measure is used to compare files. Unsupervised clustering algorithms take advantage of the similarity measure to organize a set of events into clusters. The following cluster algorithms are mentioned in the literature: K-Means, Agglomerative Hierarchical Clustering, Self-Organizing Map (SOM), and Growing Hierarchical SOM (GHSOM).

2) Supervised Classification: This approach has been most widely researched. It is based on manually labeled acoustic events, onto which individual events are automatically mapped by applying machine learning algorithms. In an initial training phase, the system is trained with some manually labeled data. Then, the trained system is used to classify unlabeled data. The classifier attempts to automatically form relationships between the features of the training set and the related categories. Under the assumption that the training data set is representative for every unknown data item to be classified, every unknown data item is classified correctly. Typical supervised machine-learning algorithms that have been used in the context of acoustic event detection are: K-Nearest Neighbour, Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), Neural Networks, Linear Discriminate Analysis, Quadratic Discriminate Analysis and Support Vector Machines.

D. Statistical Modeling

In cases where a single acoustic event does not describe a situation sufficiently, post-processing of the acoustic event detection output, i.e. a label for the audio snippet under investigation at a given point in time, and fusion with information from other points in time becomes necessary. In [12], a concept to model an emergency from its temporal characteristics is described. Both, short-term characteristics for instantaneous emergency classification and long-term characteristics for monitoring the progression of events over a longer

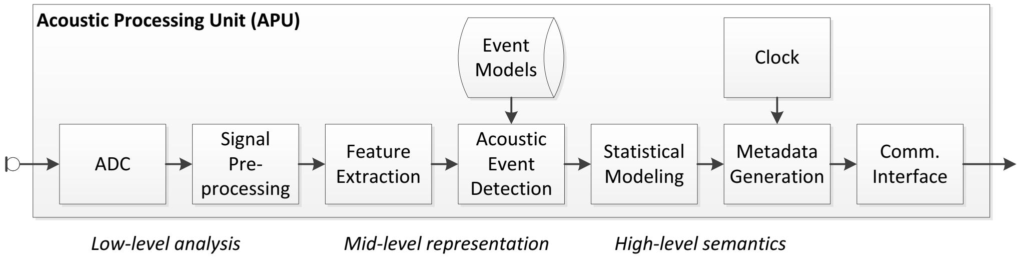

Fig. 3. Functional overview of an APU

period of time are covered by this concept. The model reliably identifies deviances between a set of parameters for an actual time interval based on the knowledge gained from previous time intervals.

E. Metadata

Since the test-beds in EAR-IT have a limited bandwidth in terms of data transmission, we defined a metadata-container instead of transmitting raw audio material. The introduction of the metadata not only heavily reduces the payload of the network, but also allows for the generation of suitable information locally that will later on be used in an application modeling environment, i.e. a service platform. In general, the size of the metadata-container will be considerably smaller compared to raw audio material. The latter will remain in the acoustic sensing unit and will be gathered manually during the training period of the acoustic event detectors. The metadata mainly consists of integer numbers, Booleans and strings. Retransmission/refresh-time of this data towards the application layer of the network depends on the use case but is likely to be irregular. With other words, emergencies should be notified immediately, otherwise notification can be done differentially. Of course, continuous/regular update of the metadata might become necessary for applications such as data visualization. Tests with the network are considered to determine how much data can be send at once and at which frequency such that already running application and services are not harmed. Within EAR-IT all the metadata is generated locally by individual acoustic sensing units. Hence, synchronization is crucial to make use of the data depending on the application. A timestamp should be provided by a global master clock in the network (the individual acoustic sensing units have their own clocks) in conjunction with jitter and delay measurements within the network. If this delay is considerably constant over time, it becomes easy to take it into account for timestamp correction. In the following, a list of potentially suitable metadata is given that will be used over the course of the project, namely: 1) Acoustic Sensing Unit ID as a global unique identifier, 2) Position/Location of the sensor to create self-awareness, 3) A globally valid time-stamp for data synchronization on an application level, 4) Sensor ID if multiple sensors are connected to a unit, 5) Status information about the sensing unit, 6) Loudness levels in dB, 7) Sound direction, 8) Car density as a quantized measure of the amount of traffic around a sensing unit, 9) Annotations of events that have been detected by the acoustic sensing unit. By using a TLV scheme for encoding metadata, less than 65bytes of data have to be transmitted within a refresh-interval, hence reducing the additional payload introduced to an existing IoT network to a minimum.

V. Energy Based Acoustic Source Localization: Experiments and Simulations

In the following, recent research, experiments and simulations are presented towards acoustic source localization using widely distributed sensors. This work is related to the EAR-IT outdoor use case for emergency vehicle detection in urban environments. Commonly used acoustic source localization algorithms based on time differences of arrival or phase information require temporally well synchronized sensors in order to provide robust and reliable estimations. However, these kinds of algorithms cannot be applied when dealing with widely spread sensors especially in wireless sensor networks due to unpredictable jitters, network delays and the lack of a master clock for all components in the network. Hence, energy based acoustic source localization approaches are favored. In general, the signal energy measured by an acoustic sensor E_i^y can be modeled as the acoustic sensor gain weighted sum of all acoustic source energies taking into account their spatial relationship to the sensors (see Eq. 1).

$$E_i^y = g_i \sum_{k=1}^{N} \frac{E_k^x}{\|r_i - x_k\|^2}, \tag{1}$$

Here, g_i defines the known acoustic sensor gain, M is the number of sensors at the known positions r_i, $i = 1...M$, N is the number of sources at unknown positions x_k, $k = 1...M$ and E_k^x determines the unknown source signal energy at location x_k. In the following, we assume that the energy is measured over a period of five seconds. In [35] and [36], an energy based acoustic source localization algorithm for meeting room scenarios is presented. Unfortunately, the use case and algorithm was too constraint to be applicable to outdoor wireless distributed sensor networks. In [34], an energy based maximum likelihood estimation scheme of the acoustic source position is presented for sensor networks. The algorithm is based on the implicit estimation of the acoustic source energies at positions x_k by minimizing the cost function

$$l(\Theta) = \|y - Ha\|^2\,, \tag{2}$$

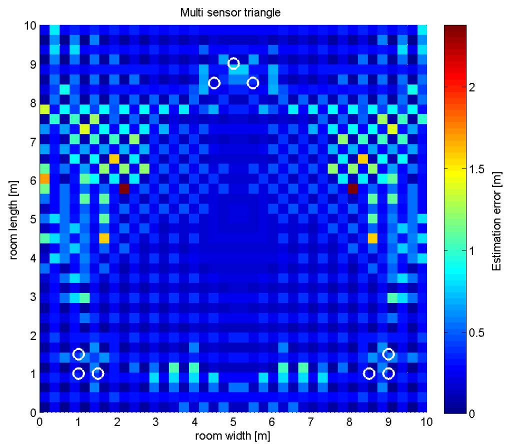

Fig. 4. Localisation accuracy for a multi-node triangular microphone array, full grid search

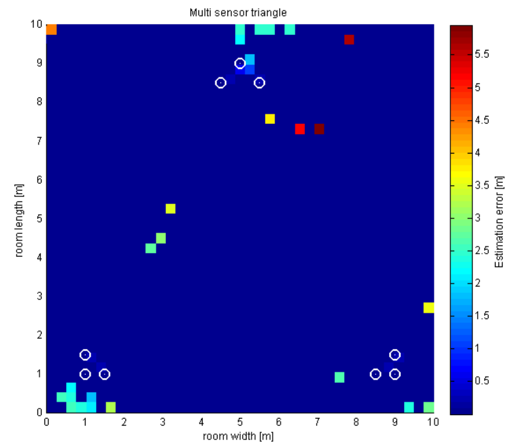

Fig. 5. Localisation accuracy for a multi-node triangular microphone array, bounded simplex-downhill search

where $y = [E_1^y, E_2^y, \ldots, E_M^y]$ is defined as a sensor energy vector, $a = [E_1^x, E_2^x, \ldots, E_N^x]$ defines the source energy vector and H is the sensor-source distance matrix of size $M \times N$. As a requirement, the algorithm needs carefully calibrated sensors and a low-complexity minimum, respectively maximum search algorithm. In the experiment presented here, two search approaches are investigated: Full grid search (brute force) and a bounded simplex algorithm.

We investigate a multi-node triangular microphone in a simulated $10 \times 10m$ room considering both search algorithms to rank them according to localization accuracy under various acoustic conditions for single source scenarios: a) clean conditions and b) a reverberant environment ($1000ms$ reverberation time). The evaluation criteria for the experiment is the Euclidian Distance between the estimated and true source position. Brute Force Full grid search is conducted based on a $55cm$ grid, resulting in approximately 330 search points across the room, whereas the sources are placed on a grid of $27cm$ in order to reduce computation computational complexity. Obviously, full grid search is computational very expensive and exceeds acceptable computation time for more than two source (exponential growth), but usually finds global minima, this, leads to reliable results. In contrast, a bounded simplex downhill algorithm is computationally less expensive and can be limited in its number of iterations. The convergence speed heavily depends on the initial starting values for the algorithm, i.e. heuristic and statistic information about potential source locations, which otherwise would lead to the output of local minima. Within our simulations, sources are placed on a $24cm$ grid across the room. As the initial value for the search a random point in a $4 \times 4m$ location in the middle of the room is selected. Both search algorithms are able to estimate the position of the acoustic source reliably and with reasonable estimation errors as shown in Figure 4 and in Figure 5. Due to the coarser search grid resolution for the full grid search compared to the potential source positions, spatial aliasing occurred. The main difference between both approaches is the significantly smaller computation time for the bounded simplex-downhill search algorithm.

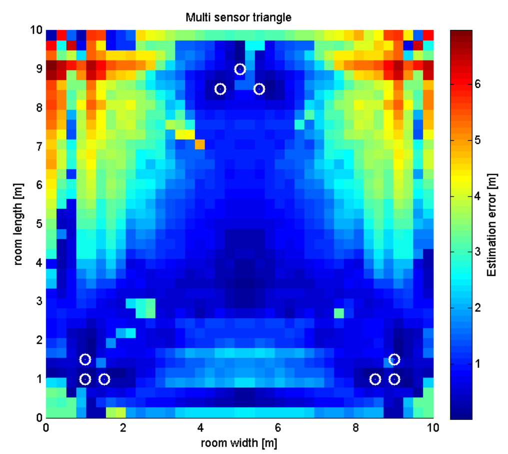

Fig. 6. Localisation accuracy in a room with $1000ms$ for a multi-node triangular microphone array

In order to investigate the influence of reverberation on the cost function and on the estimation results, two experiments with were conducted with 1000ms reverberation time. An increase of the reverberation time leads to a flattening the cost function and in the worst case, to an interference with virtual sound sources and a mismatch between the estimated and virtual number of sound sources as shown in Figure 6.

In the future, further investigations towards the influence of sensor noise will be conducted as well as an evaluation of the selected algorithm combined with the simplex-downhill search algorithm in real environments. Furthermore, the influence of varying block lengths depending on the use case will be investigated. Sensors for siren detection are already deployed, connection to central processing platform already exists and is tested (SmartSantander). Training Material gathering for emergency sirens is done and initial models have been developed. The next step is to record additional training material from the deployment area and refine the models in order to provide a robust and reliable trigger signal for the acoustic source localization algorithm to track the emergency vehicles.

VI. SUMMARY AND FURTURE WORK

The wto years project started only on October 2012 and will continue not only to carry out focus research on acoustic but will gain experience on how such additional intelligence can be smoothly integrated within large IoT network and adding new information that combined with other ones can bring new services. Some challenges will be on information synchronization, device energy efficiency and overall cost reduction to bring Acoustic Sensing Technology to current and unexpected powerful applications.

ACKNOWLEDGMENT

This work was partially funded by the European Commission under grant no. 318381 EAR-IT - Experimenting Acoustics in Real environments using Innovative Test-beds and S4ECoB - Sounds for Energy Control of Buildings under grant no. 284628.

REFERENCES

[1] M. Ziefle & C. Röcker. Human-Centered Design of E-Health Technologies: Concepts, Methods and Applications. Hershey, P.A.: IGI Global. ISBN13: 9781609601775, 2011

[2] S. Goetze, F. Xiong, J. Rennies, T. Rohdenburg & J.-E. Appell, Hands-free telecommunication for elderly persons suffering from hearing deficiencies, in Proc. 12th IEEE Int. Conf. on E-Health Networking, Appl. and Services (Healthcom10), France, 2010.

[3] N. Moritz, S. Goetze & J.-E. Appell, Ambient Voice Control for a Personal Activity and Household Assistant, In Proc. 4th German AALKongress, Berlin, Germany, Jan. 2011.

[4] S. Goetze, N. Moritz, J.-E. Appell, M. Meis, C. Bartsch & J. Bitzer, Acoustic User interfaces for ambient assisted living technologies, Accepted at Informatics for Health and Social Care, 2010.

[5] E. Hänsler & G. Schmidt (Eds.), Speech and Audio Processing in Adverse Environments, Springer, 2008.

[6] J. Schröder, S. Goetze, V.Grützmacher &J. Anemüller, Automatic acoustic siren detection in traffic noise by part-based models,"' in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'13), Vancouver, May 2013

[7] J. Rennies, J. Schröder, D. Hollosi, M. Wittorf & V. Grützmacher, S., Automatic Acoustic Event Detection in Automotive Applications, in Proc. Automotive meets Electronics 2013 (AME'13)Dortmund, 2013

[8] P.W.J. van Hengel & J. Anemüller, Audio event detection for in-home care, in Proc. Int. Conf. on Acoustics (NAG/DAGA'09), 2009.

[9] P.W.J. van Hengel, M. Huisman, & J.E. Appell, Sounds like trouble, in Human Factors - Security and Safety, D. de Waard, J. Godthelp, F.L. Kooi, and K.A. Brookhuis, Eds. Shaker Publishing, Maastricht, The Netherlands, 2009, pp. 369375.

[10] T. Rohdenburg, S. Goetze, V. Hohmann, B. Kollmeier & K.-D. Kammeyer, Combined Source Tracking and Noise Reduction for Application in Hearing Aids, in Proc. 8th ITG-Fachtagung Sprachkommunikation, Aachen, Oct. 2008.

[11] M. Wölfel and J. McDonough. Distant Speech Recognition. John Wiley and Sons Ltd, 1st edition, 2009.

[12] D. Hollosi, J. Schroder, S. Goetze & J. E. Appell, Voice activity detection driven acoustic event classification for monitoring in smart homes, 3rd Int. Symp. on Applied Sciences in Biomedical and Communication Technologies, (ISABEL'10), Rome, 2010, pp. 1-5.

[13] G. Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO Project, CUIDADO I.S.T. Project Report, 2004

[14] J. Schröder, S. Wabnik & P.W.J. van Hengel, Detection und Classification of Acoustic Events for In-Home-Care, In Proc. 4th German AAL-Kongress, Berlin, Germany, Jan. 2011.

[15] S. Goetze, V. Mildner & K.-D. Kammeyer, A Psychoacoustic Noise Reduction Approach for Stereo Hands-Free Systems, in Proc. Audio Engineering Society (AES), 120th Convention, Paris, France, May 2006.

[16] S. Goetze, J. Schröder, S. Gerlach, D. Hollosi, J.-E. Appell & F. Wallhoff, Acoustic Monitoring and Localization for Social Care, in Journal of Computing Science and Engineering (JCSE), SI on uHealthcare, 2012, 6, pp. 40-50

[17] S. Gerlach, S. Goetze & S. Doclo, 2D Audio-Visual Localization in Home Environments using a Particle Filter Speech Communication, in Proc. 10th ITG Symposium, 2012, pp. 1-4

[18] S. Gerlach, S. Goetze, J. Bitzer & S. Doclo. Evaluation of Joint Position-Pitch Estimation Algorithm for Localising Multiple Speakers in Adverse Acoustical Environments, In Proc. Deutsche Gesellschaft fr Akustik e.V. (DAGA'11), 2011, pp. 633-634.

[19] M. Haider, G. Descornet, U. Sandberg, F.G. Pratico. Road Traffic Noise Emission: Recent Developments and Future Prospects, . in Proc. 4th Int. SIIV Congress, 2007.

[20] B. Barbagli, et al. "A Traffic Monitoring and Queue Detection System Based on an Acoustic Sensor Network," Int. Journal On Advances in Networks and Services 4.1 and 2 (2011), pp. 27-37.

[21] N. Shimada, I. Akitoshi & Hiroshi Yasukawa. "A study on using linear microphone array-based acoustic sensing to detect approaching vehicles," In Proc. IEEE Int. Symp. on Communications and Information Technologies (ISCIT'10) , 2010.

[22] S.S. Yang, Y.G. Kim & H. Choi. "Vehicle identification using discrete spectrums in wireless sensor networks," Journal of Networks 3.4, 2008, pp. 51-63.

[23] A. Klausner, S. Erb & B. Rinner. "Dsp based acoustic vehicle classification for multi-sensor real-time traffic surveillance," In Proc. of 15th European Signal Processing Conference. Vol. 7. 2007, pp. 1916-1920.

[24] EU FP7 Project EAR-IT Experimenting Acoustics in Real environment using Innovative Test-beds. (ID 318381). Online: http://www.ear-it.eu.

[25] EU FP7 Project S4ECoB Sounds for Energy Control of Buildings (ID 284628). Online: http://www.s4ecob.eu.

[26] EU FP7 Project Smart Santander (ID 257992). Online: http://www.smartsantander.eu.

[27] EU FP7 Hobnet Holistic Platform design for Smart Buildings of the future InternNET (ID 257466). Online: http://www.hobnet-project.eu.

[28] BMBF KMU Innovative SonicSentinel - Intelligent Noise Guard for Detecting Critical Situations of the Elderly in Care and at Home. Online: http://www.idmt.fraunhofer.de/en/projects/expired_publicly_financed_research_projects/sonicsentinel.html.

[29] Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise.

[30] L. C. Den Boer & A. Schroten. "Traffic noise reduction in Europe," Health effects, social costs and technical and policy options to reduce road and rail traffic noise. CE, Delft, 2007.

[31] Transport and Environment - What Science Says. Online: urlhttp://www.transportenvironment.org/what-we-do/vehicle-noise/what-science-says.

[32] World Health Organization - Noise, Facts and Figures, Online: http://www.euro.who.int/en/what-we-do/health-topics/environment-and-health/noise/facts-and-figures.

[33] A. Brutti, M. Omologo, & P. Svaizer. Comparison Between Different Sound Source Localization Techniques Based on a Real Data Collection Hands-Free Speech Communication and Microphone Arrays, in Proc. Joint Workshop Hands-free Speech Communication and Microphone Arrays Trento (HSCMA'08), Trento, May 6-8, 2008, pp. 69 -72

[34] D. Ampeliotis & K. Berberidis. Low complexity multiple acoustic source localization in sensor networks based on energy measurements. in Signal Processing, 90(4), 2010, pp. 13001312.

[35] Z. Liu, Z. Zhang, L.-W. He & P. Chou. Energy-based sound source localization and gain normalization for ad hoc microphone arrays. In Proc. IEEE Int. Conf. on Acoustics, Speech and Sig. Proc., 2007 (ICASSP'07), vol.2, 2007, pp. 761.

[36] M. Chen, Z. Liu, L.-W. He, P. Chou & Z. Zhang. Energy-based position estimation of microphones and speakers for ad hoc microphone arrays. In Proc. IEEE Workshop on App. of Sig. Proc. to Audio and Acoustics, 2007, pages 2225.