Cumulative-Sum-Based Localization of Sound Events in Low-Cost Wireless Acoustic Sensor Networks

Article *in* IEEE Transactions on Audio Speech and Language Processing · August 2014

DOI: 10.1109/TASLP.2014.2351132

Cumulative sum approaches for event detection and localization in low-cost wireless acoustic sensor networks

Maximo Cobos[a,*], Juan J. Perez[a], Santiago Felici[a], Jaume Segura[a], Juan M. Navarro[b]

[a] *Computer Science Department, Universitat de València, Valencia, Spain 46100*
[b] *Advanced Telecommunications Group, San Antonio's Catholic University of Murcia, Campus de los Jeronimos, 135, Guadalupe, Spain 30107*

Abstract

Wireless acoustic sensor networks (WASNs) are known for their potential applications in multiple areas, such as audio-based surveillance, binaural hearing aids or advanced acoustic monitoring. The knowledge of the spatial position of a source of interest is usually a requirement for many of these applications. Therefore, source localization is an important problem to be addressed in WASNs. Unfortunately, most localization algorithms need costly signal processing stages that prevent them to be implemented in low-cost sensor networks, requiring additional modules for signal acquisition and processing. This paper presents a low-complexity method for acoustic event detection and localization considering a change detection statistical framework. Two possible implementation approaches based on the efficient cumulative sum (CUSUM) algorithm are presented and discussed. Results from simulations and a real deployment show that the proposed techniques can be easily implemented in low-cost sensor networks, providing good localization accuracy and making good use of the available node resources.

Keywords: Wireless sensor networks, source localization, cumulative sum (CUSUM), time difference of arrival (TDOA), statistical signal processing

1. Introduction

Digital signal processing algorithms are continuously being adapted and redirected towards new data acquisition and processing architectures. The recent advances in communications, mobile computing, nano-technology and smart sensors are bringing to the industry new devices that will probably change the traditional paradigm settled by localized sensor geometries [1, 2]. In the context of acoustic signal processing, the use of low-power, low-cost acoustic sensors with computing capabilities offers new opportunities for the development of new applications and processing strategies, different from the ones established by usual microphone array arrangements [3]. As a result, many research efforts in the acoustic signal processing field are being oriented towards wireless acoustic sensor networks (WASNs) [4, 5]. A WASN consists of a set of sensor nodes interconnected in some manner via a wireless medium. Each node consists of one or more sensors (microphones), a processing unit and a wireless communication module. The most significant advantage of a WASN over a traditional (wired) array is that it allows for an increased spatial coverage by distributing sensors over a larger volume, providing a scalable and easy to deploy structure with better signal to noise ratio (SNR) properties. However, despite these advantages, there are also many challenges to be addressed. On the one hand, the nodes are strongly limited in terms of power and communication bandwidth. On the other hand, the data from all the nodes must be properly processed to perform a given task, leading to centralized and distributed data fusion strategies [6]. In a centralized scheme, the nodes send their raw data to a master node that performs all the processing. In contrast, distributed schemes share the processing among all the nodes, which are allowed to exchange information with each other. While centralized approaches require large communication bandwidth and transmission power, distributed approaches provide more scalable solutions and, therefore, distributed algorithms are usually preferred. Hybrid schemes are also possible, where the nodes perform some local processing and communicate their results to a fusion center that deals with the most computationally expensive parts of the processing.

The high data rates and non-stationarity properties of typical audio signals make WASNs a very challenging research field. Moreover, many speech processing tasks such as blind source separation [7, 8] or acoustic echo cancellation [9] require specific time synchronization solutions. A topic with many significant contributions in WASN is acoustic noise reduction [10]. Beamforming methods based on well-known criteria such as the minimum variance distortionless response (MVDR) have been developed both for hearing-aid systems [11, 12] and more general sensor architectures [13, 14, 15]. All these approaches are aimed at producing signal estimates from the local signal observations, a task that typically requires high communication

*Corresponding author

Email addresses: maximo.cobos@uv.es (Maximo Cobos), juan.j.perez@uv.es (Juan J. Perez), santiago.felici@uv.es (Santiago Felici), jaume.segura@uv.es (Jaume Segura), jmnavarro@ucam.edu (Juan M. Navarro)

URL: http://www.uv.es/macose2 (Maximo Cobos)

bandwidth and energy consumption.

Together with acoustic beamforming, one of the most important tasks in WASNs is acoustic source localization [16, 17, 18]. Source localization has a wide range of potential applications, such as speaker detection and tracking in videoconferenceing environments or audio-based surveillance. Moreover, source localization may be used to provide supporting location information for noise reduction and speech enhancement algorithms. Source localization can be performed either in a centralized [19, 20, 21, 22] or decentralized fashion [23, 24, 25, 26]. Traditionally, most localization methods for WASNs have been focused on sound energy measurements [27]. This is motivated by the fact that the acoustic power emitted by targets usually varies slowly with respect to time. Thus, the acoustic energy time series does not require as much sampling rate as the raw signal time series, avoiding the need for high transmission rates and accurate synchronization. Sheng and Hu [22] proposed a maximum likelihood (ML) energy-based localization method for wireless networks. Energy-based methods for estimating both the position of microphones and speakers in a meeting situation were also proposed in [28]. Besides energy-based approaches, methods based on time-delay estimation (TDE) are also very popular [29, 30]. The TDE category covers all the methods based on time of arrival (TOA) and/or time-difference of arrival (TDOA) estimation, which usually require sensor synchronization [31]. TDOA-based localization methods have been widely considered both in traditional (wired) and wireless ad-hoc networks. Well-known approaches such as the steered response power phase transform (SRP-PHAT) have been shown to be very robust in the presence of reverberation [32, 33]. However, they are computationally too complex to be used in large scale wireless networks. In the context of WASNs, localization methods based on TDOA usually follow a two step approach. First, TDOA measurements are obtained from the analysis of the sensor signals [34, 35, 36]. Then, a source location estimate is obtained by merging together the TDOA measurements under a proper optimization function based on some criteria such as ML [37] or least squares [38, 39, 40]. However, the localization accuracy is very sensible to TDOA measurement errors. In fact, the estimation of time delays in WASNs is not straightforward, since the main operation underlying most TDOA-based approaches is cross-correlation. This requires a high bandwidth for data transfers from the sensors to a processing unit and/or computationally intensive signal manipulations in the frequency domain. Moreover, although some approaches consider the use of multiple microphones in each node to avoid potential synchronization problems [41, 42, 43], the hardware cost of these systems is significantly higher. Therefore, low-complexity and easy to implement solutions are needed for developing feasible localization methods adapted to large scale WASNs.

Most of the above referenced works have been discussed considering only simulation results, thus, specific hardware limitations are not usually taken into account. In fact, few works are evaluated over real network deployments, making difficult to assess the feasibility of implementing a given method and adapting it to a low-cost hardware system. In this context, commercial-of-the-shelf (COTS) hardware platforms such as smartphones [44], laptops [42] or PDAs [45, 29] are usually considered in typical WASN applications. Although these platforms are widespread and provide many powerful features in terms of computing power, communication bandwidth and system memory, the resources of most of the available devices aimed at long-term acoustic monitoring are far more limited. For example, consider the specific literature on gunshot detection and shooter localization systems for surveillance and military applications [46, 47]. Many of these systems consider Berkeley's Mica2 motes running TinyOS [48]. While this platform provides excellent hardware and software features for low-cost sensor network applications, severe resource constraints prohibit the implementation of most algorithms on the mote itself. As a result, many of these systems need additional modules for signal processing and acoustic data acquisition. In any case, it becomes apparent that the design of suitable source localization algorithms for low-cost hardware platforms is not straightforward.

This paper presents a low-complexity hybrid method for acoustic event detection and localization in WASNs, which is intended to be implemented in low-cost sensor networks without the need for additional node modules. Each node in the network analyzes its captured signal and detects the time instant corresponding to the onset of an acoustic event. This processing is performed locally at each node, avoiding the need for intense data transmissions. In a second step, the nodes send their time instants to a central node that produces a source location estimate by merging the received information with some appropriate estimator. A change detection statistical framework is considered for the estimation of time delays at each node, suggesting two different approaches that make use of the well-known cumulative sum (CUSUM) algorithm [49]. In the first one, acoustic event detection and source localization are continuously performed in an on-line (sample-by-sample) basis. In the second one, the algorithm follows a batch approach by addressing the source localization task over a set of stored data samples. While this last approach reduces power consumption and allows for better parameter estimates, some practical considerations have to be taken to assume that a relevant acoustic event has been produced. Simulations and experiments considering a real WASN deployment are discussed, showing that, in both cases, the algorithms can be efficiently implemented in nodes with very limited resources, providing an efficient solution for source localization in low-cost acoustic monitoring systems.

The paper is structured as follows. Section 2 presents the signal model and statistical framework for the proposed CUSUM-based event detection and onset time estimation methods. Section 3 proposes two different approaches for implementing this framework efficiently in WASNs. The general source localization problem to be addressed by the central node is described in Section 4. Section 5 provides a detailed description of the experiments carried out to evaluate the performance of the method with simulations and a real network deployment. Finally, the conclusions of this work are summarized in Section 6.

2. Acoustic event detection and onset time estimation

2.1. Signal Model

Consider a set of microphone nodes and one sound source. The signal captured by the mth microphone node, $x_m[n]$, can be expressed as

$$x_m[n] = h_m[n] * s[n] + v_m[n], \tag{1}$$

where $s[n]$ is the original source signal, $h_m[n]$ is the impulse response between the source signal and the mth microphone, for $m \in \{0, 1, \ldots, M-1\}$, where $M \in \mathbb{Z}$ denotes the number of microphone nodes, and $v_m[n]$ is an additive sensor noise component.

The impulse response $h_m[n]$ can be split into two parts, one corresponding to the direct-path and another part corresponding to multi-path propagation, i.e.

$$h_m[n] = h_m^d[n] + h_m^r[n] = a_m \delta[n - \tau_{ms}] + h_m^r[n], \tag{2}$$

where $h_m^d[n]$ and $h_m^r[n]$ denote, respectively, the direct-path and the reverberant part of the impulse response, $a_m \in \mathbb{R}$ is an attenuation factor and $\tau_{ms} \in \mathbb{Z}$ is the time-lag due to the propagation from the location of the source to the mth microphone. This time is usually referred to as the *time of arrival* (TOA) at microphone m. Let us assume that the source signal is zero for $0 < n < t_s$, where t_s is the absolute time at which the acoustic event emitted by the source takes place. Therefore, the signal captured by each microphone can be rewritten as:

$$x_m[n] = \begin{cases} v_m[n], & \text{if } n < \tau_m. \\ a_m s[n - \tau_{ms}] + h_m^r[n] * s[n] + v_m[n], & \text{if } n \geq \tau_m, \end{cases} \tag{3}$$

where $\tau_m = \tau_{ms} + t_s$, denotes the time at which the mth sensor notices the event, i.e. the onset time at sensor m.

2.2. Statistical framework

The signal from each microphone node can be modeled as a discrete random signal $\mathcal{X}_m[n]$, with independent and identically distributed samples. The probability density function (pdf) of each sample is given by $p(x_m[n], \theta)$, where θ is a deterministic parameter. The occurrence of an event is modeled by an instantaneous change in θ, so that $\theta = \theta_0$ before the event at $n = \tau_m$ and $\theta = \theta_1$ when $n \geq \tau_m$. Thus, the two possible hypotheses are

$$\begin{aligned} \mathcal{H}_1 &: \theta = \theta_1, \quad \textit{An event has been produced}. \\ \mathcal{H}_0 &: \theta = \theta_0, \quad \textit{No event has been produced}. \end{aligned}$$

The pdf of the signal $\mathcal{X}_m[n]$ observed between the initial sample $x_m[0]$ and the current sample $x_m[k]$ can take two forms depending on the above hypotheses. Under the *'no event'* hypothesis $\mathcal{H}_0$, the pdf is

$$p_{\mathcal{X}_m|\mathcal{H}_0} = \prod_{n=0}^{k} p(x_m[n], \theta_0). \tag{4}$$

On the other hand, under the *'event'* hypothesis $\mathcal{H}_1$, the pdf would be

$$p_{\mathcal{X}_m|\mathcal{H}_1} = \prod_{n=0}^{\tau_m - 1} p(x_m[n], \theta_0) \prod_{n=\tau_m}^{k} p(x_m[n], \theta_1). \tag{5}$$

The detection theory establishes that the *log-likelihood ratio* (LLR) test is the most powerful test (in the statistical sense) to decide between the two hypotheses [50]. The LLR is defined by

$$\Lambda_{\mathcal{X}_m} \triangleq \ln \left(\frac{p_{\mathcal{X}_m|\mathcal{H}_1}}{p_{\mathcal{X}_m|\mathcal{H}_0}} \right). \tag{6}$$

The hypothesis $\mathcal{H}_1$ is decided if $\Lambda_{\mathcal{X}_m} > \eta$, where $\eta \in \mathbb{R}$ is a threshold defined by the user. If $\Lambda_{\mathcal{X}_m} \leq \eta$, then the decision is $\mathcal{H}_0$. When looking at the current sample $x_m[k]$, the LLR taking into account Eq.(4) and Eq.(5) becomes

$$\Lambda_{\mathcal{X}_m}[k, \tau_m] = \sum_{n=\tau_m}^{k} \ln \left(\frac{p(x_m[n], \theta_1)}{p(x_m[n], \theta_0)} \right). \tag{7}$$

It is evident that it is not possible to calculate the above quantity because it depends on the unknowns θ_0, θ_1 and τ_m. Fortunately, a *generalized log-likelihood ratio* (GLLR) can be defined by replacing all the unknowns in $\Lambda_{\mathcal{X}_m}$ by their ML estimates:

$$\begin{aligned} \Gamma_{\mathcal{X}_m}[k] &\triangleq \max_{1 \leq \tau_m \leq k} \Lambda_{\mathcal{X}_m}[k, \tau_m] && (8) \\ &= \max_{1 \leq \tau_m \leq k} \sum_{n=\tau_m}^{k} \ln \left(\frac{p(x_m[n], \hat{\theta}_1)}{p(x_m[n], \hat{\theta}_0)} \right), && (9) \end{aligned}$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are the ML estimates of the parameters. When these estimates are available, the test can be used to select $\mathcal{H}_1$ if $\Gamma_{\mathcal{X}_m}[k] > \eta$. Thus, this ratio is also known as the *decision function*.

When $\mathcal{H}_1$ has been decided, the next problem is to estimate the onset time τ_m from the measured samples $x_m[0], \ldots, x_m[k]$. The ML estimate for τ_m is the value maximizing the likelihood $p_{\mathcal{X}|\mathcal{H}_1}[k, \tau_m]$ [51]:

$$\begin{aligned} \hat{\tau}_m &\triangleq \arg \max_{1 \leq \tau_m \leq k} p_{\mathcal{X}|\mathcal{H}_1}[k, \tau_m] && (10) \\ &= \arg \max_{1 \leq \tau_m \leq k} \Lambda_{\mathcal{X}_m}[k, \tau_m] && (11) \\ &= \arg \max_{1 \leq \tau_m \leq k} \sum_{n=\tau_m}^{k} \ln \left(\frac{p(x_m[n], \hat{\theta}_1)}{p(x_m[n], \hat{\theta}_0)} \right). && (12) \end{aligned}$$

2.3. Densities and ML estimates

According to Eq.(3), $x_m[n] = v_m[n]$ when $n < \tau_m$, i.e. the samples before the acoustic event belong exclusively to the noise component. We assume that it follows a normal distribution with zero mean and variance $E\left[v_m^2[n]\right] = \sigma_0^2$:

$$p(x_m[n], \theta_0) \sim \mathcal{N}(0, \sigma_0^2) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{x_m^2[n]}{2\sigma_0^2}}, \tag{13}$$

where $\theta_0 \triangleq \sigma_0^2$, i.e. its variance is the parameter modeling the pdf before τ_m.

The modeling of the samples after the acoustic event is more problematic. On the one hand, it is difficult to assume a particular distribution for every acoustic event, since many different types of audio signals might occur (speech, impulsive noises, etc.) On the other hand, different acoustical conditions may also lead to different statistical distributions. In [52], it was shown that very small time frames of speech can be approximately modeled by a Gaussian distribution. Moreover, noise and reverberation also contribute to accentuate the normality of the resulting observed sound signal due to the effect of $h^r_m[n]$. As a result, a Gaussian pdf is selected as a model for the observed signal samples after τ_m, having the advantage of providing a more general solution under different kinds of acoustic signals and environments. Moreover, the use of Gaussian pdfs results in simpler calculations in the nodes. Then, the pdf after τ_m is expressed as

$$p(x_m[n], \theta_1) \sim \mathcal{N}(0, \sigma_1^2) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x_m^2[n]}{2\sigma_1^2}}, \tag{14}$$

where $\theta_1 \triangleq \sigma_1^2 = E\left[(h^r_m[n] * s[n]) + v_m[n])^2\right]$.

As already discussed, the variances σ_0^2 and σ_1^2 are usually unknown. Moreover, in order to accurately estimate them, it is necessary to know in advance the value of τ_m, which is also unknown. This is certainly a problem in the application of Eq.(9) when working on a sample-by-sample basis. The estimation of the variance σ_0^2 might not be so critical since, at a given time k, the previous samples from $x_m[0]$ to $x_m[k]$ can be used to obtain a ML estimate. Alternatively, a calibration stage before running the algorithm can also be performed to estimate the noise variance in the absence of signal. However, the estimation of σ_1^2 is not possible without any *a priori* information. This information can be in the form of a minimum parameter change from θ_0 to θ_1. Nevertheless, the problem can still be properly addressed if the processing follows a *batch* approach instead of an *on-line* approach, i.e. if the samples are not required to be processed one-by-one as they arrive. By using a batch approach, the microphone nodes have K samples available and it is assumed that there has been some acoustic event within these samples. As will be discussed in Section 3, some signal acquisition issues must be considered in order to let the nodes know that some event has occurred. In this case, the ML estimates of the parameters are directly obtained from the observed data. The parameter σ_0^2 is estimated from the samples $x_m[0], \ldots, x_m[n-1]$, while the parameter σ_1^2 is estimated from the samples $x_m[n], \ldots, x_m[K-1]$:

$$\hat{\theta}_0^{(\mathrm{ML})}[n] \triangleq \hat{\sigma}_0^2[n] = \frac{1}{n}\sum_{k=0}^{n-1} x_m[k]^2, \tag{15}$$

$$\hat{\theta}_1^{(\mathrm{ML})}[n] \triangleq \hat{\sigma}_1^2[n] = \frac{1}{K-n}\sum_{k=n}^{K-1} x_m[k]^2. \tag{16}$$

Then, the above quantities can be used as ML estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ for each possible $n = \tau_m$ in the evaluation of Eq.(12). However, the use of these estimates may require too much processing in common WASN applications, since the pdfs involved in Eq.(12) depend on the actual estimated parameters. A practical solution to this problem is to use only one value for $\hat{\theta}_0$ and one for $\hat{\theta}_1$ so that the pdfs do not change with every possible τ_m. These values can be easily computed from different subsets of the K available samples. In fact, although the onset time of the event at each node is unknown, a set of practical considerations can be taken to guarantee that a range of T_0 initial samples, $n = 0, \ldots, T_0 - 1$, does not certainly contain the onset of the acoustic event, while the range of ending samples $n = K - T_0, \ldots, K - 1$, corresponds to the observation signal after the event has actually occurred. Under these assumptions, the estimates are given by

$$\hat{\theta}_0^{(T_0)} \triangleq \hat{\sigma}_0^2 = \frac{1}{T_0}\sum_{n=0}^{T_0-1} x_m[n]^2, \tag{17}$$

$$\hat{\theta}_1^{(T_0)} \triangleq \hat{\sigma}_1^2 = \frac{1}{T_0}\sum_{n=K-T_0}^{K-1} x_m[n]^2, \tag{18}$$

where now they can be assumed to be constant when estimating τ_m. Section 3 discusses the selection of T_0 and other signal acquisition issues. The left panel of Figure 1 shows an example where different types of parameter estimates have been considered for some observed acoustic event. The instantaneous ML parameters $\hat{\theta}^{(\mathrm{ML})}$ are shown in Fig. 1(a), while the fixed parameters $\hat{\theta}^{(T_0)}$ are shown in (b) for $T_0 = 100$ samples.

2.4. Cumulative sum algorithms

Signal detection and onset time estimation can be efficiently implemented by using the *cumulative sum* (CUSUM) algorithm, which is based on the original work from Page [49]. This algorithm has very low computational complexity and can be easily implemented in nodes with severe resource constraints. Next, two different implementations are described to fit different application needs. The first one is the on-line version of the algorithm, which must be continuously running at each node on a sample-by-sample basis to perform the detection of the event and to estimate an onset time whenever $\mathcal{H}_1$ is decided. The second one is a batch approach that only needs to be executed under the assumption that the acoustic event has just actually been produced, saving computational resources and battery life.

2.4.1. On-line CUSUM

Define the *instantaneous LLR* at time n as

$$l_m[n] \triangleq \Lambda_{\mathcal{X}_m}[n, n] = \ln\left(\frac{p(x_m[n], \theta_1)}{p(x_m[n], \theta_0)}\right). \tag{19}$$

Including Eq.(13) and Eq.(14) into Eq.(19), the instantaneous LLR becomes:

$$l_m[n] = \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)x_m^2[n] + \frac{1}{2}\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right). \tag{20}$$

Define the cumulative sum from 0 to k as:

$$S_m[k] \triangleq \sum_{n=0}^{k} l_m[n]. \tag{21}$$

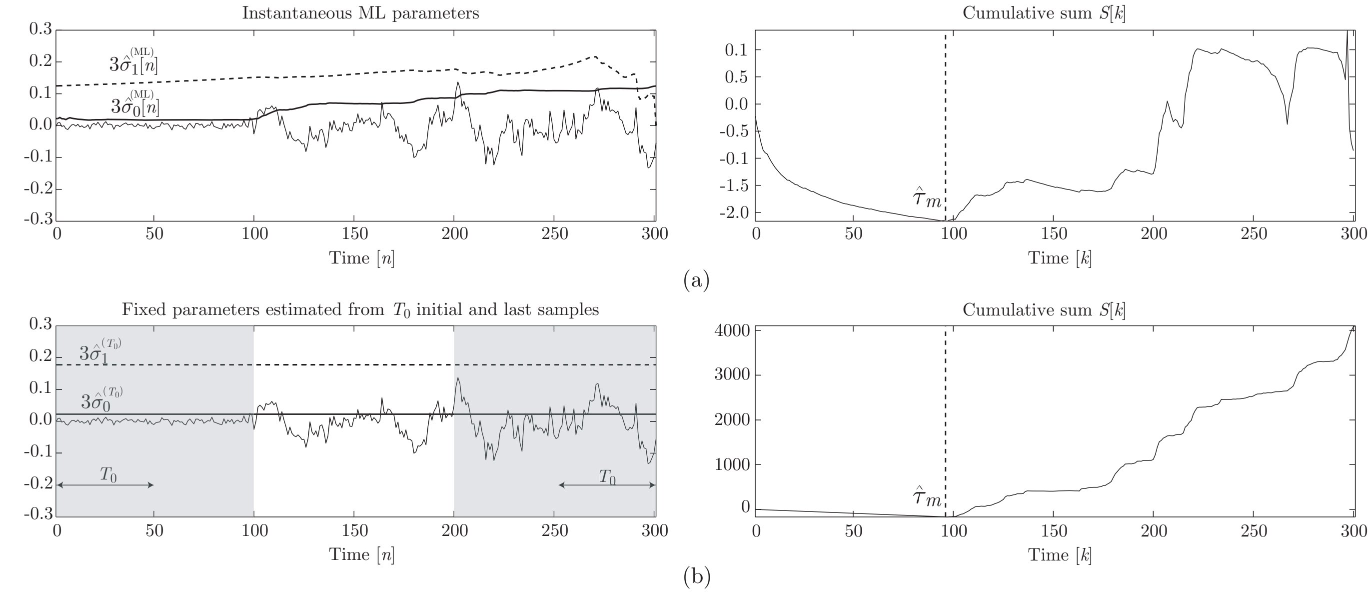

Figure 1: Signal example showing the effect of different parameter estimates and the resulting cumulative sum. For illustrative purposes, the parameters are represented as three times the estimated standard deviations for the two hypotheses. (a) ML estimates. (b) Fixed estimates considering starting and ending parts of the signal of length $T_0 = 100$ samples.

From Eq.(19) and (21), the LLR defined in Eq.(6) can be expressed as

$$\Lambda_{\mathcal{X}_m}[k, \tau_m] = S[k] - S[\tau_m - 1]. \tag{22}$$

By using the above formula in Eq.(9) and Eq.(12), we obtain

$$\Gamma_{\mathcal{X}_m}[k] = S[k] - \min_{1 \le \tau_m \le k} S[\tau_m - 1], \tag{23}$$

$$\hat{\tau_m} = \arg \min_{1 \le \tau_m \le k} S[\tau_m - 1]. \tag{24}$$

The above algorithm can be rewritten in a recursive manner. The cumulative sum can be simply calculated as

$$S_m[k] = S_m[k-1] + l_m[k]. \tag{25}$$

As shown in [53], since the decision function is compared to a positive threshold η, it can be rewritten as:

$$\Gamma_{\mathcal{X}_m}[k] = \left\{\Gamma_{\mathcal{X}_m}[k-1] + l_m[k]\right\}^+, \tag{26}$$

where $\{z\}^+ \triangleq \sup(z, 0)$.

The necessary steps to implement the on-line approach are given in Algorithm 1. It assumes the values of σ_0^2 and σ_1^2 are known. As discussed in Section 2.3, in usual applications, the variances after the event might be difficult to assess under an on-line framework, requiring some a priori information such as a minimum variance change. In any case, the following batch algorithm can be used to simplify the processing.

2.4.2. Batch CUSUM

The batch approach is based on an available number of samples K at each microphone node, all of them containing an onset from a given acoustic event. By assuming the availability of these samples, the decision function $\Gamma_{\mathcal{X}_m}$ is not necessary anymore, since the only possible hypothesis under this scenario is $\mathcal{H}_1$. Therefore, it is sufficient to estimate τ_m by using Eq.(12), evaluated from $\tau_m = 1$ to $\tau_m = K - 1$, i.e.

$$\hat{\tau}_m = \arg \max_{1 \le \tau_m \le K-1} \sum_{n=\tau_m}^{K-1} \ln\left(\frac{p(x_m[n], \hat{\theta}_1)}{p(x_m[n], \hat{\theta}_0)}\right). \tag{27}$$

By using the quantities defined in Section 2.4.1, the above formula is equivalent to

$$\hat{\tau}_m = \arg \min_{1 \le \tau_m \le K-1} S[k]. \tag{28}$$

Algorithm 2 summarizes this batch approach. While ML estimates for $\hat{\theta}_0$ and $\hat{\theta}_1$ can be obtained through the use of Eq.(15) and Eq.(16), it results easier to have one value for each parameter obtained from the initial and last T_0 observed samples, as suggested by Eq.(17) and Eq.(18). The right panel of Figure 1 shows the resulting cumulative sum for different parameter estimates. While the curves are significantly different, the minimum is the same in both cases ($\hat{\tau}_m = 97$ samples).

3. Signal acquisition, communication and synchronization

In the last section, acoustic event detection and onset time estimation were discussed by considering a LLR-based detection framework. Algorithms 1 and 2, which are based on the well-established CUSUM method, were proposed as potential solutions to perform onset time estimation in WASNs. These algorithms define two different strategies for handling acoustic events, both of them having some strengths and weaknesses. On the one hand, Algorithm 1 is able to detect acoustic events

Algorithm 1 On-line CUSUM

initialization
  set detection threshold $\eta > 0$
  set values for σ_0^2 and σ_1^2
  $C_1 = \frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)$
  $C_2 = \frac{1}{2}\ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right)$
  $S_m[-1] = \Gamma_{\chi_m}[-1] = 0$
  $k = 0$
end
while *not stopped* **do**
  $l_m[k] = C_1 x_m^2[n] + C_2$
  $S_m[k] = S_m[k-1] + l_m[k]$
  $\Gamma_{\chi_m}[k] = \{\Gamma_{\chi_m}[k-1] + l_m[k]\}^+$
  if $\Gamma_{\chi_m}[k] > \eta$ **then**
    $\hat{\tau}_m = \arg\min_{1\leq\tau_m\leq k} S_m[\tau_m - 1]$
    return $\hat{\tau}_m$ to sink and reset
  end if
  $k = k + 1$
end while

Algorithm 2 Batch CUSUM

initialization
  set $\hat{\sigma}_0^2 = \hat{\sigma}_1^2 = 0$
  for $k = 0$ to $k = T_0 - 1$ **do**
    $\hat{\sigma}_0^2 = \hat{\sigma}_0^2 + x_m^2[k]$
  end for
  $\hat{\sigma}_0^2 = (1/T_0)\hat{\sigma}_0^2$
  for $k = K - T_0$ to $k = K - 1$ **do**
    $\hat{\sigma}_1^2 = \hat{\sigma}_1^2 + x_m^2[k]$
  end for
  $\hat{\sigma}_1^2 = (1/T_0)\hat{\sigma}_1^2$
  $C_1 = \frac{1}{2}\left(\frac{1}{\hat{\sigma}_0^2} - \frac{1}{\hat{\sigma}_1^2}\right)$
  $C_2 = \frac{1}{2}\ln\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)$
  $S_m[-1] = 0$
end
for $k = 0$ to $k = K - 1$ **do**
  $l_m[k] = C_1 x_m^2[n] + C_2$
  $S_m[k] = S_m[k-1] + l_m[k]$
end for
$\hat{\tau}_m = \arg\min_{1\leq\tau_m\leq K-1} S_m[\tau_m]$
return $\hat{\tau}_m$ to sink and reset

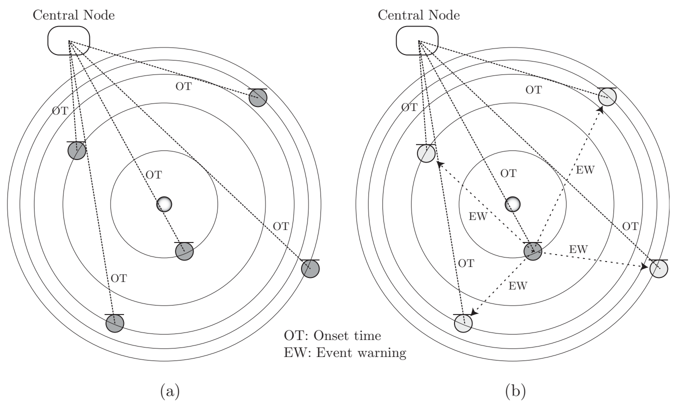

Figure 2: Detection and onset time estimation strategies. (a) On-line approach. (b) Batch approach.

automatically by looking at the statistical properties of the signal and its previous history, providing the central node or sink with an onset estimate whenever an acoustic event produces a statistical change on the acquired signal. However, the parameters that define the pdfs under the *'event'* and *'no event'* hypotheses might be difficult to set up without any a priori information. Moreover, the algorithm is continuously running on the microphone nodes, affecting significantly their battery life. On the other hand, Algorithm 2 only needs to be executed when it is assumed that some event has occurred. The decision function is not necessary in this case and better parameter estimates can be used to calculate $\hat{\tau}_m$. As it will be next explained, although the performance of this method depends on the node capability to detect an acoustic event from a basic amplitude threshold, the power consumption in the nodes is substantially reduced. The next subsection describes in detail the node communication issues involved in these two approaches.

3.1. On-line approach

The steps involved in the on-line approach are as follows:

1. Each node is configured with some user-defined parameters: η, σ_0^2 and σ_1^2. A calibration stage before operation may be needed to set adequately these parameters.
2. Algorithm 1 is run at each node. The sink is continuously listening to the onset times reported by the nodes, which are sent together with some additional clock information for synchronization.
3. Whenever the sink obtains, at least, five onset times (4 TDOAs) closely spaced in time (within a maximum allowed delay) from different nodes, they are fed to the localization algorithm to obtain an estimate of the source location (see Section 4). The maximum allowed delay $D_{\max} \in \mathbb{R}$ corresponds to the maximum TDOA for the sensor setup plus the maximum radio transmission delay, i.e.

$$D_{\max} = \kappa\left(\max_{(i,j)\in Q}\left\{\frac{r_{ij}}{c}\right\} + D_t\right), \tag{29}$$

where $Q = \{(0,1),(0,2),\ldots,(0,M-1),(1,2),\ldots,(M-2,M-1)\}$ is the set containing all the available pairs of nodes, $D_t \in \mathbb{R}$ is the maximum radio transmission delay in seconds, c is the speed of sound, r_{ij} is the distance separating the node pair $(i, j) \in Q$ and $\kappa \in \mathbb{R}$ is a user-defined constant slightly higher than 1.

Figure 2(a) depicts this approach, where all the nodes are continuously running the algorithm (represented by dark gray color) and only establish communication with the central node whenever they have detected an onset time.

3.2. Batch approach

In the batch approach, it is assumed that the node closest to the source location has the greatest SNR, defined as $SNR \triangleq \frac{\sigma_1^2}{\sigma_0^2}$. Under this assumption, the closest node may be able to decide between $\mathcal{H}_0$ and $\mathcal{H}_1$ by using a simple amplitude threshold. Then, it will notify the rest of nodes by sending an event warning alert to start the onset time estimation algorithm. The detailed steps are as follows:

1. The nodes are configured with an amplitude threshold $\gamma \in \mathbb{R}$ and a pre/post-event time slot length $T_0 \in \mathbb{Z}$. The value of T_0 must be relatively small ($T_0/f_s \approx 15$ ms, where f_s is the audio sampling frequency). This is useful for assuming signal stationarity before and after the acoustic event and to avoid signal reflections as much as possible.
2. Whenever the amplitude of the acquired signal is greater than the threshold at some node $m = i$, $|x_i[t_a]| > \gamma$, the node i sends an *event warning alert* to the rest of nodes specifying its node identifier ($m = i$) and onset time $\tau_i = t_a$.
3. When a node receives a warning alert of an event at $n = t_a$ from another node, the receiving node j obviates its own threshold and stores the samples going from $t_0 = t_a + t_{ij} - 2T_0$ to $t_e = t_a + t_{ij} + T_0$, where $t_{ij} \in \mathbb{Z}$ is the maximum TDOA between the alerting node i and the receiving node j, i.e.
$$t_{ij} = \left\lceil f_s \frac{r_{ij}}{c} \right\rceil. \tag{30}$$
This gives a total of $K = 3T_0$ stored samples at each node, containing equivalent observation time ranges. Figure 3 shows the times involved in this windowing process.
4. All the nodes run Algorithm 2, returning to the sink an estimated onset time and some additional clock information for synchronization. However, since it is likely that the signal may not arrive to all the sensors with sufficient strength (due to blocking effects, distance, etc.), only those nodes with $SNR > \zeta$ will send their $\hat{\tau}_m$, where $\zeta \in \mathbb{Z}$ is another threshold set by the user. The onset times received by the sink are used to obtain an estimate of the source location.

This approach is represented in Figure 2(b). The nodes do not run the algorithm continuously, but they compare the amplitude of the signal with a threshold (bright gray). When a node detects a high amplitude signal, it sends an event warning alert to the rest of nodes. Finally, all the nodes estimate their onset time and send it to the central node.

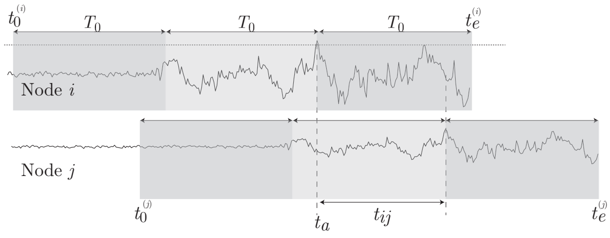

Figure 3: Diagram showing the time instants involved in the node signal windowing process.

3.3. Synchronization

Source localization is a demanding application in terms of time synchronization accuracy. Not only the onset times estimated by all the sensors must all have a common time reference, but also the time instants informed by event warning alert messages. Different types of time synchronization protocols have been proposed to solve timing issues. In the proposed method, protocols based on sender-receiver synchronization such as the the synchronization protocol for sensors networked (TPSN) can be specially useful. In this context, synchronization can be efficiently performed by embedding timing information into the information messages exchanged between nodes. This is achieved by using MAC layer time stamping [54]. Section 5.2 provides some detailed information about sensor synchronization for a real network deployment.

4. Source localization

While the processing at the wireless nodes has very strong restrictions for saving battery life and allowing for long-term monitoring, the sink is assumed to be a more powerful machine capable of performing the data fusion part of the method, which consists in generating a location estimate from the received onset times. In this paper, the selection of an appropriate location estimator fitting the needs of the system is left to the user. Nevertheless, the localization problem is next described to provide the reader with a detailed description of the task to be addressed by the central node.

4.1. Problem formulation

For the sake of simplicity, assume that an onset time from each node has been received at the master node or sink. The positions of the nodes in a 3-D Cartesian coordinate system are:

$$\mathbf{p}_m \triangleq [x_m, y_m, z_m]^{\mathrm{T}}, \quad m = 0, 1, \ldots, M-1, \tag{31}$$

which are assumed to be known. Without loss of generality, the first node ($m = 0$) is regarded as the reference and is placed at the origin, i.e. $\mathbf{p}_0 = [0, 0, 0]^{\mathrm{T}}$. The source is located at $\mathbf{p}_s \triangleq [x_s, y_s, z_s]^{\mathrm{T}}$. The distance difference of sensors i and j from the source is

$$d_{ij} \triangleq \|\mathbf{p}_i - \mathbf{p}_s\| - \|\mathbf{p}_j - \mathbf{p}_s\|, \quad i, j = 0, 1, \ldots, M-1. \tag{32}$$

This difference is proportional to the TDOA, which can be obtained from the onset time difference as

$$\tau_{ij} \triangleq \tau_i - \tau_j = (\tau_{is} + t_s) - (\tau_{js} + t_s) = \tau_{is} - \tau_{js}, \tag{33}$$

thus, $d_{ij} = c \cdot \tau_{ij}$, where c is the speed of sound.

The localization problem consists in estimating $\mathbf{p}_s$, given the set of $\mathbf{p}_m$ and τ_{ij}. In the absence of noise, the space spanned by these TDOAs is $(M-1)$-dimensional, thus, $M-1$ linearly independent TDOAs determine all the others. For simplicity, the problem is usually formulated by considering $M-1$ TDOAs as the basis for this space, considering sensor 0 as the reference. Let d_{m0} denote the source's range difference between sensor m and sensor 0, i.e.

$$d_{m0} = \|\mathbf{p}_m - \mathbf{p}_s\| - \|\mathbf{p}_s\|, \quad m = 1, \dots, M-1. \tag{34}$$

In practice, the TDOAs have measurement errors, so that the observed range differences are modeled as:

$$\hat{d}_{m0} \triangleq d_{m0} + \epsilon_{m0}, \tag{35}$$

where ϵ_{m0} is the measurement error and $\hat{d}_{m0} = c \cdot (\hat{\tau}_m - \hat{\tau}_0)$.

Most closed-form localization algorithms try to approximate the LS criterion, given by

$$J \triangleq \sum_{m=1}^{M-1} (\hat{d}_{m0} - d_{m0})^2. \tag{36}$$

In this problem, an observed range difference d_{m0} defines a half-hyperboloid in 3-D space. All points lying on this hyperboloid are potential source locations, since all these points have the same range difference d_{m0} to sensor m and the reference sensor. Note that the minimization of Eq.(36) leads to a solution that has the shortest distance to all hyperboloids associated with different microphone pairs.

4.2. Spherical equations

The LS problem can be linearized by introducing the source range $\|\mathbf{p}_s\|$ as an additional variable. It follows from (34) that

$$\|\mathbf{p}_m - \mathbf{p}_s\|^2 = (d_{m0} + \|\mathbf{p}\|)^2, \tag{37}$$

which yields the following equations in the unknown vector $\mathbf{p}_s$:

$$d_{m0}\|\mathbf{p}_s\| + \mathbf{p}_m^{\mathrm{T}}\mathbf{p}_s = b_m, \quad m = 1, \dots, M-1, \tag{38}$$

where

$$b_m \triangleq \frac{\|\mathbf{p}_m\|^2 - d_{m0}^2}{2}. \tag{39}$$

Note that, the spherical equations (38) are linear in $\mathbf{p}_s$ given $\|\mathbf{p}_s\|$ and vice versa. In order to solve the problem, several approaches have been suggested, such as the unconstrained and constrained LS [55] solutions, the spherical intersection [56] and interpolation [39] methods or the linear-correction LS [57] algorithm.

5. Experiments

5.1. Simulation Results

The performance of the proposed approaches has been evaluated under different degrees of noise and reverberation. To this end, a significant set of acoustic simulations using the image-source simulation method [58] was processed. The simulations were generated by using a sampling frequency of $f_s = 8192$ Hz, which is the same as the one used by the sensors considered in Section 5.2. Different combinations of SNR values ($SNR \in \{20, 15, 10, 5, 0\}$) and wall reflection factors ρ ($\rho \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$) have been considered, simulating 200 different sensor setups for each combination. For each run, $M = 5$ sensors were randomly distributed on a plane within a room of dimensions 10 m $\times$8 m $\times$3 m. Moreover, the experiments were repeated for two types of acoustic sources: an impulsive source (broken glass) and a male speech source, making a total of 20,000 simulations. Errors related to synchronization or to the position of the sensors were not considered. Since most localization methods are approximations of the LS solution and the resulting localization performance is also dependent on the selected method, we preferred to avoid any specific closed-form localization method and solved numerically the LS problem by using a spatial grid search with a resolution of 0.1 m. The results are presented in terms of the *Mean Absolute Error* (MAE) of the estimated location with respect to the true source position. For the batch approach, the amplitude threshold γ was set to half the maximum possible signal amplitude, while the number of samples for estimating θ_0 and θ_1 were $T_0 = 200$. For the on-line approach, the known SNR was used as a priori information to set the value of θ_1^2, while the decision threshold was set to $\eta = 500$.

Figure 4(a) and Figure 4(b) show the localization accuracy results obtained for the batch and the on-line approaches, respectively. It can be observed that the batch approach is highly robust to different SNR conditions, providing similar localization accuracy for all the tested SNR values, except for SNR= 0 dB. In this extreme case, the amplitude threshold detection stage fails in deciding the existence of an acoustic event, especially when the source is speech. In fact, the results are considerably worse for the speech source, being significantly more affected by noise and reverberation than the impulsive sound. In any case, the MAE is around 0.25 m for the broken glass source and 1.25 m for the speech source in moderate noisy and reverberant conditions.

The results for the on-line approach (Figure 4(b)) show that, while it is more sensible to the actual SNR condition than the batch approach, the robustness against reverberation is considerably better. The average localization accuracy obtained under very adverse conditions ($\rho = 0.9$, SNR= 5 dB) is better than the one provided by the batch approach. This is a somewhat expected result, since the decision between the 'event' and 'no event' hypotheses is made based on the computed statistics instead of a simple amplitude threshold. Moreover, the algorithm makes use of some a priori information (the actual simulated SNR value). Again, the differences obtained between speech and impulsive sound sources are quite significant (note that the

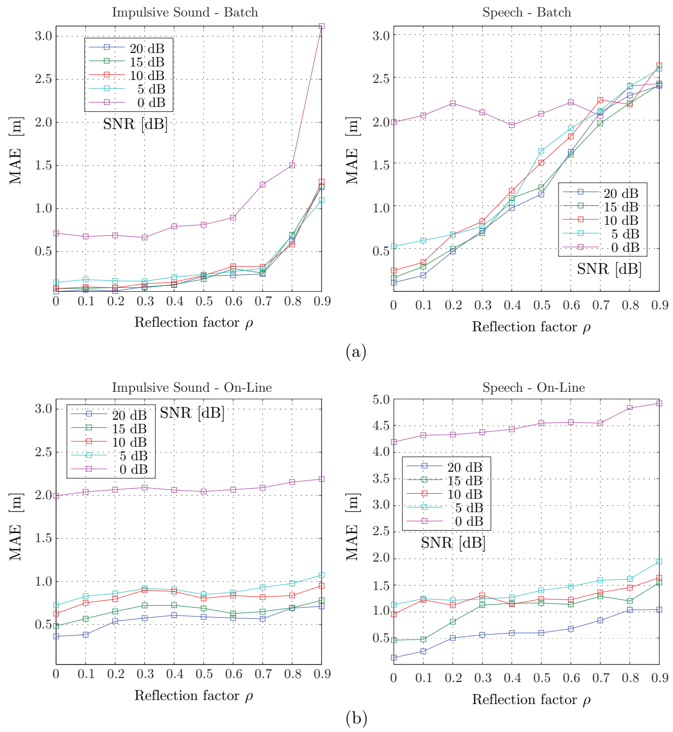

Figure 4: Localization error for different simulated environments using an impulsive sound source and a male speech source. (a) Batch approach. (b) On-Line approach

graphs have different vertical scales), being speech more problematic. The variance change is so small when SNR= 0 dB that the algorithm fails in deciding that an event has actually occurred, propagating the error to the TDOA estimates and resulting in poor localization accuracy. Nevertheless, the average location error in moderate acoustic conditions ($\approx$ 0.75 m) is acceptable for most practical applications.

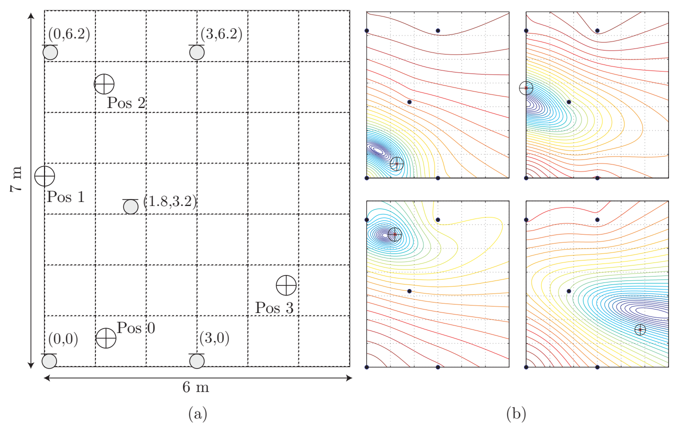

Figure 5: Real deployment set-up. (a) Room dimensions, motes and source locations. (b) Examples of cost function contours obtained from the estimated TDOAs at the four source locations.

5.2. *Real deployment*

In order to demonstrate the validity of the proposed method in a low-cost WASN, we have implemented the CUSUM batch approach in a network composed by 5 motes *Tmote-Invent*, which have very limited sound recording and processing capabilities. This mote is based on a 16 bit microcontroller (Texas Instruments MSP430), having 10KB of RAM and 48KB of Flash memory. The wireless transceiver is a Texas Instruments CC2420, which is IEEE 802.15.4 compliant. The motes run the TinyOS-1.0 operating system, which is written in a variant of C named NesC. The microphone amplifier allows for a variable dynamic compression ratio. Although compression introduces some distortion, it permits the recording of low-power signals and prevents clipping of high-power sounds. Since each captured audio sample is stored in 2 bytes of RAM memory (which is only 10KB), a small sampling frequency must be used to allow for a reasonable observation window. In our experiments, the sampling frequency was 8192 Hz, which led to an available recording buffer of 4096 samples (0.5 s). Note that this application requires all the motes to be in the same broadcast domain. The radio range provided by the motes is about 50 m outdoors or 30-40 m indoor, depending on the environment. Then, localization applications using the proposed method should be constrained to these ranges. Since a common timing reference must be available, synchronization issues must be carefully addressed. For this application, having continuous synchronized clocks among all the motes is not essential, since common time references are only needed once an event has occurred. Onset time estimation at each mote is performed according to its own local clock. Then, timing information is only sent to other nodes whenever an estimated onset time is transmitted. Both types of information are embedded into the same data packet. The translation of timing information within the central node is done by computing the instantaneous offset between the local clocks of both motes. To estimate this offset, the so-called SFD field in the 802.15.4 packet is used, allowing to get simultaneous clock time stamps in both motes. The measured mean error in the offset estimation is 3.177 μs, which does not significantly affect the translation of onset times to a common time reference.

A set of four different locations in a rectangular room of dimensions 7 m $\times$ 6 m was considered to evaluate the performance of the method. The measured reverberation time was T_{60} = 0.68 s and the background noise was below 30 dB(A). The reproduced sound source was the same impulsive sound (broken glass) used in the simulations. For each location, four different measurements were conducted, making a total of 16 experiments. The set-up is depicted in Figure 5(a), while examples of the location LS cost function obtained for the four source positions are represented as contour plots in Figure 5(b). The event was correctly detected in all the experiments, obtaining an average MAE of 0.8 m. Note that this is a very good result taking into account that there might be some measurement errors both in the mote and source positions. In any case, these results confirm the validity of the proposed approach in real situations and using low-cost hardware in the nodes.

6. Conclusion

This paper presented two signal processing strategies based on the cumulative sum framework for acoustic event detection and localization in WASNs. These methods are aimed at providing efficient solutions for long-term acoustic monitoring using low-cost hardware with constrained acquisition and processing resources. Both methods follow a hybrid scheme: all the nodes perform locally the detection of an acoustic event and the estimation of its onset time, sending their results to a central node that estimates the location of the source from the reported information. However, while these two algorithms are derived from the same change detection statistical framework, they differ significantly in terms of node communication, a priori information and parameters estimates. Simulations using different degrees of reverberation and signal-to-noise ratios have demonstrated that a relatively good localization accuracy can be achieved by means of these algorithms, even in adverse acoustic conditions. Finally, the validity of the method has been demonstrated by experiments conducted over a real network deployment using motes with very limited processing resources, providing an average localization accuracy of 0.8 m.

References

[1] D. Culler, D. Estrin, M. Srivastava, Overview of sensor networks, Computer 37 (8) (2004) 41–49.

[2] I. F. Akyildiz, T. Melodia, K. R. Chowdhury, A survey on wireless multimedia sensor networks, Computer Networks 51 (2007) 921–960.

[3] A. Bertrand, Applications and trends in wireless acoustic sensor networks: a signal processing perspective, in: Proceedings of the 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT), IBBT, Ghent, Belgium, 2011.

[4] A. Bertrand, Signal processing algorithms for wireless acoustic sensor networks, Ph.D. thesis, Katholieke Universiteit Leuven, Leuven, Belgium (May 2011).

[5] H. Wang, Wireless sensor networks for acoustic monitoring, Ph.D. thesis, University of California, Los Angeles (UCLA), Los Angeles, California (2006).

[6] A. Swami, Q. Zhao, Y. W. Hong, T. Lang (Eds.), Wireless sensor networks: signal processing and communications, Wiley, West Sussex, England, 2007.

[7] Y. Hioka, W. B. Kleijin, Distributed blind source separation with an application to audio signals, in: Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 2011, pp. 233–236.

[8] S. Wehr, I. Kozintsev, R. Lienhart, W. Kellermann, Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation, in: Proceedings of the 6th International Symposium on Multimedia Software Engineering, Miami, Florida, USA, 2004.

[9] E. Robledo-Arnuncio, T. S. Wada, B.-H. Juang, On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation, in: Proceedings of the 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2007, pp. 34–37.

[10] A. Bertrand, J. Callebaut, M. Moonen, Adaptive distributed noise reduction for speech enhancement in wireless acoustic sensor networks, in: Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC), Tel Aviv, Israel, 2010.

[11] S. Doclo, M. Moonen, T. V. den Bogaert, J. Wouters, Reduced bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids, IEEE Transactions on Audio, Speech and Language Processing 17 (2009) 38–51.

[12] S. Markovich-Golan, S. Gannot, I. Cohen, A reduced bandwidth binaural mvdr beamformer, in: Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC), Tel Aviv, Israel, 2010.

[13] A. Bertrand, M. Moonen, Distributed adaptive node-specific signal estimation in fully connected sensor networks - part I: sequential node updating, IEEE Transaction on Signal Processing 58 (10) (2010) 5277–5291.

[14] A. Bertrand, M. Moonen, Distributed node-specific LCMV beamforming in wireless sensor networks, IEEE Transactions on Signal Processing 60 (1) (2012) 233–246.

[15] S. Markovich-Golan, S. Gannot, I. Cohen, Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks, IEEE Transactions on Audio, Speech, and Language Processing 21 (2) (2013) 343–356.

[16] M. Wang, L. Ci, P. Zhan, Y. Xu, Acoustic source localization in wireless sensor networks, in: Proceedings of the 2007 IEEE Workshop on Intelligent Information Technology Application, Zhang Jiajia, China, 2007, pp. 196–199.

[17] Z. M. Saric, D. D. Kukolj, N. D. Teslic, Acoustic source localization in wireless sensor network, Circuits, Systems and Signal Processing 29 (5) (2010) 837–856.

[18] J. C. Chen, K. Yao, R. E. Hudson, Source localization and beamforming, IEEE Signal Processing Magazine 19 (2) (2002) 30–39.

[19] T. Ajdler, I. Kozintsev, R. Lienhart, M. Vetterli, Acoustic source localization in distributed sensor networks, in: Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2004, pp. 1328–1332.

[20] M. Z. Rahman, G. C. Karmakar, L. S. Dooley, Passive source localization using power spectral analysis and decision fusion in wireless distributed sensor networks, in: Proceedings of the 2005 International Conference on Information Technology: Coding and Computing (ITCC), Vol. 2, Las Vegas, NV, USA, 2005, pp. 260–264.

[21] A. Ledeczi, P. Volgyesi, M. Maroti, G. Simon, Multiple simultaneous acoustic source localization in urban terrain, in: Proceedings of the 4th Symposium on Information Processing in Sensor Networks (IPSN), Los Angeles, CA, USA, 2005, pp. 491–496.

[22] X. Sheng, Y. Hu, Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks, IEEE Transactions on Signal Processing 53 (1) (2005) 44–53.

[23] N. Kitakoga, T. Ohtsuki, Distributed EM algorithms for acoustic source localization in sensor networks, in: Proceedings of the 2006 IEEE Vechicular Technology Conference (VTC), Montreal, Canada, 2006.

[24] M. G. Rabbat, R. D. Nowak, Decentralized source localization and tracking, in: Proceedings of the 2004 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, Canada, 2004.

[25] X. Y. Pi, H. Y. Yu, A distributed and cooperative target localization algorithm in wireless sensor networks, in: Proceedings of the 6th Inernational Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Dalian, China, 2005, pp. 887–889.

[26] S. M. Williams, K. D. Frampton, I. Amundson, P. L. Schmidt, Decentralized acoustic source localization in a distributed sensor network, Applied Acoustics 67 (2006) 996–1008.

[27] C. Meesookho, U. Mitra, S. Narayanan, On energy-based acoustic source localization for sensor networks, IEEE Transactions on Signal Processing 56 (1) (2008) 365–377.

[28] 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: Proceedings of the 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2007.

[29] J. C. Chen, L. Yip, J. Elson, H. Wang, D. Maniezzo, R. E. Hudson, K. Yao, D. Estrin, Coherent acoustic array processing and localization on wireless sensor networks, Proceedings of the IEEE 91 (8) (2003) 1154–1162.

[30] J. C. Chen, R. E. Hudson, K. Yao, Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field, IEEE Transactions on Signal Processing 50 (8) (2002) 1843–1854.

[31] J. Chen, J. Benesty, Y. Huang, Time delay estimation in room acoustic environments: An overview, EURASIP Journal on Advances in Signal Processing 2006 (2006) 1–19, article ID 26503.

[32] H. F. Silverman, Y. Yu, J. M. Sachar, W. R. Patterson III, Performance of real-tiem source-location estimators for a large-aperture microphone array, IEEE Transactions on Speech and Audio Processing 13 (2005) 593–606.

[33] J. H. DiBiase, A high accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays, Ph.D. thesis, Brown University, Providence, RI, USA (2000).
[34] C. H. Knapp, G. C. Carter, The generalized correlation method for estimation of time delay, IEEE Transactions on Acoustics, Speech, and Signal Processing 24 (4) (1976) 320–327.
[35] J. Chen, J. Benesty, Y.-A. Huang, Performance of GCC- and AMDF-based time-delay estimation in practical reverberant environments, EURASIP Journal on Applied Signal Processing 2005 (1) (2005) 25–36.
[36] J. Benesty, Adaptive eigenvalue decomposition algorithm for passive acoustic source localization, Journal of the Acoustical Society of America 107 (2000) 384–391.
[37] M. Wax, T. Kailath, Optimum localization of multiple sources in passive array, IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-31 (1983) 1210–1217.
[38] J. O. Smith, J. S. Abel, Closed-form least-squares source location estimation from range-difference measurements, IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-35 (12) (1987) 1661–1669.
[39] J. O. Smith, J. S. Abel, The spherical interpolation method of source location, IEEE Journal of Oceanic Engineering 12 (1) (1987) 246–252.
[40] A. Beck, P. Stoica, J. Li, Exact and approximate solutions of source localization problems, IEEE Transactions on Signal Processing 56 (5) (2008) 1770–1778.
[41] A. Canclini, E. Antonacci, A. Sarti, S. Tubaro, Acoustic source localization with distributed asynchronous microphone networks, IEEE Transactions on Audio, Speech and Signal Processing 21 (2) (2013) 439–443.
[42] Y. Jia, Y. Luo, Y. Lin, I. Kozintsev, Distributed microphone arrays for digital home and office, in: Proceedings of the 2006 IEEE nternational Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, 2006.
[43] P. Julian, G. A. Andreou, L. R. Riddle, S. Shamma, D. Golberg, G. Cauwenberghs, A comparative study of sound localization algorithms for energy aware sensor network nodes, IEEE Transactions on Circuits and Systems 51 (4) (2004) 640–648.
[44] K. Liu, X. Liu, X. Li, Guoguo: enabling fine-grained in-door localization via smartphone, in: Proceedings of the 11th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys), Taipei, Taiwan, 2013.
[45] P. Bergamo, S. Asgari, H. Wang, D. Maniezzo, L. Yip, R. E. Hudson, K. Yao, D. Estrin, Collaborative sensor networking towards real-time acoustical beamforming in free-space and limited reverberance, IEEE Transactions on Mobile Computing 3 (3) (2004) 211–223.
[46] M. Maroti, G. Simon, A. Ledeczi, J. Sztipanovits, Shooter localization in urban terrain, Computer 37 (8) (2004) 60–61.
[47] G. Simon, M. Maroti, A. Ledeczi, G. Balogh, B. Kusy, A. Nadas, G. Pap, J. Sallai, K. Frampton, Sensor network-based countersniper system, in: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys), Baltimore, MD, USA, 2004.
[48] J. Hill, D. Culler, Mica: A wireless platform for deeply embedded networks, IEEE Micro 22 (6) (2002) 12–24.
[49] E. S. Page, Continuous inspection scheme, Biometrika 41 (1954) 100–115.
[50] S. Kay, Fundamentals of statistical signal processing, volume 2: Detection theory, Prentice Hall, 1998.
[51] S. Kay, Fundamentals of statistical signal processing: volume 1: Estimation theory, Prentice Hall, 1993.
[52] S. Gazor, W. Zhang, Speech probability distribution, IEEE Signal Processing Letters 10 (7) (2003) 204–207.
[53] M. Basseville, I. Nikiforov, Detection of abrupt changes: Theory and application, Prentice Hall, 1993.
[54] S. Ganeriwal, R. Kumar, M. B. Srivastava, Timing-sync protocol for sensor networks, in: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys), Los Angeles, CA, USA, 2003, pp. 138–149.
[55] P. Stoica, J. Li, Source localization from range-difference measurements, IEEE Signal Processing Magazine 23 (6) (2006) 63–66.
[56] H. C. Schau, A. Z. Robinson, Passive source locaization employing intersecting spherical surfaces from time-of-arrival differences, IEEE Transactions on Acoustics, Speech, and Signal Processing 35 (1987) 1223–1225.
[57] Y. Huang, J. Benesty, G. W. Elko, R. M. Mersereau, An efficient linear-correction least-squares approach to source localization, in: Proceedings of the 2001 IEEE Workshop on Applications of Signal Porcessing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2001.
[58] J. B. Allen, D. A. Berkley, Image method for efficiently simulating small-room acoustics, Journal of the Acoustical Society of America 65 (3) (1979) 943–950.