Simultaneous Ranging and Self-Positioning in Unsynchronized Wireless Acoustic Sensor Networks

Maximo Cobos, *Senior Member, IEEE,* Juan J. Perez-Solano, Óscar Belmonte, German Ramos, *Member, IEEE,* and Ana M. Torres

***Abstract*—Automatic ranging and self-positioning is a very desirable property in wireless acoustic sensor networks (WASNs) where nodes have at least one microphone and one loudspeaker. However, due to environmental noise, interference and multipath effects, audio-based ranging is a challenging task. This paper presents a fast ranging and positioning strategy that makes use of the correlation properties of pseudo-noise (PN) sequences for estimating simultaneously relative time-of-arrivals (TOAs) from multiple acoustic nodes. To this end, a proper test signal design adapted to the acoustic node transducers is proposed. In addition, a novel self-interference reduction method and a peak matching algorithm are introduced, allowing for increased accuracy in indoor environments. Synchronization issues are removed by following a BeepBeep strategy, providing range estimates that are converted to absolute node positions by means of multidimensional scaling (MDS). The proposed approach is evaluated both with simulated and real experiments under different acoustical conditions. The results using a real network of smartphones and laptops confirm the validity of the proposed approach, reaching an average ranging accuracy below 1 centimeter.**

***Index Terms*—Ranging, localization, wireless acoustic sensor networks, pseudo-noise seequences.**

I. Introduction

THE processing capabilitites of mobile computing and communication platforms such as laptops, smartphones, small single-board computers and tablets are increasing everyday. These devices, often equipped with powerful processors, are conceived as multimedia communication centers with microphones and loudspeakers. An ad hoc network of such devices is here collectively referred to as a wireless acoustic sensor network (WASN). These kind of networks are receiving the attention of the signal processing community for the wide range of applications that are currently emerging [1], such as smart conference rooms [2], source localization [3], speech enhancement [4] or environmental monitoring [5]. While no dedicated resources are needed to form such a network, there are important issues that must be addressed before applying many signal processing approaches, most of them related to the need of knowing in advance the three dimensional positions of the devices. Therefore, automatic ranging and positioning is considered to be a key enabling technology, since even relatively small uncertainties in the location of the nodes affect the overall performance of these systems.

Ranging in wireless sensor networks is typically achieved through measuring the time-of-arrival (TOA) and/or the received signal strength (RSS) of acoustic or radio signals [6], [7]. The RSS approach, while being significantly inexpensive, incurs significant errors due to channel fading, long distances and multipath [8], [9]. Since the ranging accuracy depends both on the signal propagation speed and the precision of the TOA measurement, acoustic signals are usually preferred because their relative low speed [10]–[12]. In typical WASNs, TOA measurements are often performed with pairs of nodes taking timestamps of their respective local clock at the moment when the test signal is emitted or received [13], [14]. Possible sources of error include clock skew/drift between devices, misalignment between timestamps and actual signal emissions, and errors in detecting the arrival of the sound signals [3]. While many formulations assume that all the nodes are on a synchronized setup [11], [12], [15], a typical distributed setup must explicitly account for the errors due to lack of temporal synchronization among the devices. Moreover, to achieve high ranging accuracy, it is critical to precisely locate the arrival of the test signals used in the system. This is particularly challenging in WASNs using mobile devices since the transducers typically cover a narrow band of the spectrum [16], [17]. In addition, multipath effects in indoor environments, background noise, node interference and signal distortion are very relevant aspects that motivate a proper signal design. In [17]–[20], closed-form solutions are provided for the self-localization problem considering possible errors in the measured signal arrivals. Some approaches rely on additional knowledge, given by the true emission times [19] or the existence of loudspeakers placed at known positions [18], [21]. For example, the recent work in [21] proposes a self-localization method for mobile devices that is based on the cyclical emission of known probe signals emitted from loudspeakers at known locations. While this method may be very useful for tracking purposes, simultaneous emissions must be avoided and requires the deployment of additional hardware (loudspeakers). As with our proposed approach, a peak detection strategy is also given, but taking advantage of the cyclical loudspeaker emissions. Another recent approach employs time-difference of arrivals (TDOAs) to jointly estimate sensor and source locations [22].

M. Cobos and J.J. Perez-Solano are with the Computer Science Department, Universitat de València, Burjassot, Valencia, 46100 Spain e-mail: maximo.cobos@uv.es, juan.j.perez-solano@uv.es

O. Belmonte is with the Department of Computer Languages and Systems, Universitat Jaume I, Castellón de la Plana, 12071, Spain.

G. Ramos is with the ITEAM Institute, Universitat Politècnica de València, Valencia, 46022, Spain.

A. Torres is with the I.E.E.A.C. Department, Universidad de Castilla-la-Mancha, Cuenca, 16071, Spain. This work was supported by the Spanish Ministry of Economy and Competitiveness under Grant TIN2015-70202-P, TEC2012-37945-C02-02 and FEDER funds.

However, to the best of our knowledge, all the previous work in the field has considered a non-simultaneous calibration process where test signals are emitted one at a time, reducing the node interference but significantly increasing the total calibration time.

This paper proposes a complete acoustic ranging system that allows to perform the ranging process simultaneously, with all the nodes emitting and recording at the same time. As a result, the proposed method reduces significantly the total calibration time with respect to a pair-wise non-simultaneous framework. Several contributions are proposed and evaluated in both simulated and real environments. First, we propose the design of test (calibration) signals that make use of the correlation properties of pseudo-noise (PN) sequences, which have already been shown to be especially useful in acoustic ranging applications [23]–[25]. These sequences modulate a sine wave to produce a set of test signals that can be adapted to the bandwidth of the transducers. The correlation properties of the resulting signals make the system robust to background noise, multipath effects and node interference. Second, due to the close distance between loudspeakers and sensors at each node, the own signal emitted by each node is received at a much higher level than the ones from the other nodes. To face this problem, we propose a self-interference reduction approach that allows to mitigate this effect and listen properly to the test signals arriving from the rest of nodes. Third, the modulated test signal produces a filtering effect that affects the identification of the direct-path delay corresponding to the TOA of the node signals. To this end, we propose a peak matching method based on the local signal-to-noise ratio (SNR) that allows to identify the real TOA in the presence of rejection noise. Finally, a BeepBeep [13] strategy and conventional multidimensional scaling (MDS) [26] are used to derive pair-wise ranges and absolute coordinates of the nodes up to a rotation and translation.

The intended applications of our proposed approach are related to node positioning and calibration of systems involving ad-hoc microphone arrays, such as acoustic source localization, speech enhancement and beamforming, spatial statistics or tuning of spatial audio systems. The method assumes the nodes to be stationary during the calibration process and it is not currently optimized for tracking moving nodes. While some operations might be computationally intensive for a resource-constrained device, the proposed approach offers enough flexibility to balance properly the computing load in the system by using, if necessary, a central node.

The paper is organized as follows. Section II presents the formulation of the problem to be addressed by the proposed system. Section III discusses the design of the test signals to be used by the nodes in the network by taking into account the correlation properties of PN sequences. Section IV describes the self-interference reduction part of the method, following in Section V with the proposed peak matching approach for TOA estimation. The set of estimated TOAs conform the required input for the ranging and positioning strategies in Section VI. Experiments with simulated and real setups are presented in Section VII. Finally, conclusions are summarized in Section VIII.

II. Problem Formulation

Let us consider a network of D unsynchronized acoustic nodes such as the one shown in Fig.1, where each node has its own microphone and loudspeaker and has been assigned a unique test signal. Once the calibration initialization command has been issued, each node starts the recording and playback of its own test signal, capturing as well the test signals emitted by the rest of nodes. It is here assumed that a small time is allowed in the system to let all the nodes start their recording before any of them emits any sound. We use indexes (q, d) to denote a pair of nodes in the system. The example of Fig.1 represents how the node $q = 1$ is receiving the test signals from the rest of D nodes, including its own one ($d = 1$). Similarly, the rest of nodes are simultaneously capturing all the emitted test signals. The discrete-time signal received by the microphone of the qth node can be expressed as

$$x_q(n) = \sum_{d=1}^{D} s_d(n - t_{q,d}) * h_{q,d}(n) + n_q(n), \qquad (1)$$

where $h_{q,d}(n)$ is the impulse response from the loudspeaker at node d to the microphone at node q, $s_d(n)$ denotes the test signal from the dth node and $n_q(n)$ is a noise term that consists of additive background noise. The delay term $t_{q,d}$ arises due to the fact that nodes do not start simultaneously the recording and playback of the involved test signals. In fact, $t_{q,d}$ would be zero in a synchronized setup where the nodes start their recording and transmission at the same time. However, in practical systems, $t_{q,d}$ is modeled as a non-deterministic delay given by

$$t_{q,d} = t_d^{(e)} - t_q^{(r)}, \qquad (2)$$

where $t_d^{(e)}$ denotes the emission start time at node d and $t_q^{(r)}$ is the recording start time at node q, being both discrete-time instants. The emission start time is defined as the time after which the sound is actually emitted from the speaker once the calibration initialization instruction has been sent. Similarly, the recording start time is the time after which the sound is actually captured by the microphone once the calibration command has been received. Both times include network delays, the delays in setting up the audio buffers and other physical times. These times are generally unknown and depend on the particular audio hardware and the system state such as the processor workload, interrupts, and the processes scheduled at the given instant. Note that additional sources of synchronization errors may arise in practice, such as the ones due to the small differences in the sampling rate used at each node. However, since the total calibration time is greatly reduced in a simultaneous calibration framework, the total drift caused by the sampling frequency mismatch is here considered to be negligible [27] and below the achieved accuracy.

The impulse response term can be further decomposed into a direct-path component and a reverberant component as follows:

$$h_{q,d}(n) = \alpha_{q,d}\delta(n - \tau_{q,d}) + h^{\mathrm{r}}_{q,d}(n), \qquad (3)$$

where $\alpha_{q,d}$ is an attenuation factor, $\delta(n)$ is the impulse function and $h^{\mathrm{r}}_{q,d}(n)$ is the reverberant part of the impulse

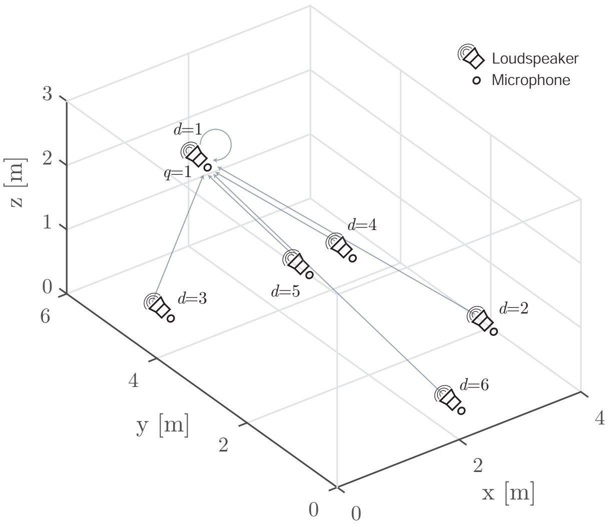


Fig. 1. An example WASN with $D = 6$ nodes.

response. Early reflections are assumed to be contained in the term $h^{\mathrm{r}}_{q,d}(n)$ and, as discussed in Section V, the proposed peak matching method is aimed at discriminating the direct-path component from peaks corresponding to early reflections. The term $\tau_{q,d}$ is the propagation delay, given by

$$\tau_{q,d} = \left\lfloor \frac{1}{c} r_{q,d} \cdot f_s \right\rceil, \tag{4}$$

where $r_{q,d}$ is the distance between the loudspeaker at the dth node to the microphone of the qth node, c is the speed of sound (≈ 340 m/s at sea level and 15°C), f_s is the sampling frequency and $\lfloor \cdot \rceil$ denotes the integer rounding operator. By including the delay $t_{q,d}$ into $h_{q,d}(n)$, Eq.(1) can be written as

$$x_q(n) = \sum_{d=1}^{D} s_d(n) * \tilde{h}_{q,d}(n) + n_q(n), \tag{5}$$

where $\tilde{h}_{q,d}(n) = h_{q,d}(n - t_{q,d})$ is a delayed impulse response. Before approaching the ranging problem, each sensor must obtain a timestamp corresponding to the TOA of each test signal, denoted as $\zeta_{q,d}$. To clarify how all these times and signals are related, Fig.2 shows schematically the times involved at two nodes q and d of the system, giving rise to their recorded signals $x_q(n)$ and $x_d(n)$. Note that the TOA at each sensor corresponds to the addition of the unknown non-deterministic delay and the propagation delay, i.e.,

$$\zeta_{q,d} = t_{q,d} + \tau_{q,d}. \tag{6}$$

The self microphone-loudspeaker distance at each node (r_{qq}) does not have to be equal for all the nodes but it is assumed to be a known parameter. The ranging estimation problem consists in estimating the node-to-node distances $\Delta_{q,d}$ for all pairs (q, d). It is assumed that the node positions $\mathbf{p}_q = [\mathrm{x}_q, \mathrm{y}_q, \mathrm{z}_q]^{\mathrm{T}}$, $q = 1, \ldots, D$ are at the center point between the microphone and the loudspeaker, so that $\Delta_{q,d} = \|\mathbf{p}_q - \mathbf{p}_d\|$.

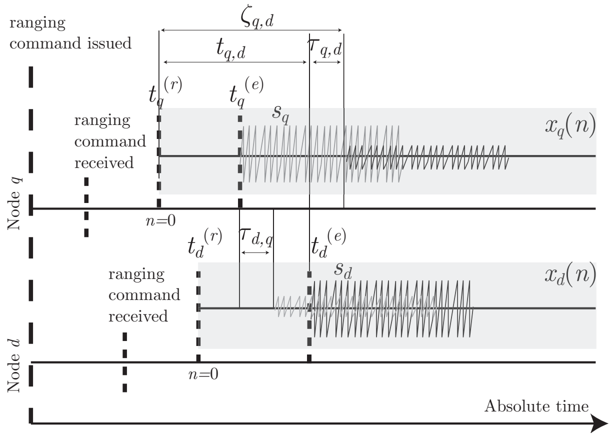


Fig. 2. Times involved at the different nodes of the system.

III. Test Signal Design

This section describes the design of the test signals used in our proposed approach. On one hand, the signals are derived from families of PN sequences in order to provide proper signal detection and interference rejection features in the proposed positioning framework. On the other hand, the signals must be adapted to the audio bandwidth of the node transducers. The following subsections discuss these issues, reviewing important aspects such as the correlation, duration and bandwidth of the node test signals.

A. Pseudo-Noise Sequence

PN sequences are very well known for their applications in spread-spectrum communications. The selection of spreading codes is of primary importance in the proposed application. In the literature [23], there are a significant variety of codes with well-known features applied principally in the development of multi-access communication systems. A PN sequence comprises an ordered set of P values forming a vector $\mathbf{g} = [g(0), g(1), \ldots, g(P-1)]^{\mathrm{T}}$. Each element in $\mathbf{g}$ represents a modulation chip that can only take two values $g(p) \in \{-1, +1\}$. In spread-spectrum communications a data symbol is combined with a PN sequence to generate a modulated communication signal that occupies a much wider bandwidth. The implementation of these systems requires the codes to provide two key features. The first one is an autocorrelation function with a narrow peak at zero time shift to enable a good detection of the code and to facilitate the code synchronization. The second one is a low cross-correlation between different codes, which is especially important in multi-access systems where the receiver must reject the interference from signals modulated with codes of other users.

Maximum length sequences (MLS or m-sequences) are a commonly used type of PN binary sequence [23]. An m-sequence is generated using a shift register of order m and feedback taps selected according to a primitive polynomial, which provides a code of length $P = (2^m - 1)$. Main features of m-sequences are: a) they are not orthogonal codes, i.e., the cross-correlation of two codes is not zero, b) their autocorrelation is a delta function with a peak value P and c) the number of $(1/-1)$ is balanced.

Another type of well-known PN sequences are Gold codes. The combination of two m-sequences of length P produces a Gold code of the same size. A family of different Gold codes of the same length can be obtained using shifted versions of the original m-sequences. The main advantages of Gold codes as compared to the original m-sequences are the increased number of family codes and slightly better cross-correlation properties, but at the expense of having a worse autocorrelation that ceases to be a delta function [28]. The minimum cross-correlation between the Gold codes can be achieved when the pair of original m-sequences constitutes a preferred pair [29].

According to all these considerations, both MLS and Gold codes are suitable to be used in the system implementation, but they present slight differences in their performance. On one hand, MLS have a delta autocorrelation function, but their cross-correlation depends on the the two specific codes selected from the whole family. On the other hand, Gold codes generated with a preferred pair have better and predictable cross-correlation, but their autocorrelation is not zero at non-zero lags. In the experiments section, both families are evaluated considering the specific conditions of the system setup. However, the formulation followed throughout the rest of the paper considers only the use of m-sequences.

B. Test Signal Duration and Bandwidth

Due to the white-noise-like properties of PN sequences, their power spectral density (PSD) covers the full sampled audio bandwidth (from $f=0$ to $f=f_s/2$) independently of the sampling frequency used for audio playback/recording. Since microphones and loudspeakers of conventional devices have usually a narrow audio bandwidth due to their small size, the PSD of PN sequences does not often match the acoustic requirements of such devices. It is a well-known fact from digital communications that the spread spectrum signal generated from a PN sequence of length P with a continuous-time chip period T_c has a bandwidth given by $W=\frac{1}{T_c}$, which is determined by the first null in its baseband power spectrum. In the time domain, the spread spectrum signal will have a total duration $T_s=P\cdot T_c$. As it will be discussed later, the length of the sequence P determines the robustness and duration of the test signal, while T_c contributes both to the duration of the test signal and to its bandwidth.

In our proposed approach, the power spectrum of the test signal is moved to a comfortable frequency range adapted to the node transducers. To this end, the PN sequence modulates a sinusoidal carrier signal with frequency f_c. As a result, the power spectrum of the test signal is accommodated into a frequency range $[f_c-W,\, f_c+W]$ by adjusting the carrier frequency and the chip duration T_c. As an example, a test signal with bandwidth $2W=4$ kHz centered at $f_c=6$ kHz can be designed by selecting a chip period of $T_c=0.5$ ms. Note that this modulation can then be used to move the PSD of the test signal to an operating region where the acoustic transfer function of the transducers is approximately flat. As described next, a conventional Binary Phase-Shift Keying (BPSK) modulation scheme is used for this purpose [23].

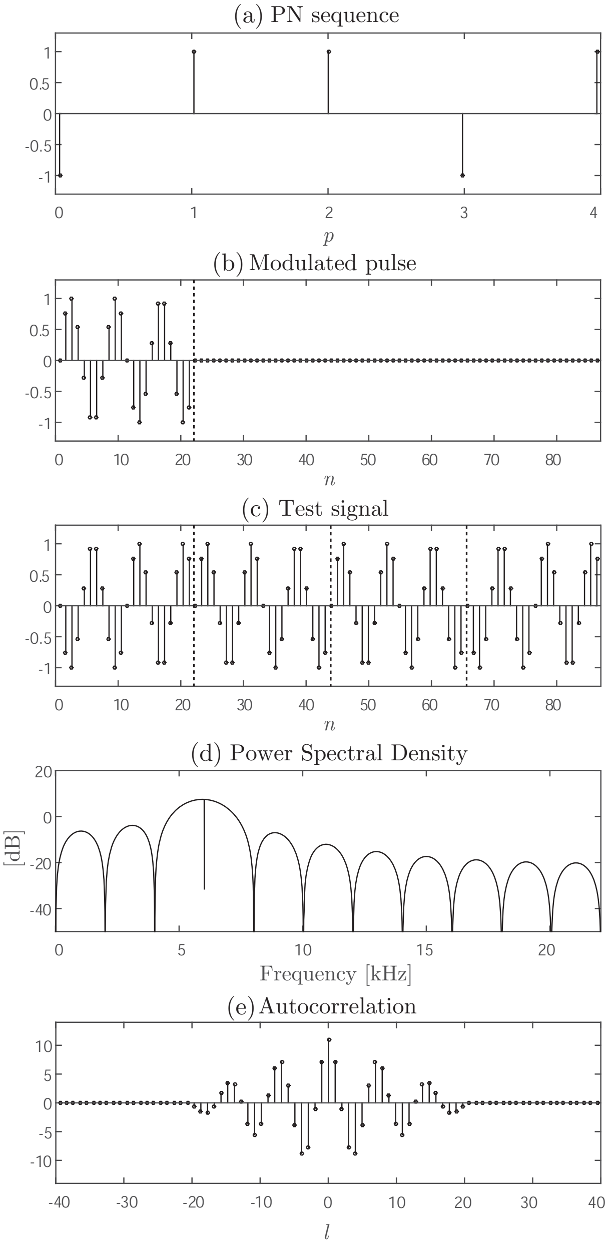


Fig. 3. Example of test signal design with $f_s=44100$, $T_c=5\cdot10^{-4}$ s, $f_c=6$ kHz. (a) Points of PN sequence $g(p)$ (only first 5 values). (b) Modulated chip pulse $z(n)$ ($N_{T_c}=22$, $m_c=3$). (c) Test signal $s(n)$ (corresponding to 4 first chip pulses). (d) Power spectral density of test signal. (e) Autocorrelation of chip pulse $z(n)$.

For convenience, the discrete-time signal generated from the PN sequence is designed with a chip period corresponding to an integer number of samples. Similarly, the carrier frequency is selected so that an integer number of cycles is contained within each chip period. As a result, the modulated chip pulse is defined as

$$z(n)=\begin{cases}\sin\left(\frac{2\pi}{N_{T_c}}m_c n\right), & n=0,\ldots,N_{T_c}-1\\ 0, & \text{elsewhere}\end{cases}\quad, \tag{7}$$

where $N_{T_c}=\lfloor T_c\cdot f_s\rceil$ is the number of samples corresponding to the chip period and $m_c = \lfloor T_c \cdot f_c \rceil$ is the number of carrier cycles contained within the chip period. The resulting test signal is therefore given by a concatenation of phase-modified chip pulses as follows:

$$s_d(n) = \sum_{p=0}^{P-1} \operatorname{sgn}\left(g_d(p)\right) z(n - pN_{T_c}), \tag{8}$$

where $g_d(p)$ is the length-P PN sequence assigned to node d and $\operatorname{sgn}(\cdot)$ is the sign function. An example test signal design is shown in Fig.3, which represents the relations among the original PN sequence (a), the modulated pulse chip (b), the final test signal (c) and its corresponding PSD (d). The chip autocorrelation sequence is also shown in (e) which, as described next, plays an important role in the proposed approach.

Note that the frequency bands selected in the design of the test signals, while being suited to the frequency response of common acoustic transducers, are within the human audible range and can be potentially annoying. However, since the calibration process is simultaneous, the calibration time is greatly reduced and the potential annoyance is minimized. As an example, consider a network with 8 sensors and a 8191-length sequence: the calibration time would be reduced from 3.82 minutes in a pair-wise non-simultaneous calibration system to only 4 seconds.

C. Correlation

An important property of PN sequences is that they can be designed to have an ideal circular autocorrelation function. For a general PN sequence $g(p)$, its circular autocorrelation is computed as

$$\mathring{R}_{gg}(l) = \mathcal{F}^{-1}\left\{G(k) \cdot G^*(k)\right\}_P = P\delta(l), \tag{9}$$

where $G(k)$ is the discrete Fourier transform (DFT) of $g(p)$ and $\mathcal{F}\left\{\cdot\right\}_P^{-1}$ denotes the length-P inverse DFT operator. The index k is the discrete frequency bin index and l is the time-lag index. The symbol $(\cdot)^*$ denotes complex conjugation. Note that $\mathring{R}_{gg}(l)$ is only non-zero for zero time lag ($l = 0$). However, the test signal generated from the PN sequence will not have an ideal autocorrelation, but will contain the effect of the modulated chip pulse $z(n)$. In fact, the circular autocorrelation of any test signal $s(n)$ coming from an ideal length-P PN sequence is given by

$$\mathring{R}_{ss}(l) = PR_{zz}(l), \tag{10}$$

where $R_{zz}(l)$ is the autocorrelation function of $z(n)$:

$$R_{zz}(l) = \begin{cases} \frac{N_{T_c}-|l|}{2}\cos\left(\frac{2\pi}{N_{T_c}}m_c|l|\right) + \\ \frac{1}{2}\cot\left(\frac{2\pi}{N_{T_c}}m_c\right)\sin\left(\frac{2\pi}{N_{T_c}}m_c|l|\right) & |l| < N_{T_c} \\ 0, & \text{elsewhere.} \end{cases} \tag{11}$$

Note that the length of $R_{zz}(l)$ is determined by the chip period N_{T_c}, which at the same time determines the bandwidth of the test signal. As expected, a wider frequency bandwidth results in a narrower autocorrelation function, with a maximum value of $N_{T_c}/2$, as shown in the example of Fig.3(e).

On the other hand, two different PN sequences g and g' are assumed to have a low cross-correlation $\mathring{R}_{g,g'}(l)$, so that their derived test signals s and s' retain the low cross-correlation properties: $\mathring{R}_{ss'}(l) = R_{zz}(l) * \mathring{R}_{gg'}(l) \ll P\frac{N_{T_c}}{2}\,\forall l$.

IV. Self-Interference Reduction

Due to the proximity between the loudspeaker and the microphone at each node, the emitted test signal will be received with high level at the same node. This fact produces a high-power self-interference signal that prevents the node from listening properly to the test signals coming from the rest of nodes, especially in reverberant conditions. To illustrate this problem, Fig.4(a) shows the contributions captured by the microphone of the first node in a network of several acoustic sensors emitting simultaneously (the one of Fig. 1). As representative examples, we show the contributions corresponding to the nodes $d = 2, 3$, so that it can be clearly observed that the power corresponding to the own test signal ($d = q = 1$) is greater than the one of the test signals arriving from other distant sensors. This power difference has also an effect in the resulting cross-correlation signals, as observed in Fig.4(b) (with an adjusted vertical scale to observe the relative rejection noise level). While the self-interference path can be clearly observed at the top, the impulse responses of the other sensors are barely above the noise level. This problem motivates the use of the self-interference reduction step discussed below.

A. Circular Cross-Correlation

Each node $q \in \{1, \ldots, D\}$ computes its circular cross-correlation between its recorded signal $x_q(n)$ and the known test signals corresponding to the rest of nodes:

$$\mathring{R}_{x_q s_d}(l) = \mathcal{F}^{-1}\left\{X_q(k) \cdot S_d^*(k)\right\}_{L_q}, \quad d = 1, \ldots, D. \tag{12}$$

where $X_q(k)$ and $S_d(k)$ are the L_q-point DFTs of $x_q(n)$ and (zero-padded) $s_d(n)$, being L_q the length of x_q. By taking the DFT of Eq.(5) and inserting it into the above equation, we get:

$$\begin{aligned} \mathring{R}_{x_q s_d}(l) &= \mathcal{F}^{-1}\left\{\left(\sum_{d'=1}^{D} S_{d'}(k)\tilde{H}_{q,d'}(k) + N_q(k)\right) S_d^*(k)\right\}_{L_q} \\ &= \sum_{d'=1}^{D} \mathcal{F}^{-1}\left\{S_{d'}(k)S_d^*(k)\tilde{H}_{q,d'}(k)\right\}_{L_q} + \\ &\quad \mathcal{F}^{-1}\left\{N_q(k)S_d^*(k)\right\}_{L_q}, \end{aligned} \tag{13}$$

where $\tilde{H}_{q,d'}(k)$ and $N_q(k)$ are the DFTs of $\tilde{h}_{q,d}(n)$ and $n_q(n)$, respectively. By extracting from the summation the target term corresponding to $d' = d$:

$$\begin{aligned} \mathring{R}_{x_q s_d}(l) &= \mathcal{F}^{-1}\left\{S_d(k)S_d^*(k)\tilde{H}_{q,d}(k)\right\}_{L_q} + \\ &\quad \sum_{d' \neq d} \mathcal{F}^{-1}\left\{S_{d'}(k)S_d^*(k)\tilde{H}_{q,d'}(k)\right\}_{L_q} + \\ &\quad \mathcal{F}^{-1}\left\{N_q(k)S_d^*(k)\right\}_{L_q} \\ &= \mathring{R}_{ss}(l) * \tilde{h}_{q,d}(l) + n_{q,I}(l) + n_{q,d}(l), \end{aligned} \tag{14}$$

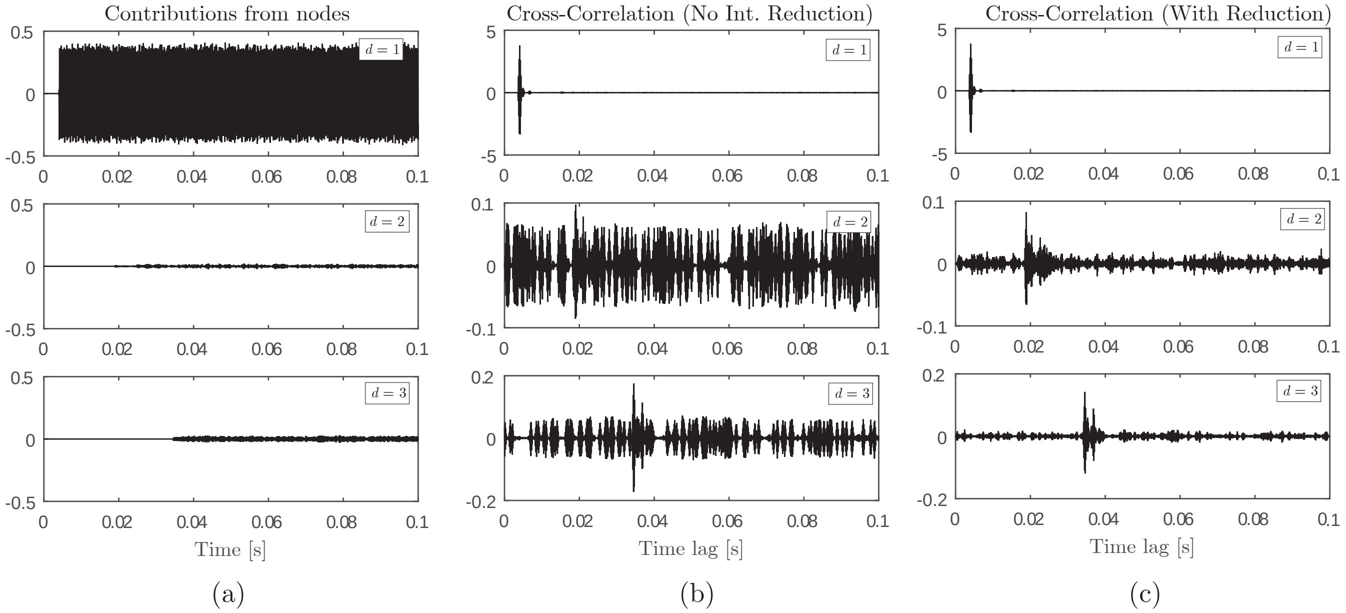


Fig. 4. Contributions received at the first node ($q = 1$) of the Fig.1 example, coming from the node itself ($d = 1$) and two more nodes ($d = 2, 3$), when emitting simultaneously. (a) Contributions from the three nodes captured by $x_1(n)$. (b) Cross-correlation $\mathring{R}_{x_1 s_d}(l)$ with known test signals without any processing. (c) Cross-correlation $\tilde{R}_{x_1 s_d}(l)$ after applying self-interference reduction.

where $n_{q,I}(l)$ is an interference rejection noise term and $n_{q,d}(l)$ is another noise term arising from the sensor background noise. This last term can be neglected, since $n_q(n)$ and $s_d(n)$ are uncorrelated. Taking into account Eq.(10), the above equation can then be expressed as a filtered impulse response with an interference rejection noise term:

$$\mathring{R}_{x_q s_d}(l) = P R_{zz}(l) * \tilde{h}_{q,d}(l) + n_{q,I}(l). \tag{15}$$

We can here define the signal-to-noise ratio (SNR) of the (q, d) sensor pair as:

$$\mathrm{SNR}_{q,d} = 10 \log_{10} \left(\frac{E_t \left\{ \left(P R_{zz}(l) * \tilde{h}_{q,d}(l) \right)^2 \right\}}{E_t \left\{ \left(n_{q,I}(l) \right)^2 \right\}} \right), \tag{16}$$

where $E_t\left\{\cdot\right\}$ denotes temporal averaging. The interference rejection term can be further decomposed into an autopath ($d' = q$) and cross-path interference term:

$$\begin{aligned} n_{q,I} &= \mathcal{F}^{-1}\left\{ S_q(k) S_d^*(k) \tilde{H}_{q,q}(k) \right\}_{L_q} + \\ &\quad \mathcal{F}^{-1}\left\{ \sum_{d' \neq q,d}^{D} S_{d'}(k) S_d^*(k) \tilde{H}_{q,d'}(k) \right\}_{L_q} \\ &= \mathring{R}_{s_q, s_d}(l) * \tilde{h}_{q,q}(l) + \sum_{d' \neq q,d} \mathring{R}_{s_{d'}, s_d}(l) * \tilde{h}_{q,d'}(l) \end{aligned} \tag{17}$$

Given the small distance between the microphone and the loudspeaker at each sensor, the autopath interference term clearly dominates over the cross-path term ($\alpha_{q,q} \gg \alpha_{q,d}\ \forall d \neq q$). Therefore, $\tilde{h}_{q,q}(n)$ must be estimated in order to remove the sensor's own contribution from the captured signal and increase the node pair SNR.

B. Estimation of Self-Interference Acoustic Path

It can be readily observed that the self-interference acoustic path $h_{q,q}(n)$ is contained within the circular cross-correlation of Eq.(15) when the own test signal is considered in its computation ($d = q$). In fact, when $d = q$, the first term of Eq.(15) will dominate over the interference term $n_{q,I}(n)$ leading to:

$$\mathring{R}_{x_q s_q}(l) \approx P \tilde{h}_{q,q}(n) * R_{zz}(n). \tag{18}$$

This can be clearly seen in Fig.4(b), where the rejection noise level for $d = q = 1$ is much lower than the one of the impulse response, i.e. $\mathrm{SNR}_{q,q} \gg \mathrm{SNR}_{q,d}\ \forall (q, d)$. Note, however, that the filtering effect caused by the chip signal must be removed in order to get an accurate estimate of the self-interference acoustic path. This filtering effect can be removed by using a properly designed inverse filter matrix. Due to the closeness between the microphone and loudspeaker at each sensor, the direct-to-reverberant energy ratio of the self-interference path is very high, so that it can be well assumed that $h_{q,q}(n)$ is a short-length response with duration N much shorter than L_q, i.e., $N \ll L_q$. In other words, due to the proximity between the loudspeaker and the microphone at one node, the impulse response corresponding to the acoustic path between them will concentrate most of its energy in the direct path peak of the response. Consequently, peaks corresponding to room reflections will be almost negligible. Even if $h_{q,q}(n)$ is not strictly zero for $n > N$, most of the energy will be concentrated in the direct path and the first reflections, which will be shown to be sufficient to reduce significantly the interference.

Let us rewrite Eq.(18) in matrix form as

$$\mathring{\mathbf{r}}_{x_q s_q} = \mathbf{R}_z \mathbf{h}_{q,q}, \tag{19}$$

where $\mathring{\mathbf{r}}_{x_q s_q} = [\mathring{R}_{x_q s_q}(0), \ldots, \mathring{R}_{x_q s_q}(N-1)]^{\mathrm{T}}$, $\mathbf{h}_{q,q} = [\tilde{h}_{q,q}(0), \ldots, \tilde{h}_{q,q}(N-1)]^{\mathrm{T}}$ and $\mathbf{R}_z$ is an $N \times N$ banded symmetric Toeplitz (non-causal) filter matrix, given by Eq.(20) (top of the next page).

The matrix $\mathbf{R}_z$ can be inverted [30], so that the estimated delayed impulse response is obtained as

$$\hat{\mathbf{h}}_{q,q} = \mathbf{R}_z^{-1}\mathring{\mathbf{r}}_{x_q s_q}, \tag{21}$$

with $\hat{\mathbf{h}}_{q,q} = [\hat{\tilde{h}}_{q,q}(0), \ldots, \hat{\tilde{h}}_{q,q}(N-1)]^{\mathrm{T}}$. Note that the inverted matrix only depends on $R_{zz}(n)$, thus, it can be computed offline and stored beforehand without adding any computational load to the system. Due to the small distance between the loudspeaker and the microphone the direct path will be dominant over the total response, thus, the node self-timestamp can be directly obtained as the location of the maximum:

$$\hat{\zeta}_{q,q} = \underset{n}{\operatorname{argmax}} \left\{ \hat{\tilde{h}}_{q,q}(n) \right\}. \tag{22}$$

C. Self-interference Cancellation

Once the inverse filter matrix has been applied to $\mathring{R}_{x_q,s_q}(l)$ to recover $\tilde{h}_{q,q}(n)$, the self-interference contribution can be highly attenuated from $x_q(n)$, resulting in

$$\tilde{x}_q(n) = x_q(n) - s_q(n) * \hat{\tilde{h}}_{q,q}(n), \tag{23}$$

where $\tilde{x}_q(n)$ is the self-interference-free processed signal at node q. When all the sensors remove their own contribution from their captured signal, the cross-correlation of the resulting $\tilde{x}_q(n)$ with the known test signals $s_d(n)$ will reveal the impulse response structure of the different acoustic paths between each pair of sensors with higher SNR. These are obtained as in Eq.(12), but considering the DFT of the processed signals $\tilde{x}_q(n)$. To avoid confusion, we denote these new cross-correlation signals as $\tilde{R}_{x_q s_d}(l)$, which are computed for all pairs (q, d) with $q \neq d$ (the cross-correlation for the autopath impulse response already has a high SNR, as discussed in the previous section).

Following the same example of Fig.4, it can be clearly observed in panel (c) that the structure of the impulse response between sensors $d = 2, 3$ and the first node $q = 1$ has emerged from the rejection noise after canceling the self-interference path. As it will be analyzed in Section VII, an average SNR gain of more than 20 dB is achieved. Note that the cross-correlation is the same for the case $d = 1$, since interference rejection is only applied for pairs where q and d are different.

V. TOA Estimation

The effect of self-interference reduction has already been discussed, but there is still the need to estimate the location of the direct path peaks from the resulting cross-correlation signals $\tilde{R}_{x_q s_d}(l)$. This is not an easy task, since rejection noise and the filtering effect of $R_{zz}(n)$ affects the detection of the direct path delay. While this could seem a serious issue for inverting the filtering effect, the objective now is only to detect the time delay corresponding to the TOA. In fact, detecting the TOA peak of the impulse response will be sufficient and the need to estimate the impulse responses as in the self-interference case is avoided. Note also that selecting the location of the cross-correlation maximum as in Eq.(22) is not possible, since the direct path peak is not always the peak with highest amplitude in a room impulse response. However, while not being the peak with highest amplitude, it will be one of the most prominent peaks, so that estimating the location of the most relevant peaks will help us determine the correct one. To this end, we propose a peak matching strategy that makes use of the known $R_{zz}(n)$ to extract the most important peaks from the node-to-node acoustic paths.

A. Peak Matching Algorithm

The next peak matching method has been proposed as an effective way to detect the TOA of the test signal from the computed cross-correlations. Although the TOA will be reflected in the cross-correlations as a prominent peak, selecting the right peak is not an easy task due to multiple effects: the true impulse responses have multiple peaks due to multipath; the true impulse responses have additional peaks due to the rejection noise arising from the correlation of PN sequences; all the above peaks are filtered due to the modulation of test signals, causing ripples that can be confused with the true response peaks.

The motivation behind the algorithm is to iteratively approximate the observed cross-correlation signals as a linear combination of delayed autocorrelation pulses R_{zz}. Thus, the algorithm is applied to the cross-correlation signal $\tilde{R}_{x_q s_d}(l)$ of each pair (q, d) with $q \neq d$, by following the next steps:

- *Step* 1: Normalize to unit norm $R_{zz}(n)$, obtaining $\bar{R}_{zz}(n) = R_{zz}(n)/\|R_{zz}(n)\|$. Initialize a residual signal $e_i(n) = \tilde{R}_{x_q s_d}(l)$ for the first iteration $i = 0$.
- *Step* 2: Find the maximum of the residual $\mathcal{E}_i = \max\{e_i(n)\}$ and its location $\upsilon_i = \operatorname{argmax}_n \{e_i(n)\}$. Take the inner product of $\bar{R}_{zz}(n)$ centered at the location of the maximum, i.e. $a_i = \langle \bar{R}_{zz}(n - \upsilon_i), e_i(n) \rangle$.
- *Step* 3: Update the residual by $e_{i^+} = e_i - a_i \cdot \bar{R}_{zz}(n - \upsilon_i)$.
- *Step* 4: Go to *Step* 2 and iterate until a maximum number of iterations has been reached or the value of $\mathcal{E}_i$ is below a given threshold. The smallest location index υ_i obtained from all the iterations is selected as the direct path peak, resulting in a stored timestamp

$$\hat{\zeta}_{q,d} = \min\{\upsilon_i\}. \tag{24}$$

The first step takes R_{zz} as a normalized basis where the residual signal will be projected throughout the different iterations, starting from the current cross-correlation signal. The second step performs the projection of the residual onto the basis, centered at the position of the current maximum (highest peak). Then, the third step eliminates from the residual the contribution of the basis to the current peak. The last step checks if a number of iterations has been reached or if the current maximum of the residual is sufficiently small, indicating that the current peak probably belongs to the rejection noise rather than to the impulse response. Note that the smallest location index υ_i corresponds to the first peak in the response above the estimated threshold, which will likely belong to the direct-path.

$$\mathbf{R}_z = \begin{pmatrix} R_{zz}(0) & R_{zz}(1) & \cdots & R_{zz}(N_{T_c}-1) & 0 & 0 & \cdots \\ R_{zz}(1) & R_{zz}(0) & \cdots & R_{zz}(N_{T_c}-2) & R_{zz}(N_{T_c}-1) & 0 & \cdots \\ R_{zz}(2) & R_{zz}(1) & \cdots & R_{zz}(N_{T_c}-3) & R_{zz}(N_{T_c}-2) & \ddots & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & R_{zz}(N_{T_c}-1) \\ R_{zz}(N_{T_c}-1) & R_{zz}(N_{T_c}-2) & \cdots & \cdots & \cdots & \ddots & R_{zz}(N_{T_c}-2) \\ 0 & R_{zz}(N_{T_c}-1) & \cdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & R_{zz}(1) \\ \vdots & \vdots & \ddots & R_{zz}(N_{T_c}-1) & \cdots & R_{zz}(1) & R_{zz}(0) \end{pmatrix}. \quad (20)$$

B. Stop Criterion

The first stop criterion suggested in *Step* 4 (maximum number of iterations) can be used to specify the number of peaks to be extracted from the cross-correlation signals, while the second one can be used to store all the peaks above a certain SNR. In fact, the SNR of a given cross-correlation signal can be approximated as

$$\widehat{\mathrm{SNR}}_{q,d} = 20\log_{10}\left(\frac{\max_{0\leq l<L_t}(\tilde{R}_{x_q s_d}(l))}{\max_{L_t-\nu\leq l<L_t}(\tilde{R}_{x_q s_d}(l))}\right), \quad (25)$$

where ν is a proper number of samples obtained from the tail of the cross-correlation signal assumed to pertain to the rejection noise. For example, in Fig.4(c), it can be clearly observed that the part of the cross-correlation signals going from 0.08 to 0.1 seconds is clearly dominated by the rejection noise. Samples in this range can be used to determine the rejection noise floor. Then, by using Eq.(25), the algorithm can be stopped when $\mathcal{E}_i < 10^{-\widehat{\mathrm{SNR}}_{q,d}/20}$

The advantage of using the threshold criterion is that the stopping rule is adapted to the considered pair of nodes as a result of the amount of rejection noise present in the signal. Following the example discussed throughout the paper, Fig.5 shows the values a_i and locations υ_i obtained from the cross-correlation signals of Fig.4(c). In both cases ($d = 2, 3$), five peaks have been extracted before reaching the SNR-based threshold. Note that the resulting peaks reveal the most prominent acoustic paths in the node-to-node responses.

VI. Ranging and Positioning

Once the timestamps corresponding to all the node pairs have been obtained, the system is ready to estimate ranges and node locations. The combination of two well-known techniques are used for this purpose, namely BeepBeep and multidimensional scaling (MDS). For the sake of completeness, we here briefly review their fundamentals.

A. BeepBeep

In the last years, the BeepBeep strategy has become a popular choice for estimating the range between two devices having acoustic emitters and receivers [13], [14]. The main advantage of the BeepBeep technique is that it does not require node synchronization. It is based on the simultaneous emission and recording of a specially designed sound signal ("Beep"). Each recording should contain two beeps, one coming from the speaker of the own node and one coming from the speaker of its peer node. To solve the ranging problem, each device counts the number of samples between these two beeps and exchanges the time duration information with its peer, deriving the two-way time of flight of the beeps with an accuracy limited by the sampling rate of the system. We use the BeepBeep method to extract the node-to-node ranges from the estimated TOAs. The use of modulated PN sequences as "beeps" allows to apply the BeepBeep algorithm in a network of multiple devices with all the nodes emitting simultaneously. This is achieved by exploiting the low cross-correlation properties of the sequences.

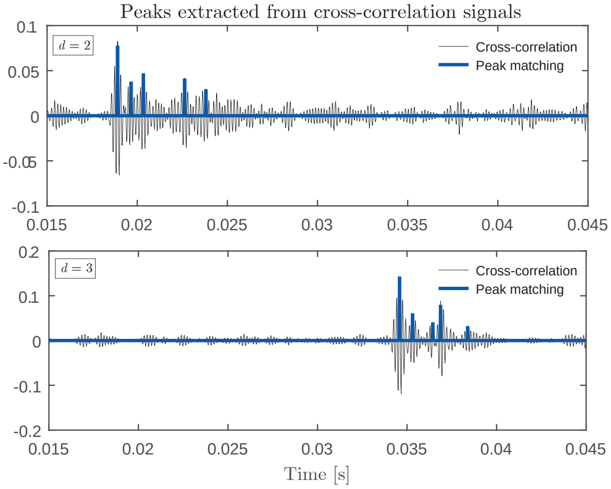


Fig. 5. Peaks obtained from the proposed iterative matching algorithm, corresponding to the cross correlation signals $d = 2, 3$ of Fig.4(c). Five peaks are extracted before reaching the SNR-based threshold.

By using the timestamps obtained from each pair of nodes (q, d) and assuming that all the nodes use the same sampling frequency, their distance can be straightforwardly estimated as [13]:

$$\hat{\Delta}_{q,d} = \frac{c}{2f_s}\left((\zeta_{q,d}-\zeta_{q,q})-(\zeta_{d,q}-\zeta_{d,d})\right)+r_{q,q}+r_{d,d}. \quad (26)$$

The $1/2$ factor in Eq.(26) comes from the fact that the range between devices q and d is approximated by $\frac{1}{2}(r_{q,d}+r_{d,q})$. Note that the addition of these two distances is approximately two times the distance between the center of the nodes and, therefore, must be divided by two. The key advantage of the

BeepBeep strategy is that the use of the self-timestamps $\zeta_{q,q}$ and $\zeta_{d,d}$ avoids the need for having synchronized nodes.

B. Multidimensional Scaling

After collecting the set of ranges corresponding to the different node pairs, MDS is applied for the final sensor positioning. The goal of MDS is to find a low dimensional representation of a group of objects (e.g. sensor positions), such that the distances between objects fit as well as possible a given set of measured pairwise "dissimilarities" that indicate how dissimilar objects are. In our sensor localization context, MDS is applied to find a map of node positions, where dissimilarities are range measurements. When the measured dissimilarities are equal to the true distances between sensors, classical MDS provides a closed-form solution by singular value decomposition of the centered squared dissimilarity matrix. On the other hand, when dissimilarities are measured in noise, other techniques should be used, usually based on iteratively minimizing a loss function between dissimilarities and distances [26]. In our work, we use Kruskal's raw-Stress [31], also referred to as the least-squares MDS model:

$$\sigma^2(\mathbf{P}) = \sum_{q=2}^{D} \sum_{d=1}^{q-1} w_{qd} \left(\hat{\Delta}_{q,d} - \mathrm{d}_{q,d}(\mathbf{P}) \right)^2 , \tag{27}$$

where $\mathbf{P} = [\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, \ldots, \hat{\mathbf{p}}_D]$ is the matrix of estimated sensor coordinates and $\mathrm{d}_{q,d}(\mathbf{P}) = \|\hat{\mathbf{p}}_q - \hat{\mathbf{p}}_d\|$. The weights $w_{q,d}$ are non-negative values that affect the contribution of the input dissimilarities (ranges) when computing and minimizing the stress. We use the computed SNRs as weights, so that $w_{q,d} = \hat{\mathrm{SNR}}_{q,d}$. The minimization of the stress function is performed by iterative gradient descent, as recommended by Kruskal [31].

Typical smartphone microphones and similar commercial devices exhibit considerable directional responses at frequencies above 1 kHz [32]. Owing to such limitations, line-of-sight losses may appear if the setup design does not consider this issue. The proposed SNR-based weighting in MDS helps to combat this effect, letting the optimization algorithm take into account which node pairs are more reliable. Modifications to the proposed peak-matching method could also be used to discard those peaks that are not due to direct-path propagation as in [21]. This aspect may be considered in a future work.

Finally, it is important to note that, because Euclidean distances do not change under rotation, translation and reflection, these operations may be freely applied to the MDS solution without affecting the raw-Stress. Procrustes analysis provides a way to transform one set of points to make it more comparable to another specified set [33], and this is the tool we use in our experiments to evaluate the performance of the proposed approach. In practice, the way to solve the rotation, translation and reflection ambiguities may be different depending on the application at hand. Geometric constraints can be introduced in the system, such as in [17]. For example, in order to remove the rotation and translation ambiguities, three nodes can be selected to lie on a plane, such that the first one is at $[0, 0, 0]^{\mathrm{T}}$, the second at $[\mathrm{x}_1, 0, 0]^{\mathrm{T}}$ and the third at $[\mathrm{x}_2, \mathrm{y}_2, 0]^{\mathrm{T}}$. The reflection ambiguity can be solved by specifying one more node to lie in the positive $\mathrm{z}-$ axis.

VII. Experiments

The performance of the proposed method was investigated through a series of room acoustics simulations with different degrees of interference and reverberation using the image-source method [34]. A shoe-box-shaped enclosure of dimensions 10 m×8 m×3 m was defined and different combinations of number of sensors ($D \in \{2, 4, 8\}$) and wall reflection factors ($\rho \in \{0.0, 0.5, 0.7, 0.9\}$) were considered, with average reverberation times $T_{60} \in \{0.00, 0.110, 0.225, 0.383\}$ seconds.

A total of $N_t = 500$ random sensor setups were simulated for each combination. All the experiments considered a sampling frequency of $f_s = 44100$ Hz in the nodes. The parameters affecting the signal design were the same as in the paper example ($T_c = 5 \cdot 10^{-4}$ s, $f_c = 6$ kHz), shown in Fig.3. The non-deterministic delays were simulated by adding random delay values uniformly distributed in the range [0, 0.015] seconds in the synthesized node-to-node impulse responses obtained by the image-source method. To minimize Eq.(27), we used Matlab's `mdscale` function.

The different stages of the proposed approach were evaluated as follows. First, we analyzed the SNR of different spreading codes (Gold codes and MLS of different length), with and without applying the proposed interference reduction approach. Second, we considered a medium-length MLS to evaluate the localization rate of the system and the proposed TOA estimation approach. Third, we evaluated how TOA estimation errors are propagated forward to ranging errors and to node positioning errors. Finally, the performance was compared to the ideal synchronized non-simultaneous case where node interference does not exist, taking it as an upper performance limit of our system.

In all the above cases, the performance was evaluated by computing the mean absolute error (MAE) between the real and estimated quantities, defined as:

$$\mathrm{MAE} = \frac{1}{N_t} \sum_{j=1}^{N_t} e_j , \tag{28}$$

where j is an index corresponding to the results from the jth simulated topology. The error e_j is defined as a function of the evaluated aspect:

$$e_j^{\mathrm{TOA}} = \frac{1}{N_p} \sum_{(q.d)} |\hat{\zeta}_{q,d} - \zeta_{q,d}|, \tag{29}$$

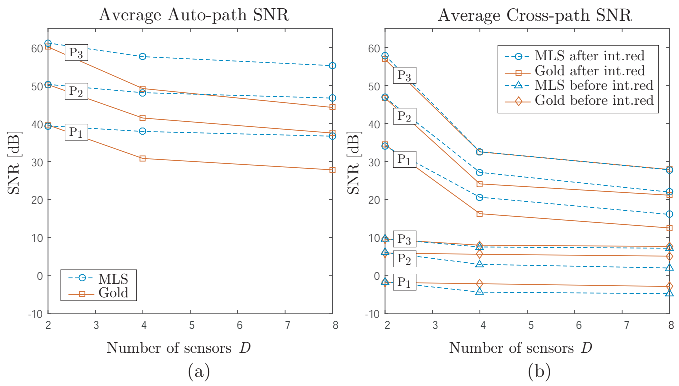
$$e_j^{\mathrm{RNG}} = \frac{1}{N_p} \sum_{(q.d)} |\hat{r}_{q,d} - r_{q,d}|, \tag{30}$$

$$e_j^{\mathrm{POS}} = \frac{1}{D} \sum_{d=1}^{D} \|\hat{\mathbf{p}}_d - \mathbf{p}_d\|, \tag{31}$$

where N_p is the total number of node pairs in the topology and superscripts TOA, RNG and POS denote TOA, ranging and positioning, respectively. We discuss the obtained results in the next subsections.

A. Spreading Codes and Interference Reduction

This experiment analyzes the selection of appropriate spreading codes used in the design of the node test signals.



(a)

(b)

Fig. 6. Average SNR found in the auto-path and cross-path correlation signals for different number of sensors and different code orders.(a) Autopath SNR. (b) Cross-path SNR.

We consider different families of PN sequences with different lenghts, namely Gold codes and MLS, with lenghts $P_1 = 2047$, $P_2 = 8191$ and $P_3 = 32767$. Fig.6 shows the average SNR obtained both from the auto-path (a) and cross-path (b) correlation signals, computed as in Eq.(16). In panel (b) we also present the results after applying the interference reduction approach, which show a significant SNR gain in all cases. By looking at the auto-path results in (a), it can be clearly observed that MLS outperform Gold codes independently of the number of sensors and the sequence length. This is due to the better autocorrelation properties of MLS with respect to Gold codes. In fact, the increased SNR in the auto-path case results in better interference reduction in (b), since the auto-path impulse response is estimated with higher accuracy and can be more easily removed from the node signal in the interference reduction stage. This is the reason why, although Gold codes perform better before interference reduction (they are known to have better cross-correlation properties), MLS outperform Gold codes when interference reduction is applied. As a result, the next subsections only consider MLS of length $P_2 = 8191$.

B. Localization Rate and TOA Estimation Accuracy

Localization rate describes the capability of the system to perform successfully the node localization task. It is defined as the ratio between successful trials and the total number of trials. A trial was considered to be successful when the average TOA error among all the node pairs was below 15 samples (≈ 0.3 ms). When this is not the case, the trial was considered to be unsuccessful because the TOA estimation error leads to average location errors that are above 10 cm (as shown in the next subsection). Note that no restrictions were considered in the simulated topologies, so that some unsuccessful trials might be due to non-advisable sensor placements (such as too closely positioned nodes). Fig.7(a) and (b) shows the localization rate and TOA MAE for the selected MLS-based signal as a function of the number of sensors and the wall reflection factor/reverberation time. As expected, the localization rate degrades significantly with reverberation and with the number of sensors, although it is above 80% even in the worst reverberant case with 8 simultaneous sensors. Similarly, the TOA error increases with reverberation, especially for $\rho = 0.9$. Nevertheless, TOA errors for 8 sensors under moderate reverberation are below 2 samples, showing the validity of the proposed peak matching algorithm.

C. Ranging and Positioning Accuracy

Obviously, TOA errors are propagated forward to ranging and positioning errors. Fig.7(c) shows the MAE of the estimated ranges after the BeepBeep approach. The error behavior is quite similar to the TOA error, showing a robust performance for moderate reverberation even when there are 8 emitting nodes, with a ranging error below 2 cm. Similarly, Fig.7(d) shows the MAE of the estimated node locations after MDS. The Procrustes method was used to align the estimated locations from MDS to the ground truth locations, filtering out translation and rotation effects [33]. Note how positioning errors are very similar to ranging errors, confirming that the accuracy of the final positioning is very dependent on the TOA estimation stage.

D. Comparison

This subsection compares the performance of the proposed system with respect to the ideal case where the nodes emit their test signals one at a time. Consequently, the rest of nodes capture all the test signals without any interference from other nodes, avoiding also their own self-interference. Moreover, they are assumed to be synchronized as in a wired system, so that they all know accurately the emitting and recording instants under a common time line. As a result, the errors in the system are only given by the TOA estimation stage, i.e. they still need to detect the arrival time for each emitted test signal. Table I compares our proposed approach with 4 unsynchronized simultaneous sensors with respect to this ideal case. The results are only slightly worse for our proposed approach, being comparable in accuracy with the evaluated non-simultaneous case. Note that our simultaneous positioning framework provides a calibration time reduction by a factor $D(D-1)$. Thus, for the case $D = 8$, the calibration time is reduced 56 times with respect to a pair-wise calibration (from 3.82 minutes to 4 seconds). While our simultaneous calibration framework requires additional complexity with respect to a non-simultaneous framework, the additional processing time is negligible with respect to the total gain in calibration time, specially when the number of sensors is considerably high.

E. Real Deployment

The applicability of the proposed approach was evaluated in a real heterogeneous WASN comprised of 3 smartphones (Galaxy S4, Motorola G and Sony Xperia S) and one laptop (Asus Zenbook), the last one being the central sink node. The devices were placed on a table inside a medium-size meeting room (6 m × 6 m), adjusting their orientation and assuring a line-of-sight condition among all of them. The reverberation time of the room was $T_{60} = 0.29$ s. The experiment was carried out by developing an ad hoc Android application for audio playback and recording ($f_s = 44.1$ kHz) that listens to a

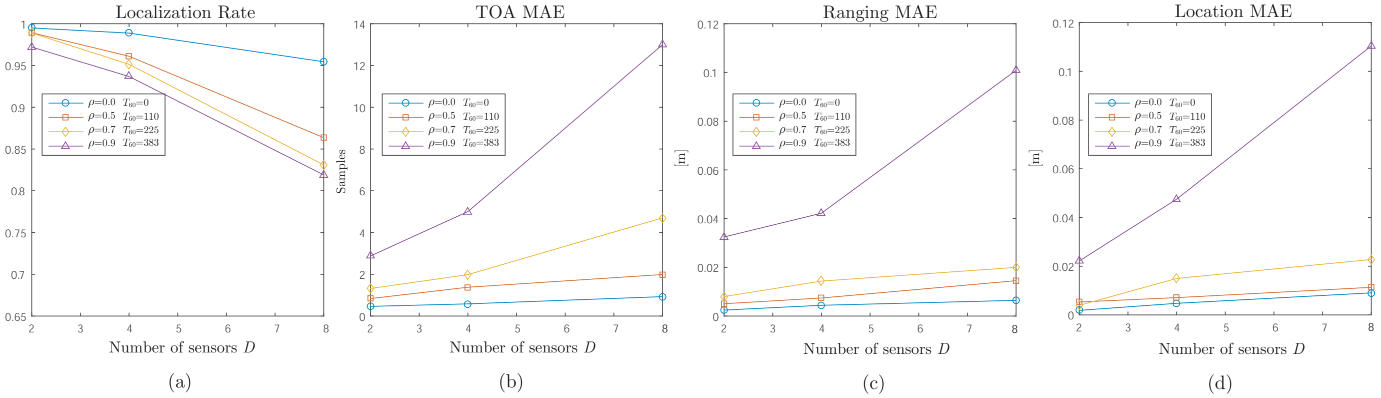


Fig. 7. Simulation results for MLS and $P_2 = 8191$ as a function of the wall reflection factor/reverberation time (ρ and T_{60} [ms]). (a) Average localization rate. (b) TOA Mean Absolute Error. (c) Ranging Mean Absolute Error. (d) Node Location Mean Absolute Error.

TABLE I
COMPARISON WITH IDEAL NON-SIMULATENAOUS SYNCHRONIZED NETWORK

	Loc. Rate				TOA MAE [samples]				Range MAE [cm]				Loc. MAE [cm]			
ρ	0.0	0.5	0.7	0.9	0.0	0.5	0.7	0.9	0.0	0.5	0.7	0.9	0.0	0.5	0.7	0.9
T_{60} [ms]	0	110	225	383	0	110	225	383	0	110	225	383	0	110	225	383
Sync. Non-simultaneous	1.00	1.00	1.00	0.98	0.82	1.32	2.04	3.31	0.35	0.81	1.43	2.55	0.17	0.40	0.91	1.25
Proposed ($D = 4$)	0.99	0.96	0.95	0.94	0.76	1.39	1.99	4.98	0.48	0.96	1.51	4.16	0.52	0.83	1.47	4.64

broadcast calibration command sent through the network. The average network delay (802.11n WiFi) when transmitting the calibration command was 15.444 ms, with a standard deviation of 2.219 ms. Regarding the initialization of the audio device, the average delay was 12.715 ms, with a standard deviation of 1.275 ms. In order to assure that all the devices start to record before any of them emits a test signal, we used a guard interval of 50 ms before emissions. In our setup, the smartphones sent their recordings to the laptop, which was used to process all the collected signals and its own one in Matlab. The experiment was repeated for 10 different device placements, using the same test signals as the ones used in the simulated experiments ($P = 8191$, $T_c = 0.5$ ms and $f_c = 6$ kHz).

The results were very similar to the ones of the simulations: the TOA MAE was 1.1 sample, leading to a ranging MAE of 0.86 cm and a location MAE of 0.81 cm. Note that although the devices were placed on a table and they can be assumed to lie on a plane, the error was computed by taking into account all three spatial coordinates $(\mathrm{x}, \mathrm{y}, \mathrm{z})$.

Additionally, the 10 tested set-ups were also numerically simulated by tuning the wall reflection factors with the aim of approximating the measured reverberation time. The average results were slightly worse than in the real case, obtaining a TOA MAE of 1.8 samples, a ranging MAE of 1.27 cm and a location MAE of 1.10 cm.

VIII. Conclusion

A complete framework for node ranging and positioning in unsynchronized WASNs has been presented. Its main advantage over other state-of-the-art approaches is that ranges and node locations are obtained by following a simultaneous playback/recording process, significantly reducing the total calibration time. PN sequences with good autocorrelation and cross-correlation properties are used in the design of the node test signals with this purpose. Important aspects affecting the performance of the method in practical situations, such as the adaptation of the node signals to the node transducers, the reduction of the node self-interference and the selection of appropriate PN sequences, have been considered and discussed. Moreover, a novel peak matching approach for TOA estimation has been proposed, which allows to use BeepBeep ranging and multidimensional scaling for estimating the final node locations. Experiments in both simulated and real environments have been conducted considering different acoustical conditions, showing that our proposed approach is a valid alternative for high-accuracy node positioning.

References

[1] A. Bertrand, S. Doclo, S. Gannot, N. Ono, and T. Van Waterschoot, "Special issue on wireless acoustic sensor networks and ad hoc microphone arrays," *Signal Process.*, vol. 107, pp. 1–3, February 2015.

[2] M. Parviainen, P. Pertila, and M. Hamalainen, "Self-localization of wireless acoustic sensors in meeting rooms," in *Proc. 4th Joint Workshop on Hands-free Speech Commun. and Microphone Arrays (HSCMA)*, 2014, pp. 152–156.

[3] M. Cobos, J. Perez-Solano, S. Felici-Castell, J. Segura, and J. Navarro, "Cumulative-sum-based localization of sound events in low-cost wireless acoustic sensor networks," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 22, no. 12, pp. 1792–1802, 2014.

[4] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 21, no. 2, pp. 343–356, 2013.

[5] J. Segura-Garcia, S. Felici-Castell, J. Perez-Solano, M. Cobos, and J. Navarro, "Low-cost alternatives for urban noise nuisance monitoring using wireless sensor networks," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 836–844, 2015.

[6] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Commun. Surveys Tutorials*, vol. 11, no. 1, pp. 13–32, First 2009.

[7] H. Liu, H. Darabi, P. Banerjee, and J. Liu, “Survey of wireless indoor positioning techniques and systems,” *IEEE Trans. Syst., Man, Cybern*, vol. 37, no. 6, pp. 1067–1080, 2007.

[8] J. Costa, N. Patwari, and A. Hero III, “Distributed weighted-multidimensional scaling for node localization in sensor networks,” *ACM Trans. on Sensor Networks*, vol. 2, no. 1, pp. 39–64, 2006.

[9] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, and N. Correal, “Locating the nodes: cooperative localization in wireless sensor networks,” *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, July 2005.

[10] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The cricket location-support system,” in *Proc. of the 6th Annual Int. Conf. on Mob. Comp. And Net. (MOBICOM)*, 2000, pp. 32–43.

[11] C. Sertatl, M. A. Altnkaya, and K. Raoof, “A novel acoustic indoor localization system employing CDMA,” *Digit. Signal Process.*, vol. 22, no. 3, pp. 506 – 517, 2012.

[12] J. Prieto, A. Jimenez, J. Guevara, J. Ealo, F. Seco, J. Roa, and F. Ramos, “Performance evaluation of 3D-LOCUS advanced acoustic LPS,” *IEEE Trans. Instrum. Meas*, vol. 58, no. 8, pp. 2385–2395, 2009.

[13] C. Peng, G. Shen, and Y. Zhang, “Beepbeep: A high-accuracy acoustic-based system for ranging and localization using COTS devices,” in *ACM Trans. on Embed. Comp. Syst.*, vol. 11, no. 1, 2012, pp. 4:1–4:29.

[14] C. Tan, X. Zhu, Y. Su, Y. Wang, Z. Wu, and D. Gu, “A low-cost centimeter-level acoustic localization system without time synchronization,” *Measurement*, vol. 78, pp. 73 – 82, 2016.

[15] J. Sachar, H. Silverman, and I. Patterson, W.R., “Position calibration of large-aperture microphone arrays,” in *IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP)*, vol. 2, 2002, pp. 1797–1800.

[16] M. Hennecke and G. Fink, “Towards acoustic self-localization of ad hoc smartphone arrays,” in *Proc. of Joint Workshop on Hands-free Speech Commun. and Microphone Arrays (HSCMA)*, 2011, pp. 127–132.

[17] V. Raykar, I. Kozintsev, and R. Lienhart, “Position calibration of microphones and loudspeakers in distributed computing platforms,” *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 13, no. 1, pp. 70–83, 2005.

[18] D. Haddad, L. Nunes, W. Martins, L. Biscainho, and B. Lee, “Closed-form solutions for robust acoustic sensor localization,” in *Proc. of the IEEE Int. Workshop on App. of Signal Process. to Audio and Acoust. (WASPAA)*, 2013, pp. 1–4.

[19] M. Crocco, A. Del Bue, and V. Murino, “A bilinear approach to the position self-calibration of multiple sensors,” *IEEE Trans. on Signal Process.*, vol. 60, no. 2, pp. 660–673, 2012.

[20] R. Moses, D. Krishnamurthy, and R. Patterson, “A self-localization method for wireless sensor networks,” *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 4, pp. 348–358, 2003.

[21] D. B. Haddad, W. A. Martins, M. V. M. da Costa, L. W. P. Biscainho, L. O. Nunes, and B. Lee, “Robust acoustic self-localization of mobile devices,” *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, 2016.

[22] L. Wang, T. Hon, J. D. Reiss, and A. Cavallaro, “Self-localization of ad-hoc arrays using time difference of arrivals,” *IEEE Trans. on Signal Process.*, vol. 64, no. 4, pp. 1018–1033, 2016.

[23] J. Proakis and M. Salehi, *Digital Communications, 5th Edition*, 5th ed. McGraw-Hill, 2007.

[24] L. Girod and D. Estrin, “Robust range estimation using acoustic and multimodal sensing,” in *IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, vol. 3, 2001, pp. 1312–1320.

[25] L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin, “Locating tiny sensors in time and space: a case study,” in *Proc. of IEEE Int. Conf. on Comp. Design: VLSI in Comp. and Process.*, 2002, pp. 214–219.

[26] I. Borg and P. Groenen, *Modern multidimensional scaling: theory and applications (2nd ed)*. New York: Springer-Verlag, 2005.

[27] M. Guggenberger, M. Lux, and L. Böszörmenyi, “An analysis of time drift in hand-held recording devices,” in *Proc. of 21st Int. Conf. MultiMedia Model. (MMM 2015)*, Sydney, 2015, pp. 203–213.

[28] D. Kedia, “Comparative analysis of peak correlation characteristics of non-orthogonal spreading codes for wireless systems,” *Int. Journal of Dist. and Parall. Syst.*, vol. 3, no. 3, p. 63, 2012.

[29] R. Gold, “Optimal binary sequences for spread spectrum multiplexing (corresp.),” *IEEE Trans. on Inf. Theory*, vol. 13, no. 4, pp. 619–621, 1967.

[30] D. A. Lavis and B. W. Southern, “The inverse of a symmetric banded toeplitz matrix,” *Rep. on Math. Phys.*, vol. 39, no. 1, pp. 137–146, 1997.

[31] J. B. Kruskal, “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis,” *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[32] N. D. Gaubitch, J. Martinez, W. B. Kleijin, and R. Heusdens, “On near-field beamforming with smartphone-based ad-hoc microphone arrays,” in *Proc. of the 14th Int. Workshop on Acoust. Signal Enhancement (IWAENC 2014)*, Juan les Pins, 2014.

[33] D. Kendall, “A survey of the statistical theory of shape,” *Statistical Science*, vol. 4, no. 2, pp. 87–89, 1989.

[34] J. B. Allen and B. D. A., “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Amer.*, vol. 65, no. 3, pp. 943–950, 1979.

Maximo Cobos received a Ph.D. in telecommunications engineering from the Universitat Politècnica de València (UPV) in 2009. He was the recipient of the Ericsson Best Ph.D. Thesis Award from the Spanish National Telecommunications Engineering Association (COIT). In 2010, he received a “Campus de Excelencia” post-doctoral fellowship to work at the iTEAM Institute of the UPV and T-Labs Berlin. Since 2011, he has been an Assistant Professor with the Computer Science Department of the Universitat de València. His research interests include signal processing and machine learning algorithms for audio and wireless sensor network applications, where he has published more than 80 technical papers. Dr. Cobos is a senior member of the IEEE.

Juan J. Perez-Solano received the M.Sc. in physics and the Ph.D. degree in electrical engineering from the Universitat de València in 1994 and 2002, respectively. In 1996, he joined the Computer Science Department of the Universitat de València, where he is currently an Associate Professor. His research interests include multiresolution techniques and wavelet transform applications, spread spectrum communications, wireless data transmission and wireless sensor networks.

Óscar Belmonte received the M.Sc. and the Ph.D. degrees in physics from the Universitat de València in 1992 and 2002, respectively. He is currently an Associate Professor with the Computer Languages and Systems Department at Universitat Jaume I, Spain. His research areas include wireless sensor networks, multiresolution modeling, real-time visualization and Java technology.

German Ramos received the M.Sc. and the Ph.D degrees in telecommunications engineering in 1997 and 2006, both from the Universitat Politècnica de València (UPV), Spain. He is an Associate Professor with the Electronics Department of the UPV. Over the last 15 years he has developed the software and hardware of several digital audio processors for professional audio companies around the world, covering digital equalizers and cross-overs. His research interests are in the areas of digital filter design and implementation. Dr. Ramos is a member of the AES and the IEEE

Ana M. Torres received the M.Sc. and the Ph.D degrees in telecommunications engineering from Universitat Politècnica de València in 2003 and 2012, respectively. Currently she is an Assistant Professor with the IEEAC Department at the University of Castilla-La Mancha, Spain. Her work is focussed on the area of signal processing for audio and communications. Her research interests include algorithms for room acoustics analysis, microphone arrays and spatial audio.