Indoor localization with audible sound — Towards practical implementation

J.N. Moutinho [a,b,∗], R.E. Araújo [a,b], D. Freitas [b]

[a] *INESC TEC, 4200-465 Porto, Portugal*

[b] *Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal*

ARTICLE INFO

ABSTRACT

This paper presents an innovative evaluation and comparison of several methods and techniques necessary to implement an indoor localization system based on audible sound. Experiments were conducted in a room with conditions very close to possible practical application demonstrating that time delay estimation using generalized cross-correlation phase transform provides the best estimate to the distance to fixed anchors, and highlight the benefits of a new localization method entitled circle shrinking based on an optimization methodology. Of the three optimization methods tested, Gauss–Newton proves to be the most adequate, and among the three medium access methods evaluated, code division multiple access acoustic transmission provided the best results. A localization system combining these components and using only off-the-shelf hardware reached an average accuracy of 1.3 cm in the central area of the test room with an excitation signal-to-noise ratio as low as 7.2 dB, an almost unperceivable noise like audio signal. These results represent an advance of the state-of-the-art in indoor localization systems, pointing towards the possibility of widespread practical implementation with everyday use components.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

With the recent technological advances in mobile devices, there has been a strong growth in context-aware applications that passively or actively determine their locations [1]. Using appropriate algorithms these intelligent systems estimate the location and provide to the users some kind of information or service regarding specific geographically adapted content. Global Navigation Satellite Systems (GNSS) is nowadays seen as a standard service and assures this task in outdoor environments where there are satellites in line of sight. In contrast, indoor environments do not have a consensual technology established to achieve indoor localization with similar simplicity of use or implementation. All the available methods require at least one aspect that is not easily achievable in most practical situations. Some are still in a concept development phase, others are simply too expensive to implement. A rational approach to indoor localization would be one that uses information available everywhere without the need for additional specific setups. The presented approach uses public address (PA) loudspeakers as signal emitters that will allow a mobile device to localize itself. The typically present PA sound systems can be leveraged to provide additional functionality to allow users to determine their location through sound. Small modifications can be performed centrally minimizing large scale implementation costs, in essence, achieving a localization system like the established global positioning system (GPS), but this time, for indoor environments using acoustic signals. In a GNSS, the infrastructure is available in the sky worldwide. Conversely, in an indoor environment there is no suitable global signal available to receivers. Therefore, indoor spaces require specific signals, using special techniques, requiring

---

∗ Corresponding author at: INESC TEC, 4200-465 Porto, Portugal. Tel.: +351 225081600; fax: +351 225081440.
  E-mail address: jnm@fe.up.pt (J.N. Moutinho).

custom hardware suitable to the indoor environment, based on the desired requirements of cost, precision, robustness or granularity. There is no "standard" RF signal like in the GPS example. Each type of environment or application has its own approach. In industrial environments, for example, buried cable systems emit RF signals, use laser beams or optical-based floor markers. For domestic, administrative, commercial or public environments proposed solutions rely, for instance, on radio technologies like Wi-Fi, optical signals, ultrasound waves or simply artificial vision. Up to now, no method, technology or technique has demonstrated the possibility to become a "standard" in indoor localization. Some are too costly, others are not robust enough and others require add-ons, tags or devices that will not facilitate an easy large scale dissemination.

The work presented here aims to reach a novel indoor localization system suitable for most situations and approaching the ease of use of GPS. To accomplish that, it employs audible sound (almost unperceivable) as the localization enabling signal considering that most public spaces already possess a PA sound system infrastructure. Smartphones can act as the localization enabling device offering acquisition hardware to capture audio without any patch, tag or other accessories. With both of these nearly ubiquitous components, an indoor localization system may be assembled without having any additional major cost or introduction of new equipment in the users' daily lives. Such a universal and user convenient approach, that considers these aspects simultaneously, has not been found in the literature so far.

For this type of application, it is of great importance to consider a simultaneous multi-user scenario. A user "timeslot" multiplexing approach for the localization is not desirable as it imposes smaller position refresh rates and may cause a waiting queue among users. Complementarily, when considering this signal type and the channel, one can understand that each audio contribution of a loudspeaker will be mixed with all the others in the receiver's point of view. There is no reserved single-access channel where communication would be much simpler. Thus, one of the key aspects when considering such a system is to evaluate which channel access method is more suitable for this audible sound approach in a shared medium. Three cornerstone communication methods used typically in radio frequency communication systems were evaluated: Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA) and Code Division Multiple Access (CDMA). An innovative comparison is performed among several methods required for achieving localization through the use of sound in the audible range. Audio signals are mechanical waves generated from vibrations within a medium. These signals travel through the channel very differently from the way electromagnetic waves do. In fact these types of signals do not interact much with people or objects (furniture, boundaries, etc.). Therefore, the known results of radio frequency approaches cannot be generalized without an additional study like the one provided in this paper. In an indoor scenario, this difference is reinforced due to the presence of extra physical boundaries (wall, floor and ceiling, for example). Channel effects like multipath, fading, interference are phenomena that behave quite differently while using audio signals. It is also very important to consider the drastically smaller propagation speed of sound, almost a million times slower. All these differences justify the need to evaluate methodologies without assuming that previously obtained results with other types of signals are valid for audible sound.

To achieve a functional setup valid for a real test bed comparison, the most relevant correlation methods to perform time delay estimation (TDE) are also evaluated. Localization estimation is discussed considering error minimization approaches based on classical least square estimation methods like Gauss–Newton, Newton–Raphson and steepest descent. Even a slight difference in performance may be important due to its frequent repetitive use in a localization system. This particular study can be of great interest depending on the application and in the conditions where processing capabilities are limited or localization estimation update rate needs to be higher.

The evaluation of all these requirements is necessary to study the possibility of using audible sound as the type of signal that will allow achieving localization. The experimental setup described in the paper emulates a possible real usage situation in almost all aspects. A 60 square meter room with 4 public address loudspeakers, people working, furniture, plaster walls, glass windows and environmental noise, is similar in many aspects to most possible applications where a pervasive indoor localization technology would be useful. Possible applications like museums, malls or public transportation hubs, are examples on where public address loudspeakers are typically present and would require minimum adaptation to such technology, lowering costs and improving the possibility of implementation. Larger areas like industrial spaces, where limited sound coverage and loud noise conditions could limit its usage, may have their own adjusted localization technology based on their specific requirements. Experimental tests were conducted statically to allow comparison between methods. Similar results are expected when moving at walking speeds as the Doppler Effect is not relevant because frequency changes in the received signals do not greatly affect the methodologies here employed regarding TDE, the most important component in localization based on time-of-flight.

This paper continues as follows: Section 2 presents the related work providing a general view on the state-of-the-art of indoor localization and explains the place and motivation of this novel approach; Section 3 introduces Indoor Audible Sound Localization, covering some of its most important issues and providing some possible answers; Section 4 describes the conducted experiments concerning the evaluation of all the components necessary to accomplish this type of localization. It also contains the results and comparisons between the choices to converge on a feasible audible sound localization system. Finally, Section 5 presents the conclusions and some future work on the subject.

2. Related work

Historically, radar and underwater sonar were the first modern localization applications. In the former, the distance from a faraway object at a certain angle is determined by the travel time necessary for an electromagnetic wave to go from

Table 1
Comparison of indoor localization technologies.

	Accuracy (m)	Coverage (m)	Cost	Reliability	Practical use
Computer vision	10^{-3}–10^{-1}	1–10	High	Low	Medium
Infrared	10^{-2}–1	1–5	Medium/High	Low	Low
Ultrasound	10^{-2}	2–10	Medium	High	Low
Wi-Fi	1–10	20–50	Low/Medium	Low	High
RFID	10^{-1}–1	1–10	Low/Medium	High	Low
Bluetooth	1–10	1–30	Low/Medium	High	Low
Inertial	10^{-1}–1	1–10	Low	Low	High
Audible sound	10^{-2}	1–20	Low	High	High

the source to the target object and back, usually in the air. The second uses a sound wave and tries to achieve the same objective underwater. Many localization technologies were derived from these approaches in military application. Many developments have revolutionized the way ranges and consequently position has been determined. The well-known GPS is just an example on how military applications have, in time, become useful and available to civilians. However, there was no strong military interest in indoor localization until recent history [2] and indoor localization remained underdeveloped until some years ago.

Nowadays, many point out indoor localization as one of the next big things in pervasive and mobile computing [3]. Indoors is where people stay long periods, meet friends, and where business happens. Knowing locations indoors will surely be at least as popular as outdoor localization and all the major technological companies are researching on it [4].

The Indoor Positioning Systems (IPS) may be distinguished by the signal used to determine localization. The existing approaches have explored almost every type of signal and sensor: image (Artificial Vision), Infrared, Ultrasound, Radio Frequency, Inertial and Magnetic, and also Audible Sound [5]. However, most of these approaches require specific custom hardware, have costly installations and may have some reliability issues [6]. Table 1 presents a comparison between some of the technologies regarding accuracy, coverage, cost, reliability and practical use.

Many of the approaches found in the literature are still conceptual and do not account for the hard requirements and challenges of implementation in realistic conditions. It is difficult to imagine using infrared signal-based localization systems in people's daily life, as it requires line-of-sight (LOS) between transmitter and receiver without interference from strong light sources. Or it may be difficult to convince people to start carrying devices beyond what they already have.

Audible sound is lately being studied as an emergent approach due to its inherent characteristics. It shares the ultrasound characteristics but is even less expensive and more appropriate to usage in everyday life due to existence of off-the-shelf devices with wider coverage, like loudspeakers and cell phones, which may be used with relatively minor modifications. Audible Sound clearly was not intensely approached before due to the assumption that it would disturb the acoustic environment in an undesirable way. Ultrasound systems, for instance, use the human upper hearing limitation of 20 kHz to conceal the necessary signal. However, the use of ultrasound signals requires custom made hardware [7–11]. Some projects cleverly explore the upper frequency hardware limits of off-the-shelf devices to achieve low ultrasound frequencies while a few do that in the upper part of the hearing frequency band (18–20 kHz). The former approaches are expensive and therefore not applicable in a large scale, because they require a large number of ultrasound transducers to provide uniform area coverage. They also rely on the user carrying a custom made device capable of receiving ultrasound. The latter are more promising and fulfill most of the requirements to succeed, but still depend on having a large number of sensors (loudspeakers or microphones) to achieve sufficient area coverage and good enough performance, due to the high frequency of sound and consequent high directivity and attenuation with distance [12–14]. To minimize this problem, one may use the regular frequency range of common portable devices by using wide band audible sound. The mobile device (to be localized) will receive the signal more easily as it will be in its characteristic frequency response. Using lower frequencies, lower loudspeaker directivity will also allow better area coverage and will have smaller impact on the problem of physical obstructions between the transmitters and the receivers (nonexistent LOS).

In the literature, there are not many proposals for sound based approaches for indoor localization systems. One that deserves mention is a 3D indoor positioning system named Beep that was designed as a cheap positioning solution using audible sound technology [15]. Beep uses a standard 3D multilateration algorithm based on time of arrival (TOA) measured by the Beep's system sensors while a Personal Digital Assistant or another device emits sound signals. However, Beep requires custom hardware and specially designed emitters (beacons) carefully placed across an indoor space. This is a major drawback considering large scale implementation. Very recently, an acoustic indoor localization experiment employing Code Division Multiple Access (CDMA) was developed [16]. Binary-phase-shift-keying modulated Gold code sequences using direct-sequence (DS) spread spectrum (SS) technique are used to measure TOA. Their results confirm the validity of using audible sound and off-the-shelf components (ordinary loudspeakers and microphones) to achieve accurate and reliable indoor localization. However, their experiments were not focused in fulfilling the necessary requirements for use as a real indoor localization system, and their work stops short of considering problems related with implementation in realistic operating conditions. In [17] environmental sounds are used with very promising results and off-the-shelf components. Even though some of the experimental results found in [17] are interesting, usually a network of microphones is not present
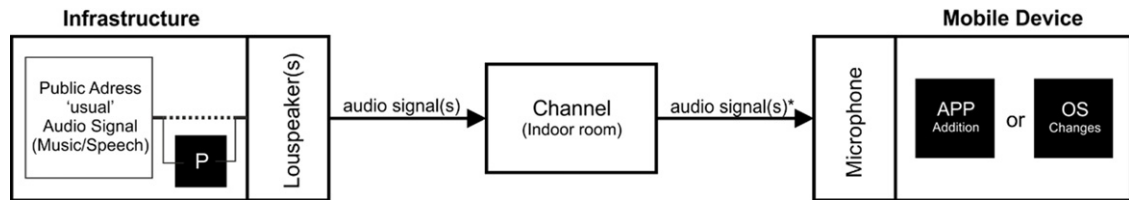
Fig. 1. Illustration of the two main components that support the presented approach: infrastructure and mobile device. The black-filled boxes represent the necessary processing elements to transform the "usual" audio signal being emitted and the software changes in the mobile device on the left and right, respectively.

in indoor spaces, and therefore its practical use is questionable. Nevertheless, the validity of some principles used in position determination and signal coding are useful and are taken as references in some of the subjects under analysis here.

The major apparent drawback of using audible sound is the effect that it may have on people. The proposed solution in this paper is to use psycho-acoustic techniques [18–20] to produce signals that are, or may be, unperceivable by the users. Given the possibility to avoid people's perception of this audio signal, one can take advantage of all its previously mentioned benefits.

In the system architecture illustrated in Fig. 1, it is possible to observe that this approach has great practical use as it requires few and inexpensive additions to the existing sound system and mobile device. The black-filled boxes in the infrastructure and in the mobile device represent the new necessary parts. On one hand, a processing unit to add the necessary psycho-acoustically hidden audio signals, and on the other hand a piece of software on a pre-existent mobile device (possibly a smartphone).

If a public address sound system is properly designed, it will have good sound coverage in all areas and no costly change will have to take place in the physical setup of the PA sound system. Some public spaces are already legally bound (e.g., EN 54) to have a "properly designed" (EN-60849) public address sound system to be used in emergency situations. Others already use it as a tool for public information (transport hubs, commercial spaces, etc.) and as a way to set a pleasant acoustic environment with music or nature sounds. This leads to the assumption that most indoors public spaces have a capable public address sound system. It is also expected that people will have a smartphone. These two realistic assumptions are in the basis of the practicality of the proposed approach and when combined constitute one of the major differences from the previously discussed possibilities.

3. Indoor audible sound localization

Considering the previously mentioned references regarding localization systems, it may be concluded that to address the concrete issues of indoor localization using sound signals, three key aspects should be taken into account: to emit the signal into the shared channel medium, to determine distances with that signal and to calculate position using some of the distances (with a minimum of three for 2D localization). The choice on how to emit into the shared channel is usually taken between a time division multiple access (TDMA), a frequency division multiple access (FDMA) or a code division multiple access (CDMA) scheme. No previous work was found performing a comparison between these methods or justifying the use of one of them in this application. A comparison is presented in this paper. Additionally the other requirements necessary to implement such a system are evaluated, like the use of time delay estimation techniques based on Cross Correlation Methods and position estimation using error minimization approaches based on nonlinear least squares estimation methods.

3.1. Beacon based localization system

A beacon based localization system relies on devices emitting signals from known locations to determine the position of mobile devices that travel in between. Position may be characterized in two or three dimensions, depending on the application. Typically, the "altitude" z coordinate is not considered in applications for people or object localization in indoor environments because maps are often in 2D and there is not much need for that. However some robotics applications require positioning on the three axes. This more complete scenario requires a low vertical dilution of precision (VDOP) only possible with a carefully designed beacon distribution architecture and other optimal conditions. Considering that one of the requirements for universal adoption of a localization system is minimizing installation costs, it is desirable to consider the use of already installed beacons. Therefore one cannot rely on a favorable beacon architecture but instead on one that hopefully was implemented to allow good sound coverage for the frequency range of voice. This is usually present in public spaces for emergency purposes, as said. Therefore the use of two-dimensional location (x, y) may be a good starting point as it simplifies representation and decreases computational cost. Upgrading afterwards to a third dimension is still possible.

Localization algorithms considering the audible sound approach require measurements of distance between the sound beacons and the mobile device. The distance vectors d_i can be obtained using different signal measurements techniques such as time of arrival (TOA) or time difference of arrival (TDOA):
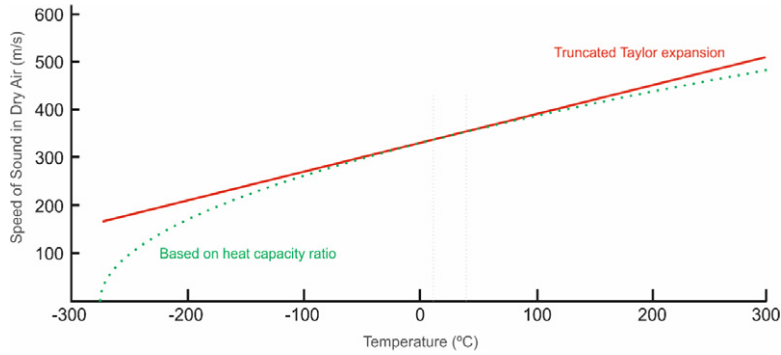
Fig. 2. Speed of sound approximation in dry air based on the heat capacity ratio (in the dashed lower line) versus approximation of using the truncated Taylor expansion (in the solid upper line).

(1) TOA—The time of reception of the signal by the mobile device. Knowing the signal arrival time and subtracting it from the departure time, will provide the duration the signal took to reach the receiver and allow distance calculation. It is the most efficient and usual choice among the options, as it uses the minimum number of beacons (3 for two-dimensional localization).

(2) TDOA—The time difference between a first arrival and subsequent arrivals. This does not require synchronization with the emitter as it uses one beacon as a reference to measure time. It can be measured at least in three ways. In the first case, at least two signals with different propagation velocities are used concurrently to measure TDOA between the signals. In the second case, the differences in time of arrivals measured between different emitters and a receiver can be taken as TDOA. In the third case, the time differences between the arrivals of a signal at different receivers are also defined as TDOA.

From the above described choices, TOA is simpler and less computationally demanding to process than TDOA. However it requires synchronization between beacons and the mobile device. On the other side, TDOA uses the first (closest) beacon to set the time reference and loses its contribution in determining position. TOA localization estimates are a trilateration problem with the simple intersection of the circumferences of the measured distances. TDOA on the other hand uses multilateration and requires the calculation of the hyperbolic intersection of each of these distance difference curves.

Time of Flight (ToF) is represented by t_{TOF}. If a signal propagates with a constant velocity v, the distance d can be calculated by the simple expression:

$$d = t_{TOF} \times v. \tag{1}$$

Measuring the time of flight of an RF signal, which travels with a velocity that is approximately the speed of light $(3 \times 10^8 \text{ m s}^{-1})$ can provide high accuracy [21]. However, very expensive and high resolution clocks must be used. Considering nanoseconds time resolution brings up synchronization problems since transmitters and receiver must be synchronized at that level. For instance, GPS operates with RF signals and measures time using very high resolution atomic clocks in the transmitters.

In the current experiments, measuring ToF with a sound signal still provides a relatively high accuracy using only inexpensive clocks, since the velocity of sound (approximately 343 m s^{-1} at normal conditions) is much smaller than the wave velocity at RF.

In fact, it is necessary to consider that sound velocity is influenced by air temperature and humidity. Considering only the temperature variation, it increases at a 0.6 m s^{-1} °C^{-1} rate in a typical temperature range between 0° and 50 °C. Sound velocity v_0 will therefore be

$$v_0 = 331.45 \sqrt{1 + \left(\frac{T}{273.15} \right)}, \tag{2}$$

where T represents the temperature in Celsius degrees.

As can be seen in Fig. 2, it is appropriate to use the truncated Taylor expansion on the ambient temperature range of interest with no associated significant error in dry air. Humidity has a small but measurable effect on sound speed as well (causing it to increase by about 0.1%–0.6%), because oxygen and nitrogen molecules on the air are replaced by lighter molecules of water. This is a simple mixing effect. Although high humidity causes a higher sound attenuation and fading and therefore causes sound to travel only smaller distances, this is not relevant in indoor spaces and can be neglected.

The reference for determining the d_i distances as a function of the time measurements is depicted in Fig. 3 where the t_i time instants of arrival of each of the beacon's signals to the m receiver may be observed.

Distances d_i between beacons and the mobile device, are defined by

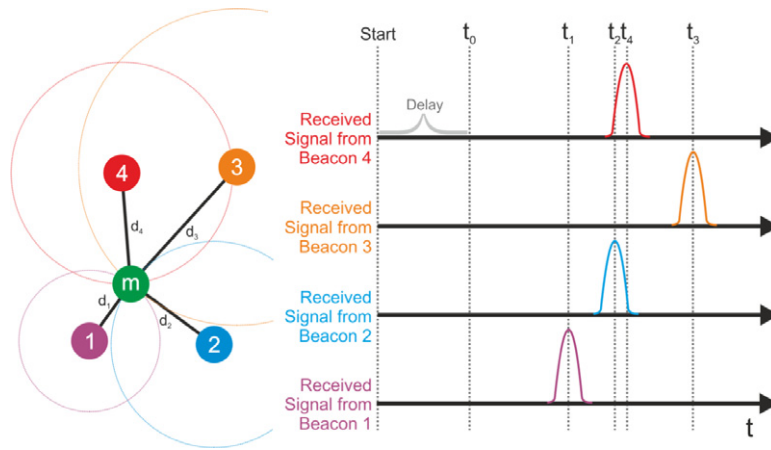$$d_i = v_0 (t_i - t_0) = \sqrt{(x - X_i)^2 + (y - Y_i)^2} \tag{3}$$

Fig. 3. Trilateration method for finding position based on TOA by measuring ToF.

Fig. 4. Illustration of circle shrinking that resulted in 25% reduction. The overestimated distance vectors on each beacon (left) are iteratively reduced to minimize the solution space (right).

where X_i and Y_i are the ith beacon's known and fixed coordinates, t_i the arrival time of each i beacon's wave and t_0 the simultaneous emission time, common to all the beacons. The variables x and y are the unknown coordinates to be determined.

Signals arrival times t_i may be estimated using correlation methods as is explained later in the text. However, solving this problem is more difficult than Eq. (3) suggests because the physical system, which includes the sound production and sensing software and hardware, is not linear. Time instant t_0 is different from zero due to varying delays in processing and transmission from start (as Fig. 3 depicts). Using TDOA would allow clock synchronization between the parts, however it would require one more beacon than TOA and would be more computationally demanding. Similarly, Global Navigation Satellite Systems (GNSS) deal with the same situation of using TOA instead of TDOA as synchronization between the very precise satellite atomic clocks and the cheap quartz ones of the receivers cannot be achieved. Considering that the speed of sound is about a million times slower than the GNSS signals, one can assume that synchronization is easier to achieve as synchronization errors will have far smaller consequences.

Determining t_0 is critical to correctly evaluate distance between the anchors and the mobile device. If the emission is simultaneous in all anchors, t_0 can be determined by optimization techniques. The vectors d_i to all beacons are affected by the same Δd error. In these conditions, this delay may be subtracted using a specially developed technique entitled "Circle Shrinking" that is now explained. In this technique, distance is usually overestimated (because of latency in the emission). Considering the overestimated distances d_i as the radii of circles centered on the beacon's positions, as illustrated on the left side of Fig. 4, the method iteratively does a shrinking of the circles until the intersection area between them is minimized as shown on the right side of Fig. 4.

The local search halt criterion can be a numeric threshold or simply a "stop when there is no intersection". However, performing "circle shrinking" can be computationally demanding. It requires calculation of the intersection area of the circles at each iteration of this minimization problem. The application requirements in precision and accuracy must be taken into account to evaluate what is the reasonable halt criterion. Frequently, a small estimation error in the distance vectors may be acceptable and the source localization algorithm may eventually deal with it very well. For example, one-sample error in
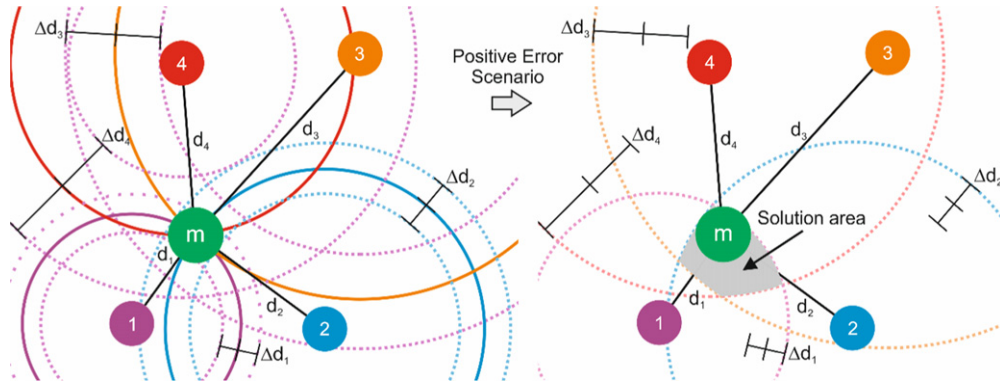
Fig. 5. Location estimation problem based on noisy ToF measurements.

ToF estimation at 44.1 kHz sampling rate represents less than one centimeter error in a distance vector, and an even smaller error in the final position estimation.

Depending on the latency variation (Δt_0) or the application itself, this technique can also be performed only when synchronization between the emitters and the receiver is lost. This will avoid a constant heavier processing and will increase position refresh rate.

To synchronize the infrastructure and the mobile device, a possible approach may be to send the t_0 information on the radiated signal as a time stamp. In a scheme where DS-CDMA is used, the signal information can be the exact time of emission spread with a code interpreted in the receiver. Another possibility is to use a clock (sync) signal together with the signals at every cycle [15]. A previous work [16] has used a dedicated microphone in a known position to calculate the delay in each run. It is a simple solution but it requires additional hardware with implied additional costs. In the conducted experiment presented ahead, a computer sound board with a fixed latency was used to avoid the use of the calibration microphone. Assuring fixed latency in sound emission, and therefore a constant delay, will allow to use circle shrinking only once for the first delay measurement. From that point beyond, delay is considered constant and is simply subtracted resulting in $t_0 = 0$. This strategy avoids the need for additional hardware and does not increase computational complexity.

3.2. Localization estimation

When redundant beacons are employed, an overdetermined equation system applies. Localization can be achieved by trilateration with various triplet combinations of land Ref. [16]. In this case, a direct method can be employed for different sets of reference units, thus generating multiple solutions [7]. The final estimate may then be chosen by computing the average, or by selecting the set that yields the highest signal to noise ratio (SNR). Another possibility is disregarding the least reliable ones with the smallest correlation peak [16], or the ones with the smallest loudspeaker quality parameter q_i [15]:

$$q_i = \max R_{rs_i} / \sigma (R_{rs_i}), \tag{4}$$

where R_{rs_i} represents the cross-correlation between the received signal r and the radiated signal s_i from the ith beacon and σ ( ) represents the standard deviation operator.

Due to the presence of noise, Δd_i, in the d_i measurements, the desired and unknown mobile device's position (x, y) cannot be obtained just by solving the system of equations and expecting to have one optimal solution. The solution is no longer the circles' intersection point m that Fig. 5 ideally describes on the left side. To illustrate this situation, Fig. 5 depicts a case where different errors in d_i will create a larger solution area for the problem (on the right), creating the need for an optimization process.

Therefore, the location will be determined through an algorithm of source localization that considers an error minimization approach. An optimization method can be used to estimate the rectangular coordinates position (x, y) by minimizing the following objective function concerning the error:

$$f(x, y) = \sum_i \left[\sqrt{(x - X_i)^2 + (y - Y_i)^2} - v_0 \cdot t_i \right]^2, \tag{5}$$

where f represents the error function, and one considers the typical constrains in the variable's domains.

Iterative nonlinear least square estimation methods like the Newton–Raphson, Gauss–Newton or Steepest Descent appear in the literature as good methods to calculate solutions to this problem [15].

Newton–Raphson successively finds better approximations to the roots (or zeros) of a certain function given an initial value. The Gauss–Newton algorithm, also used to solve non-linear least squares problems, can be seen as a simplification of the Newton–Raphson method for finding a minimum of a function. Unlike the Newton–Raphson, the Gauss–Newton

Fig. 6. Experiments' physical layout. The larger corner rings represent the beacons (loudspeakers) while the 23 ground truth points are denoted by the small circles in between. The considered central points are 1, 2, 4, 18 and 21 represented in a highlighted circle. All the other points are considered to be outside the central area.

algorithm can only be used to minimize a sum of squared function values, but it has the advantage that second derivatives, harder to compute, are not required. Steepest descent has a slower convergence and can be seen as an approximation of the Newton–Raphson where the second order partial derivatives are not used. It performs a local search by taking steps proportional to the symmetric of the gradient (or of the approximate gradient) of the function at the current point. However it has the lowest computational complexity. To choose which method is more suitable it is necessary to explore the tradeoff between performance in solution convergence and computational effort.

4. Experiments and results

4.1. Experimental setup description

The experiments were performed in a 7 m × 9 m × 3 m research laboratory room. From this total space only a 6 m × 7 m area was used as is depicted in Fig. 6. The room has plaster reverberant walls in which two are outer walls with four large windows, and is occupied by a set of furniture, computers and persons. Twenty three "ground truth" points were marked on the floor as landmarks to allow error estimation.

The sound emission and capture was performed at 44.1 kHz sampling rate using a Pre-Sonus low latency/low noise IEEE1394 interface sound board with ASIO drivers. Four ordinary satellite desktop computer loudspeakers were used as audio anchors (beacons) and were wall mounted at ear level (1.65 m). The mobile omnidirectional condenser microphone on a stand was connected to the previously mentioned sound board. All processing was performed in a 1.6 GHz Dual Core PC Laptop with 2 GB RAM running Windows 8 with Matlab2012b. An illustrative view of the receiving, transmitting and processing devices used in the experiments, is presented in Fig. 7.

The experiment indoor space was sealed, with stable temperature and humidity, so that no effect on the speed of sound was considered regarding humidity and air speed variations. Room reverberation time was measured to be 0.67 s. The room "standard" environmental noise resulting from computers fans, lights, and people, was not altered in any way.

4.2. Evaluation

The experimental activities were centered in performing an innovative comparison between channel access methods for the shared medium by using an audible sound indoor localization system. Experimental settings were defined in such a

Fig. 7. Experimental setup devices (processing, audio capture and emission).

way that a ground base comparison could be applied to the different methods evaluated without affecting the comparison results.

Two experiments were conducted to establish the best methodology to estimate range and position:

- Comparison on three correlation techniques to perform Time Delay Estimation (TDE): cross-correlation, generalized cross-correlation phase transform and maximum likelihood;
- Performance comparison on nonlinear minimization methods to calculate location.

Furthermore a third experiment was done for evaluation of position estimation error and reliability with a sufficient SNR considering the following multiple access techniques: TDMA (with unit pulses), FDMA (with chirps) and CDMA (with a coded pseudorandom noise sequence).

The consequent results will allow establishing a base on which future developments necessary to the implementation of an audible sound indoor based localization system may be built.

4.2.1. Comparison on three pre-selected correlation techniques to perform time delay estimation

Poor ToF measurements will cause bad range estimation and will probably result in bad localization estimation, especially if the error is not systematic. Therefore emphasis should be placed in determining the best possible methodology to evaluate ToF accurately.

A "comparison" between the sent signal and the received one will allow the estimation of the delay and the associated distance. Among other possibilities, three correlation methods were selected to be tested: basic cross-correlation (CC), generalized cross-correlation maximum likelihood (GCC-ML) and generalized cross-correlation phase transform (GCC-PHAT) [22]. They were selected among others because of their computational simplicity and interesting performance in detecting delay [23].

Time delay estimation determines the t_i values in Eq. (3), the most problematic aspect of the distance vectors determination.

Cross-correlation is the simpler tool, as may be read in Eq. (6). Depending on the noise and signal similarity, it can provide a good enough peak allowing determination of delay τ:

$$R_{r_1 r_2}(\tau) = E\left[r_1(t) r_2(t - \tau)\right], \tag{6}$$

where $R_{r_1 r_2}$ represents the cross-correlation between r_1 and r_2 and $E\{\cdot\}$ is the expected value. The delay τ is the value that maximizes this function in delay estimation. The time delay D_{CC} is calculated as

$$D_{CC} = \arg_\tau \max\left[R_{r_1 r_2}(\tau)\right]. \tag{7}$$

The sharper the peak of $R_{r_1 r_2}(\tau)$ will be, the better or easier TDE will be.

The generalized cross-correlation Maximum Likelihood technique is based on the formulation [24] of Eq. (8):

$$R_{r_1 r_2}(\tau) = \int_{-\infty}^{+\infty} \psi_{ML}(f) G_{r_1 r_2}(f) e^{j2\pi f \tau} df, \tag{8}$$

where $G_{r_1 r_2}(f)$ is the cross-spectrum of the received signal and $\psi_{ML}(f)$, the maximum likelihood weighting function defined as

$$\psi_{ML}(f) = \frac{1}{|G_{r_1 r_2}(f)|} \frac{\left|\gamma_{r_1 r_2}(f)\right|^2}{1 - \left|\gamma_{r_1 r_2}(f)\right|^2}, \tag{9}$$
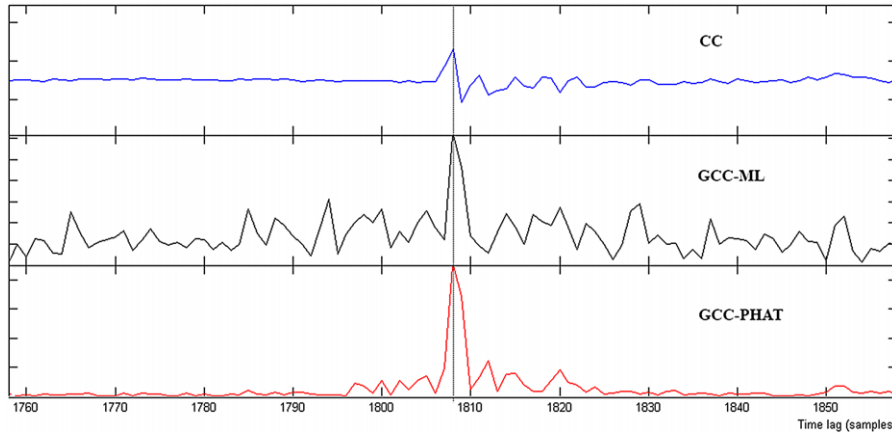
Fig. 8. Correlation methods comparison with an approximate 5 dB SNR for an 1808 samples delay. Cross-correlation (CC), Maximum Likelihood (GCC-ML) and Generalized Cross-Correlation Phase Transform (GCC-PHAT).

and:

$$\left|\gamma_{r_1 r_2}(f)\right|^2 = \frac{\left|G_{r_1 r_2}(f)\right|^2}{G_{r_1 r_1}(f)\, G_{r_2 r_2}(f)} \tag{10}$$

is the squared magnitude coherency.

Time delay D_{ML} is similarly determined by the generalized cross-correlation peak value:

$$D_{ML} = \arg_\tau \max\left[R_{r_1 r_2}(\tau)\right]. \tag{11}$$

The $\gamma^2/(1-\gamma^2)$ term of the weighting function will provide greater weight on frequency bands that result in coherence close to one. On the other hand, frequencies in which coherence is low are deemphasized. The GCC-ML process weighs the cross-spectral phase according to its estimated value when the variance of the estimated phase error is the lowest. Therefore, the GCC-ML weighting function $\psi_{ML}(f)$ is chosen to improve the accuracy of the estimated delay by attenuating the signals fed into the correlator in the spectral region where the SNR is the lowest.

Alternatively, the Generalized Cross-correlation Phase Transform (GCC-PHAT) method has been demonstrated to provide a better delay detection in low white noise environments [8,20]. It was thought as a way to sharpen the cross correlation peak by whitening the input signals with a weighting function. The PHAT is a GCC procedure which has received considerable attention due to its ability to avoid spreading of the peak of the correlation function. It is popular for its excellent performance in noisy environments, even under relatively heavy reverberation [25]. It can be expressed mathematically by:

$$R_{r_1 r_2}(\tau) = \int_{-\infty}^{+\infty} \psi_P(f)\, G_{r_1 r_2}(f)\, e^{j2\pi f\tau}\, df, \tag{12}$$

where $\psi_P(f)$ is the PHAT weighting function which is defined by;

$$\psi_P(f) = \frac{1}{\left|G_{r_1 r_2}(f)\right|}. \tag{13}$$

Time delay D_P is again determined by finding the correlation maximum peak as

$$D_P = \arg_\tau \max\left[R_{r_1 r_2}(\tau)\right]. \tag{14}$$

Only the phase information is preserved after $G_{r_1 r_2}(f)$ is divided by $\left|G_{r_1 r_2}(f)\right|$. This causes the effect of sharpening of the correlation function peak ideally by resembling a delta function at D_P.

When the environment noise is low, PHAT is indeed a special case of the GCC-ML algorithm, which explains its good performance under those circumstances. Also, as long as the noise stays low, PHAT remains optimal in the GCC-ML sense even when the room reverberation is intense, which explains its robustness over reverberation [26].

The experimental comparison has provided confirmation of literature results and demonstrates that GCC-PHAT performs better, especially in reverberant situations like in the conducted experiment. Furthermore, in Fig. 8 it is possible to visually interpret the sharper peak area that occurs even in a rather low SNR scenario. This illustrative situation of TDE would be correctly determined in the three methods. However, in less correlated situations, peak detection is not so clear, and TDE errors will occur if there is no clear peak area separation in amplitude.

Automatic peak detection mechanisms, used for determining TDE, will therefore perform better if the correlation signal has a higher crest factor. This improvement will become even clearer in situations where other peaks with similar amplitudes

Table 2

Comparison between the non-linear estimation methods for localization.

	Gauss–Newton	Newton–Raphson	Steepest descent
Runtime of each iteration (μs)	35.70	40.82	34.30
Iterations to convergence	3	3	8
Total time (μs)	107.10	122.46	285.60
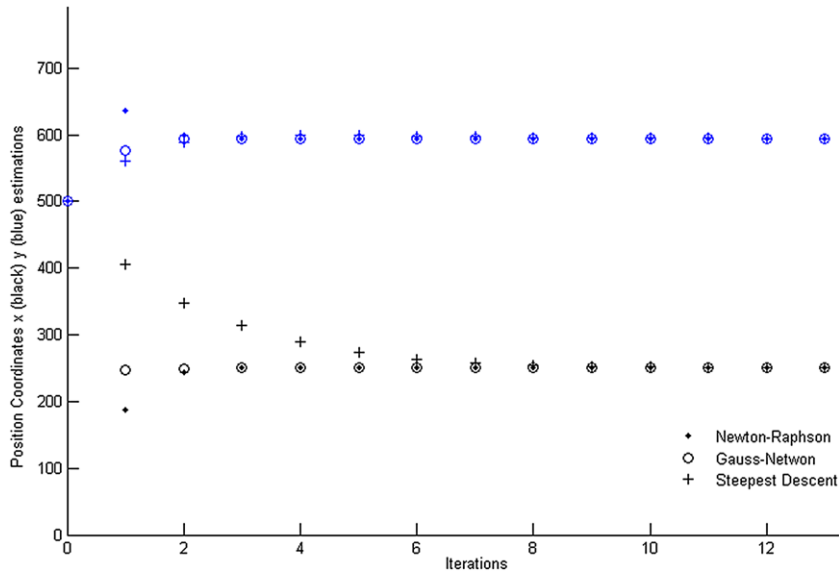
Fig. 9. Non-linear square estimation methods experimental comparison. With the same initial first guess position (500, 500) searching for (250, 600) with a 20 dB SNR signal (defined as the mean of squared ranges over noise variance).

occur in the vicinity due to reverberation. As observed, GCC-PHAT significantly reduces the spreading of the correlation peak. When comparing to the ML method one can also notice a significant difference in the baseline noise which is attributed to the reverberant environment of the experiment.

4.2.2. Applied comparison on non-linear square estimation methods for position estimation

The combination of noisy range measurements require a local search for a solution in the location space regarding the minimization of the error objective function. Among several nonlinear estimation methods, a selection was made of those well suited to the error distribution and to the data type, while maintaining low computational complexity and not requiring adjustable parameters. The selection consists in the Newton–Raphson, Gauss–Newton and Steepest Descent methods [17].

Among these three possibilities it is important to evaluate the most appropriate one for minimizing the localization estimation processing time and to obtain the best tradeoff between the smallest possible error and computational complexity. In this particular case, this may be a bottleneck in the performance of the localization system as it will have to be repeated many times at each localization cycle (position refresh).

Considering the runtime duration per iteration presented in Table 2, the methods confirmed the expectations regarding their respective computational complexity. Gauss–Newton runs each iteration in an average time of 35.70 μs, Newton–Raphson in 40.82 μs and Steepest Descent in the fastest with 34.30 μs. Second derivative calculations in Newton–Raphson are slightly noted. Absolute timings have a small practical interest as they rely on the processing platform, however, a relative analysis reveals differences that will influence the localization estimation processing time with consequences that will be as large as the number of times the method will be used in each localization update.

A comparison among these three possibilities was conducted regarding the number of necessary iterations to achieve precise localization and results are presented in Fig. 9. Using an initial guess position and 4 relatively good range measurements, it is possible to observe that at the 3rd iteration, Gauss–Newton and Newton–Raphson converge to very accurate and precise results (error smaller than 5 m). The same cannot be said about the Steepest Descent method that requires 6 to 8 iterations to reach a similar kind of performance even with a carefully selected step size.

The number of iterations necessary to reach the stopping criterion on the local search will also depend on the measurements error. In this experiment with 20 dB SNR signal, defined as the mean squared distance over noise variance, convergence happens in 3 to 6 iterations. Noisier signals, with smaller SNR, would require more iterations to converge and would also emphasize the methods' performance differences.
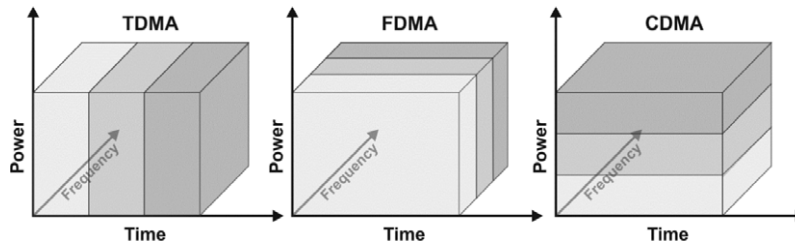
Fig. 10. Graphical comparison between TDMA, FDMA and CDMA methods.

Table 3
Results comparison between TDMA, FDMA and CDMA.

	TDMA	FDMA	CDMA
Average error (cm)	5.3	5.4	4.5
Average error inside center area (cm)	1.5	3.0	1.3
Average error outside center area (cm)	6.4	6.0	5.4
Standard deviation (cm)	8.1	6.4	5.4
Standard deviation inside center area (cm)	0.7	1.3	0.5
Standard deviation outside center area (cm)	8.9	7.2	5.8
Reliability filtered inside center area (%)	100.0	100.0	100.0
Reliability filtered outside center area (%)	98.3	98.9	97.8
Reliability inside center area (%)	100.0	80.0	100.0
Reliability outside center area (%)	86.7	51.9	63.9
Minimum SNR (dB)	24.7	11.4	7.2

The requirements of an indoor localization system impose taking into account the processing time of the methods as an important variable to the position update frequency. As may be read in Table 2, the most computationally simple method is the one that requires more time to calculate an accurate position (2.7 times more than the fastest one). Therefore, from these pre-selected methodologies one can conclude that, among these possibilities, the Gauss–Newton is the best choice for handling the optimization problem of minimizing the error, and consequently finding the best position using noisy ToF measurements.

4.3. Multiple access shared channel techniques comparison

The conducted experiment evaluates the use of three different methods to convey the excitation audible sound signal to a receiver so that TDE can be as accurate, precise and reliable as possible in real conditions (the noisy reverberant space). These three multiple access techniques, explore differently how the signal is conveyed to the receiver and Fig. 10 describes their conceptual differences.

In TDMA, a short pulse (an almost unperceivable spike in the loudspeakers) was sent by each beacon in a dedicated time window. A guard period was employed to avoid erroneous pulse detection originated from previous reverberations or echoes. In the FDMA experiment, each beacon was assigned with a non-overlapped range of logarithmic chirps between 4 and 15 kHz uniformly distributed to reproduce a resulting sound similar to a bird singing, and therefore subjectively less artificial. All sounds are emitted simultaneously in different frequency bands. In CDMA, simultaneous emission also occurs. However, a specific coded pseudorandom noise sequence is used for each beacon resulting in a white noise like signal. Each of these sequences has a high auto-correlation and a low cross-correlation. This allows better detection by the receiver.

The comparison between these three multiple access techniques was performed considering 20 runs in each of the 23 positions depicted in the floor plan of Fig. 6. The average accuracy error, defined as the Euclidean distance between the estimated and the ground truth positions, was then calculated. Robustness was evaluated considering two possible scenarios inside or outside center area: using all four range estimations with the same weights or not considering the worst range measure. Table 3 summarizes the results comparing average error, robustness and minimum SNR on these three methods considering an average minimum SNR so that reliability of the distance vectors does not decrease below 50% in the worst measurement position.

Results in Table 3 demonstrate that the CDMA method has performed slightly better than the other two. Achieving a 1.3 cm accuracy in the central points may be considered in the range of the best state-of-the-art results. The chirp (bird like) FDMA approach had difficulties in estimating some positions because of the signal's relatively small bandwidth. Loudspeaker directivity at these frequencies has caused that reverberations were interpreted as the direct signal (that is not detectable) in the positions where there is no line of sight to beacons. This problem has an even more pronounced effect when the mobile position is close to walls and corners. In these cases, with a directionality factor greater than one, higher intensity reverberations may be erroneously interpreted as direct signals causing distance d to be overestimated.
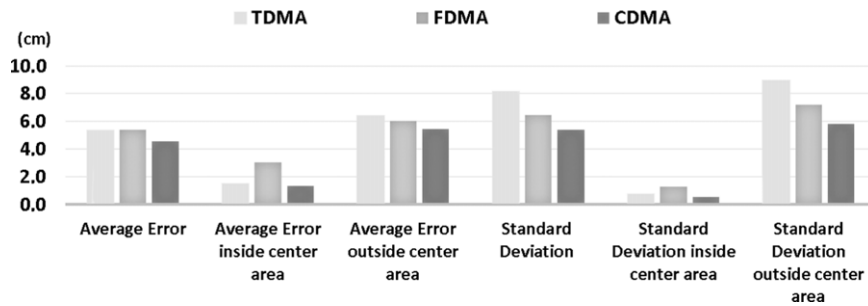
Fig. 11. Techniques comparison depending on the mobile's device location.

The other wide band approaches, TDMA and CDMA, were not affected in the same way. Yet, data also demonstrates significantly better results in central points. The directivity issues in the loudspeakers or in the mobile microphone may be considered together with the frequency response of all the parts, as it will affect the ability to perform TDE. Channel's equalization may also have to be considered to reduce TDE errors.

One of the most meaningful observations is related with the minimum SNR that each method requires. It is an additional possible measure of effectiveness. As previously mentioned, TDMA pulse method is very demanding. In this case it only performed well (with a robustness criteria of 30 cm error in distance vectors) above 24.7 dB SNR. On the other hand, CDMA performed very well with 7.2 dB minimum SNR. At this sound level, the noise like signal was found to be almost undetectable in the acoustic environment by a panel of subjects. This complies with the localization requirements performance without being acoustically annoying. A detailed psychoacoustic study considering the annoyance factor and other metrics was not yet done at the time. Nevertheless, it was consensual amongst the panel that all the sounds used were not particularly annoying. This is an important information considering the use of audible sound in an indoor localization application where people may co-exist.

TDMA method has shown not to be robust enough in a noisy environment. Even though reliability is one of the greatest, the required minimum SNR is quite larger than the others. The pulse detection technique used was based on peak detection and therefore, a simple impulsive masking noise is sufficient to make the TDE fail and consequently everything else. It is also possible to conclude that extra beacons may not contribute to a better position if they are considered without other criteria. In the FDMA case, where loudspeaker directivity has clearly affected some positions, there was an improvement from 51.9% to 98.9% on non-central positions just by disregarding the worst beacon measurement and considering only the necessary three.

The comparison from Fig. 11 reveals that the center positions (points 1, 2, 4, 18 and 21 in Fig. 6) have very good precision and accuracy. FDMA performs precisely but with a considerable smaller accuracy than TDMA and CDMA. Outside the central points, average errors and standard deviations increase significantly and some differences may be observed between methods. CDMA has the smaller average error and standard deviation. FDMA follows and finally TDMA, with the worse results, confirms the susceptibility of shorts pulses in a situation where wall reverberations visibly degrade precision and accuracy.

Standard deviation for all points follows a similar tendency although with lower values. This leads to the conclusion that results are more precise than accurate. This smaller accuracy may be the result of some systematic errors relative to the ground truth measurements. Nevertheless, centimeter level precision is in general an interesting result and, for comparison purposes, the absolute measurement error is not as important as the low standard deviation which demonstrates interesting consistency in the results.

In Fig. 12 the Horizontal Dilution of Precision (HDOP) performance of the conducted experiment may be observed, where the poorer results in average error are on non-central points and especially close to walls or sound reflective objects. The HDOP performance of a certain indoor space is of great interest when comparing methods and when considering the economical and simplicity advantages of using off-the-shelf pre-installed loudspeakers. It may not be reasonable to assume changing beacon positions in some indoor spaces as they were previously installed for other purposes. Therefore, it is of great interest to choose the best possible way to avoid great influence of the HDOP in the indoor localization system. This can be achieved by using wider band lower frequency signals that will reach a larger area providing more coverage and diversity of beacons in the frequency range of the microphone's mobile device. Having more available signals from beacons will diminish the negative effects of a non-favorable beacon distribution for some positions in the horizontal plane. This is of particular importance in situations where the loudspeakers are pre-existent and the loudspeaker installation design was based in some criteria like sound coverage, voice intelligibility or even security, and not in geometrically useful properties (non-collinearity, distance and angular separation, etc.) that would be more favorable for a smaller HDOP. This will surely be an advantage when using audible sound as signal and is one of the most observable results of this experiment.

The bar graph presented in Fig. 13 presents the localization estimation error average and the standard deviation regarding the 23 test positions ordered from the geometrical center of the test area (on the left) to the furthest anchor (on the right), in the CDMA experiment. From this figure it is possible to observe the relation between the distances to the center of the room,
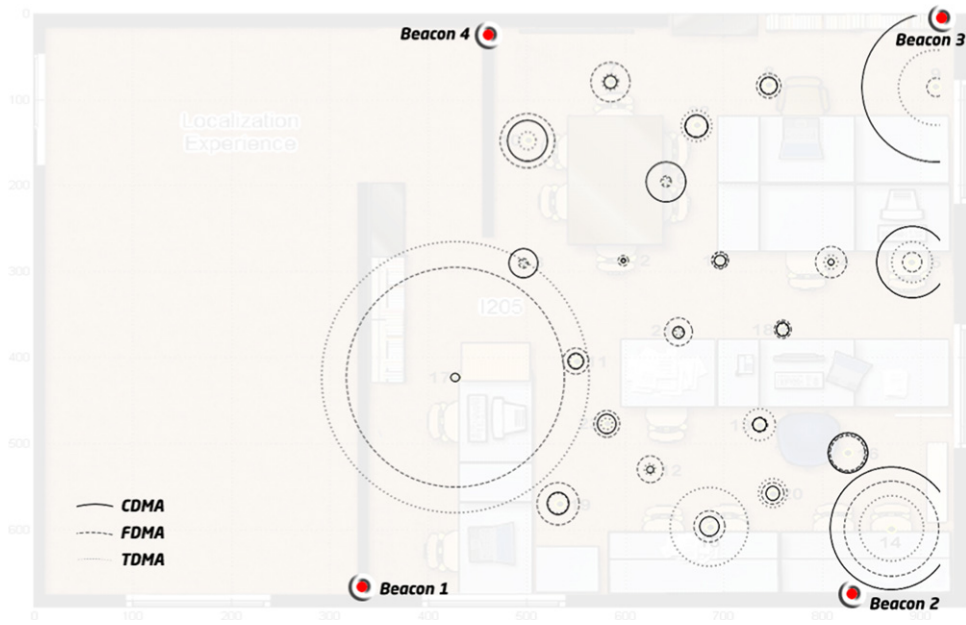
Fig. 12. Graphical Interpretation of the system's Horizontal Dilution of Precision (HDOP) with a 3.4× scale in average error. Continuous darker circles—CDMA, dashed medium gray—FDMA and light gray—TDMA.
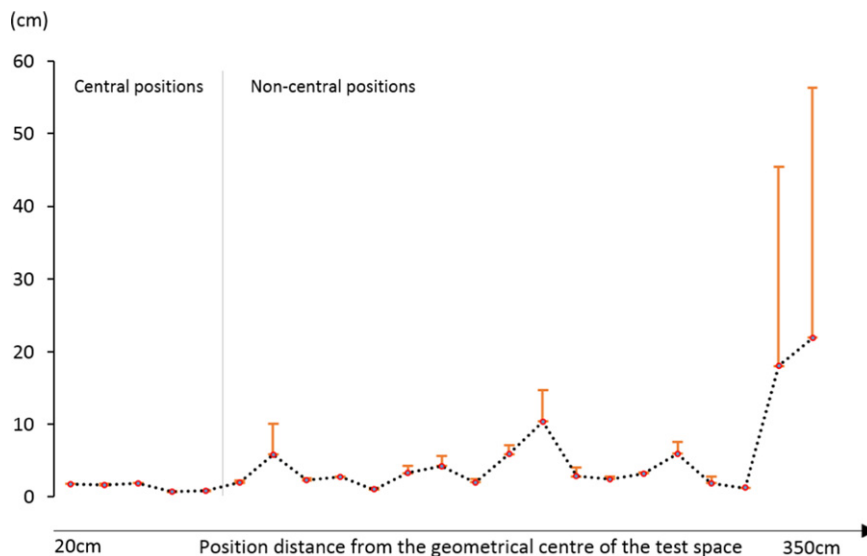

Fig. 13. Error bar graph discriminating central and non-central positions regarding localization estimation error average and standard deviation in the CDMA experiment. The 23 test positions were ordered according to the Euclidean distance to the geometrical center of the test area.

where the walls are further away, and the accuracy and precision of the localization estimates. This graph confirms visually the inverse proportionality between the precision and accuracy and the Euclidean distance to the center of the room which represents the positions where multipath effects may be less experienced.

Considering the best performing multiple access shared channel technique, CDMA, Fig. 14 shows the percentage of realizations compared with accuracy (Tolerance). It is possible to observe that 95% of all localization estimates are placed within 10 cm radius anywhere in the room.

If only central positions are considered (at least 1.5 m away from walls), 95% of the time estimates are below 2 cm error. Another observable result is that all localization estimates have localization estimation errors below 1 m. This maximum error is by itself compatible with most possible context-aware applications. Moreover, considering that in 95% of time, the localization is within 10 cm, it is possible to assume that in random situations the remaining 5% estimates may be filtered by using recursive Bayesian estimators such as Kalman or Particle Filters. Their use could also be explored to help in minimizing Doppler Effect consequences that may occur in a moving device at higher speeds [27].
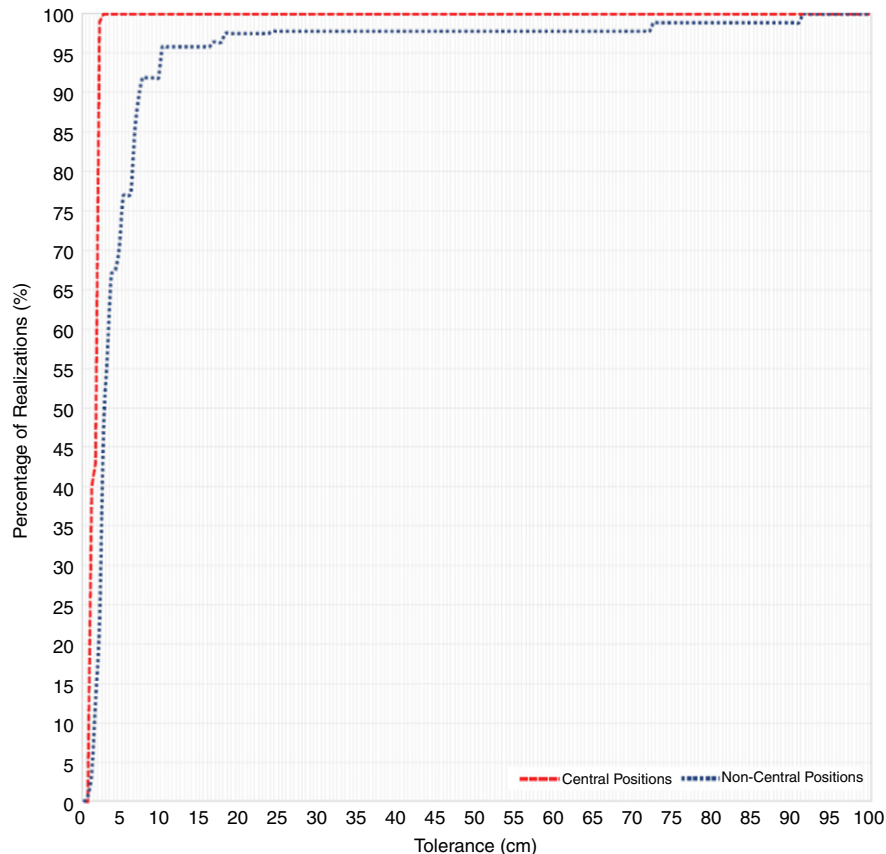
Fig. 14. Graphical plot of percentage of realizations while considering tolerance for two scenarios: localizations estimates in central and non-central positions in the CDMA experiment.

5. Conclusions and future work

An audible sound-based indoor localization system using off-the-shelf pre-installed loudspeakers was implemented, tested and evaluated. Its methods and techniques were chosen based on innovative comparisons concerning this specific application and its possible wide dissemination in real applications.

Using GCC-PHAT to determine ToF from the mobile device to the beacons was found to be the best choice for this audible sound-based approach. A sharper peak eases the synchronization process and allows more accurate ToF estimations especially in noisy environments. Also, the Gauss–Newton estimation method has proven to be the best choice for determining the position based on imprecise range measurements. The tradeoff between its fast convergence and its low computational complexity makes it the best choice to compute localization, as it is performed many times in each localization cycle. Even a small difference when comparing with the other methods will result in relevant improvement in the localization update frequency.

The results of the performed experiment on multiple access shared channel techniques demonstrated that CDMA performed better in order to achieve accurate and precise positioning even with a low SNR. Its wide band characteristic, and the wider radiation aperture, makes it very appropriate especially for cases where beacons are in fixed pre-determined (not chosen) positions. Using CDMA also fulfills the objective of minimizing any caused disturbance in the acoustic environment with an almost unperceivable signal at 7.2 dB SNR.

It may be concluded that it is necessary to establish criteria on how to use redundant beacons. It is of capital importance not to use all the beacons while estimating localization. That will seriously deteriorate results, especially in low HDOP situations, as it was demonstrated. Using a weighting strategy based on the relevance of each beacon or choosing the most favorable minimum set of beacons has shown to improve results in a significant way.

Audible sound has been shown to be a viable signal to estimate position indoors. It is also noteworthy that the physical setup where the experiments took place has successfully achieved the status of a working audible sound localization system in almost its full dimension. The only concession concerns the mobile device which was emulated by means of an omnidirectional microphone. Addressing the mobile device itself can be seen as an independent problem to be treated.

In the future the work will be focused in improving the CDMA through the acoustic channel so that information other than just the signal identification can travel from the beacons to the mobile device and be used to improve localization.

Considering the effects of a moving receiver will also be studied as a part of the evolution to a possible indoor position tracking application.

Acknowledgments

References

[1] R. Ferraro, M. Aktihanoglu, Location-Aware Applications, first ed., Manning, Shelter Island, NY, 2011.
[2] J. Rantakokko, J. Rydell, P. Stromback, P. Handel, J. Callmer, D. Tornqvist, F. Gustafsson, M. Jobs, M. Gruden, Accurate and reliable soldier and first responder indoor positioning: multisensor systems and cooperative localization, IEEE Wirel. Commun. 18 (2) (2011) 10–18.
[3] S. Chan, Toward better indoor localization: cooperative localization and estimation fusion, in: Proceedings of the 1st International Workshop on Mobile Location-Based Service, (MLBS'11), ACM, New York, NY, USA, 2011, pp. 51–52.
[4] J. Brown, Indoor Location Market [Indoor Positioning and Indoor Navigation (IPIN); Indoor Mapping; Indoor LBS; Indoor Analytics; By Positioning Systems (Network-based, Independent, Hybrid)]: Global Advancements, Market Forecasts and Analysis (2013–2018), Marketing Report, June 2013.
[5] G. Deak, K. Curran, J. Condell, A survey of active and passive indoor localisation systems, Comput. Commun. 35 (16) (2012) 1939–1954.
[6] L. Mainetti, L. Patrono, I. Sergi, A survey on indoor positioning systems, in: 2014 22nd International Conference on Software, Telecommunications and Computer Networks, SoftCOM, 17–19 September 2014, 111–120.
[7] N.B. Priyantha, The cricket indoor location system (Ph.D. dissertation), Massachusetts Institute of Technology, 2005.
[8] V.T. Pham, Q. Qiu, A.A.P. Wai, J. Biswas, Application of ultrasonic sensors in a smart environment, Pervasive Mob. Comput. 3 (2) (2007) 180–207.
[9] C. Brignone, T. Connors, G. Lyon, S. Pradhan, SmartLOCUS: An autonomous, self-assembling sensor network for indoor asset and systems management, Mobile Media Syst. Lab., HP Laboratories, Palo Alto, CA, Tech. Rep., Vol. 41, 2003.
[10] J. Prieto, A. Jiménez, J. Guevara, Subcentimeter-accuracy localization through broadband acoustic transducers, Intell. Signal Process. (2007).
[11] L. Segers, J. Tiete, A. Braeken, A. Touhafi, Ultrasonic multiple-access ranging system using spread spectrum and MEMS technology for indoor localization, Sensors 14 (2) (2014) 3172–3187.
[12] S. Knauth, C. Jost, A. Klapproth, iLOC: a localisation system for visitor tracking & guidance, Ultrasound 2 (2) (2009) 3.
[13] K. Liu, X. Liu, L. Xie, X. Li, Towards accurate acoustic localization on a smartphone, in: INFOCOM, 2013 Proceedings IEEE, April 2013, pp. 495–499.
[14] T. Aguilera, J. Paredes, F. Alvarez, J. Suarez, A. Hernandez, Acoustic local positioning system using an iOS device, in: International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2013, pp. 1–8.
[15] C. Lopes, A. Haghighat, A. Mandal, T. Givargis, P. Baldi, Localization of off-the-shelf mobile devices using audible sound: architectures, protocols and performance assessment, ACM SIGMOBILE Mob. Comput. Commun. Rev. 10 (2) (2006) 38–50.
[16] C. Sertatil, M. Altinkaya, K. Raoof, A novel acoustic indoor localization system employing CDMA, Digit. Signal Process. 22 (3) (2012) 506–517.
[17] Ish Rishabh, Don Kimber, John Adcock, Indoor localization using controlled ambient sounds, in: Indoor Positioning and Indoor Navigation (IPIN) International Conference, vol. 13, IEEE, 2012, p. 15.
[18] F. Hatfull, Watermarking audio data—A survey and comparison of techiques for audio steganography, Case Western Reserve University, April 2011.
[19] W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding, IBM Syst. J. 35 (34) (1996) 313–336.
[20] A. Garcia, Digital watermarking of audio signals using a psychoacoustic auditory model and spread spectrum theory, in: Proc. 107th Conv. Aud. Eng. Soc., 1999.
[21] K. Srinivasan, T. Sundararajan, S. Narayanan, T.J.S. Jothi, Effects of acoustic source and filtering on time-of-flight measurements, Appl. Acoust. 70 (8) (2009) 1061–1072.
[22] W. Xinwang, W. Zhenyang, Sound source localization based on discrimination of cross-correlation functions, Appl. Acoust. 74 (1) (2013) 28–37.
[23] H.C. So, Source localization: Algorithms and analysis, in: S.A. Zekavat, M. Buehrer (Eds.), Handbook of Position Location: Theory, Practice and Advances, Wiley-IEEE Press, 2011 (Chapter 2).
[24] Y. Huang, J. Benesty, J. Chen, Time delay estimation and source localization, in: J. Benesty, M.M. Sondhi, Y. Huang (Eds.), Springer Handbook of Speech Processing, Springer-Verlag, 2008, pp. 1043–1064. EiCs (Chapter 5.1) Part I.
[25] H. Khaddour, A comparison of algorithms of sound source localization based on time delay estimation, Elektrorevue 2 (1) (2011).
[26] C. Zhang, D. Florêncio, Z. Zhang, Why does PHAT work well in lownoise, reverberative environments? in: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference, April 2008, pp. 2565–2568.
[27] S. Widodo, T. Shiigi, N. Hayashi, H. Kikuchi, K. Yanagida, Y. Nakatsuchi, Y. Ogawa, N. Kondo, Moving object localization using sound-based positioning system with doppler shift compensation, Robotics 2 (2) (2013) 36–53.