Guoguo: Enabling Fine-grained Indoor Localization via Smartphone

Kaikai Liu, Xinxin Liu, and Xiaolin Li
Scalable Software Systems Laboratory
University of Florida, Gainesville, FL 32611
kaikailiu@ufl.edu, xinxin@cise.ufl.edu, andyli@ece.ufl.edu

ABSTRACT

Using smartphones for accurate indoor localization opens a new frontier of mobile services, offering enormous opportunities to enhance users' experiences in indoor environments. Despite significant efforts on indoor localization in both academia and industry in the past two decades, highly accurate and practical smartphone-based indoor localization remains an open problem. To enable indoor location-based services (ILBS), there are several stringent requirements for an indoor localization system: highly accurate that can differentiate massive users' locations (foot-level); no additional hardware components or extensions on users' smartphones; scalable to massive concurrent users. Current GPS, Radio RSS (e.g. WiFi, Bluetooth, ZigBee), or Fingerprinting based solutions can only achieve meter-level or room-level accuracy. In this paper, we propose a practical and accurate solution that fills the long-lasting gap of smartphone-based indoor localization. Specifically, we design and implement an indoor localization ecosystem Guoguo. Guoguo consists of an anchor network with a coordination protocol to transmit modulated localization beacons using high-band acoustic signals, a realtime processing app in a smartphone, and a backend server for indoor contexts and location-based services. We further propose approaches to improve its coverage, accuracy, and location update rate with low-power consumption. Our prototype shows centimeter-level localization accuracy in an office and classroom environment. Such precise indoor localization is expected to have high impact in the future ILBS and our daily activities.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication; H.5.5 [**Sound and Music Computing**]: Signal analysis, synthesis, and processing

Keywords

Localization, Smartphone, Anchor Network, Acoustic Signal

MobiSys'13, June 25-28, 2013, Taipei, Taiwan

1. INTRODUCTION

Current coarse-grained (room-level or meter-level) indoor localization on a smartphone has enabled a lot of mobile services, such as location-based services, indoor maps, and indoor navigation systems. However, these services are severely limited due to low resolution. Indoor users can hardly navigate like using outdoor GPS services. The major difference is as follows: meter-level (e.g. GPS with five-meter accuracy) localization is sufficient to navigate a car (meter-level footprint) on a street (several-meter footprint); but it is far from sufficient to navigate a user (foot-level footprint) in a library (with half-meter-wide isles and inch-level books) or a shopping mall (with inch-level items). Smartphone-based accurate "indoor GPS" or IPS (indoor positioning system) has long been waited to improve indoor mobile services and enable new services. Despite significant efforts on indoor localization in both academia and industry in the past two decades [2, 3, 4, 5, 23, 26], highly accurate and practical smartphone-based indoor localization remains an open problem. Some accurate localization solutions cannot be readily converted to smartphone-based ones due to various constraints.

The technology that enables centimeter or decimeter level location accuracy and integrates with context-aware mobile services has the potential to revolutionize users' mobile experiences. Using ranging, angle, displacement or fingerprinting based sensing is required for accurate indoor localization. However, none of existing solutions could achieve desired performance at low cost and low complexity. Time-of-Arrival (TOA) based ranging approaches, e.g., special devices using ultra-wideband or ultrasound signals, are more reliable and accurate. However, the increased complexity and additional devices make them not practically useful for conventional users. Other low-complexity approaches rely on existing sensors in smartphones to infer current location, e.g., WiFi, fingerprinting, compasses and accelerometers. However, their low accuracy and prerequisites like site survey or pairing limits their applications.

Recent research on leveraging ubiquitous microphone sensors in a smartphone introduces a convenient approach of ranging by using the audible band acoustic signal (less than 20kHz). A microphone sensor is inexpensive and has potential for high accurate ranging due to the low transmission speed of acoustic signals. However, the limited bandwidth of a microphone, strong attenuation of aerial acoustic signals, as well as various interferences in the audible band, poses significant challenges in using acoustic signals for indoor localization. Although using acoustic signal to perform

ranging-based localization suffers from various issues, such as short operation distance, low update rate, and sound pollution, the potential of centimeter-level localization accuracy motivates us to design better solutions to overcome its drawbacks.

In this paper, we propose an indoor localization system called "Guoguo" [1], and further improve its performance in terms of coverage, accuracy, update rate, and sound pollution. We make its acoustic beacons imperceptible to humans and improve detection sensitivity for better location coverage. Rather than simply using "Beep" signals, to enable passive sensing for multiple smartphone users, we design the transmission waveform, wide-band modulation, and one-way synchronization and ranging schemes. We design a transmission scheme of the acoustic beacon that follows the high-density pseudo-codes to enable anchor node identification without radio assistance on a smartphone. We propose a *symbol-interleaved* beacon structure to overcome the drawback of the low transmission speed of acoustic signal and improve the *location update rate*. To improve the accuracy of ranging, we propose a fine-grained adaptive time-of-arrival (TOA) estimation approach that exploits the details of the beacon signal and perform Non-line-of-sight (NLOS) identification and mitigation. By combining all these techniques together, we implement the prototype system with anchor nodes and localization processing app in a smartphone, and make it work in realistic environment.

The rest of the paper is organized as follows. Section 7 summarizes the related work. Section 2 introduces the design consideration and system architecture. Section 3 presents the acoustic beacon and anchor network design. Section 4 presents the ranging and localization approaches by using smartphone. Section 5 presents the experimental evaluation of our proposed approach. Section 9 discusses the future work. Section 8 concludes the paper.

2. SYSTEM DESIGN

2.1 Motivation

Various existing localization solutions rely on different assumptions and are dedicated to specific applications. For most existing solutions using only built-in sensors in a smartphone, rough position information can be obtained for building- or room-level navigation. For indoor mobile services for shelf-to-shelf navigation, location-aware and context-aware information recommendation and virtual-reality interaction, fined-grained indoor localization with the foot-level accuracy is critical.

In terms of system complexity, users' side is more stringent especially in hardware requirement. It is very hard to persuade the consumers to purchase special devices for indoor localization. Even with special devices, the integration with the mobile services provided by smartphone is also a difficult problem. To enable centimeter or decimeter level localization accuracy, and only require smartphone on the user side, the investment on indoor infrastructure is necessary and potential for other interesting applications, e.g., surveillance and monitoring. For normal retail stores, several hundred dollars investment on indoor infrastructure to enable indoor "smart" shopping, could attract more consumers to enjoy the mobile services, and in turn boost their business.

With these two points in mind, we focus on developing a fine-grained smartphone-based indoor localization system leveraging indoor low-complexity anchor network.

2.2 Design Challenges and Consideration

To meet the localization accuracy, TOA-based ranging is more appropriate [21]. Limited by the smartphone, using acoustic signal for TOA-based ranging is the only solution. However, making the acoustic signal for ranging and localization in real situations has a lot of challenges.

First, the coverage of the anchor network should be sufficient for a room or a retail store with less than 10 anchor nodes placed to minimize the infrastructure cost. The possible solution is to increase the transmit power or detection sensitivity. However, increase the transmit power could cause sound pollution and high energy cost. [17, 22] all suffer from sound pollution caused by the audible signal. H. Liu et al. [14] achieves less pollution, but only works within 3 meters. Maintaining a low transmit power and keeping the acoustic beacon unnoticeable are the two prerequisites in the anchor network design. Thus, improving the detection sensitivity in the smartphone by using advanced signal processing technique is the feasible solution to ensure sufficient coverage.

Second, sufficient update rate is required to track users' movement. If the localization update rate is too sluggish, no user could be patient enough to wait their location results. There are also not enough time margin for mobile services. Existing acoustic ranging solution proposed in [14] shows 7.8s delay in one location calculation; authors in [19] perform localization for desktop and calculates the running time for 5 nodes in the range of 3.7s to 285s. Possible solutions for improving the location update rate includes better design of the anchor network protocol, real time implementation of the algorithm, and eliminate the use of WiFi assistance (takes about 2s for one scan).

Third, the system should support multiuser simultaneously. However, solutions used in [22, 14, 19] utilize the two-way approach for ranging. Two-way ranging eliminates the synchronization requirement, but suffer from the limitation of only allowing one user to access at one time. Sound source localization [25] is another way to eliminate the synchronization by using active transmit mode for users, however, the random access time of the user making the sensor network could only handle limited users at the same time. One-way passive ranging should be utilized to maximize the multiuser capacity, which could support unlimited number of users in theoretical.

2.3 System Architecture

In order to meet the requirements mentioned in Section 2.1 and 2.2, we propose to the develop an anchor network with better coverage and make the beacon unnoticeable. With several low-complexity anchor nodes mounted on indoor places, the smartphone can localize itself by receiving more than three beacon signals from the anchor nodes. The simplified architecture of our proposed localization system is shown in Fig. 1. In the receiver side, we implement advanced signal processing and ranging algorithm in smartphone to achieve accurate localization in passive mode. The cloud server provides mobile services, e.g., localization, naviga-

[1]"Guoguo" means katydid/bush-cricket in Mandarin Chinese. Guoguo, famous for its beautiful sound, represents a pet culture in ancient China for thousands of years

tion, for the accessed users in real-time response mode and minimize the computation and storage cost in the smartphone.

Specifically, the anchor network transmits modulated acoustic beacon signals based on the token of the controller periodically. By detecting and extracting the desired information bits embedded in beacon signals, the smartphone demodulates the symbols and calculates the relative TOA. The algorithm in smartphone eliminates erroneous measurements using statistical pruning methods, and access the anchor position by matching the pseudocodes. Finally, the smartphone can be aware of its fine-grained position by accessing the localization results.

3. ANCHOR NETWORK DESIGN

3.1 Design Criterion

In our anchor network design, an anchor transmits the spatial beacon signal to inform its unique location to a smartphone without relying on additional radio communication devices. Such basic design criterion requires the anchor network to support acoustic communication in addition to ranging. Moreover, anchor nodes and mobile phones are distributed without explicit connection or pairing, and use acoustic communication to realize synchronization. Compared with systems using a radio signal for synchronization and communication to assist acoustic ranging, e.g., Cricket [23] and Beep [17], our system implements both acoustic communication and localization. With acoustic communication and ranging, our solution features the following salient advantages: no need of special radio signal (e.g. ZigBee) that is not available on a smartphone; no need of special devices for ranging assistance; enabling one-way passive ranging to support massive users like the outdoor GPS. Ranging based on the acoustic communication channel improves efficiency and scalability, and reduces the hardware requirements of system implementation. Using such configuration, we can simply extend our system to other devices that contains a microphone and computation resources rather than limited to the smartphone platform.

3.2 Transmitter Waveform Design and Modulation

Using audible-band acoustic signal as beacon for ranging and communication, we must contend with a variety of noises to ensure its accurate ranging due to the highly populated frequency band below 20kHz. The ranging accuracy direct depends on the signal-to-noise ratio (SNR) and effective bandwidth. However, higher transmitter signal bandwidth and power will generate sound noise that disturbs to users. The standard microphone in mobile phone can only support bandwidth of 200Hz $\sim$ 20kHz. To minimize the sound noise while perform ranging, we choose 15kHz $\sim$ 20kHz as the operating band. The reason is that our ear is less sensitive to the high frequency signal, while the microphone in smartphone could still receive signal in this boundary band. With well controlled transmitter signal power, the acoustic beacon could be unnoticeable to humans while still detectable for smartphone. We use spread-spectrum and ultra low-duty-cycle pulse sharping techniques to ensure the proper operation of the acoustic beacon under realistic environment, as well as improving the ranging accuracy. Unlike the less sophisticated single-frequency acoustic signal used in Cricket and Beep, the beacon signal used in *Guoguo* is wide-band modulated and unnoticeable to humans.

The task of anchor nodes that transmitting modulated acoustic sharp pulse to the mobile receiver is to realize synchronization and ranging, as well as inform its unique pseudocode $\mathbf{p} = [p_{m,i}]$, $m = 1, \ldots, M$, $i = 1, \ldots, L$, where M is the number of the anchor nodes; L is the pseudocode length. The pseudocode $\mathbf{p}_{m,i} = 0, 1$ is also the information bits carried by the beacon signal. For the symbol waveform, we choose to use the second derivative of the Gaussian (Doublet) Pulse [24] and multiple it to the carrier wave. The waveform could be written as

$$g(t) = \frac{A}{\tau}\left[1 - 4\pi\left(\frac{t}{\tau}\right)^2\right]\exp\left[-2\pi\left(\frac{t}{\tau}\right)^2\right]\cos(2\pi f_c t) \quad (1)$$

where τ is the pulse width parameter, f_c is the carrier wave frequency. We truncate the Doublet pulse by their 3τ to approach the real-time condition.

The center frequency f_c of the modulated pulse is controlled by the on-chip timer and working at 18kHz. In addition to f_c, τ is tuned to ensure the effective bandwidth of (1) lies in between 15kHz $\sim$ 20kHz. The multiplication with the carrier wave in (1) is also a spread spectrum process that extend the initial narrow band f_c to wide band signal $g(t)$ by using the Gaussian Doublet Pulse sharping. The use of ultra low-duty cycle pulse (1) has the feature of higher data rate, higher location refresh rate, better multipath resolution, lower energy consumption and smaller sound pollution. To balance between the system complexity and the sophisticated modulation scheme, we choose to perform 2-PAM modulation with symbol duration as T_s; i.e., transmit sharp pulse $g(t)$ represents symbol '1', no pulse for symbol '0'.

3.3 Pseudocode Selection

One design challenge faced in selecting the pseudocode is to utilize the proper pseudo-codes with enough code distance redundancy to separate different anchor nodes, e.g., utilize orthogonal codes. In addition to communication, every symbol '1' in $\mathbf{p}$ will contribute to one ranging measurement, thus the high density bit '1' could improve the location update rate.

Conventional communication process relies on the preamble part of a frame to perform synchronization, then follows the data bits. For our designed anchor network, the beacon signal should only contain the pseudocodes and transmits cyclically to save round time for higher location update rate. The stringent requirement on efficiency making the remove of unnecessary part of bits especially important for Guoguo due to the low transmission speed of acoustic signal. However, most orthogonal pseudo-codes are not cyclic orthogonal; it will loss orthogonality due to the cyclical transmission. Using Walsh-Hadamard codes, for example, only $L+1$ codes with length L are orthogonal to each other in all phases among Hadamard Matrix $H(2^L, L)$. Moreover, the balanced '1' and '0' in these pseudo-codes does not comply with our requirement on high '1' density sequence.

In order to meet the special requirement of pseudocode for our anchor network, we select three distinct maximum sequence as $\{ms_1, ms_2, ms_3\}$ with length of 2^L, L is the pseudocode length. To increase the '1' density, we perform plus and decision process to combine these three m-sequences as

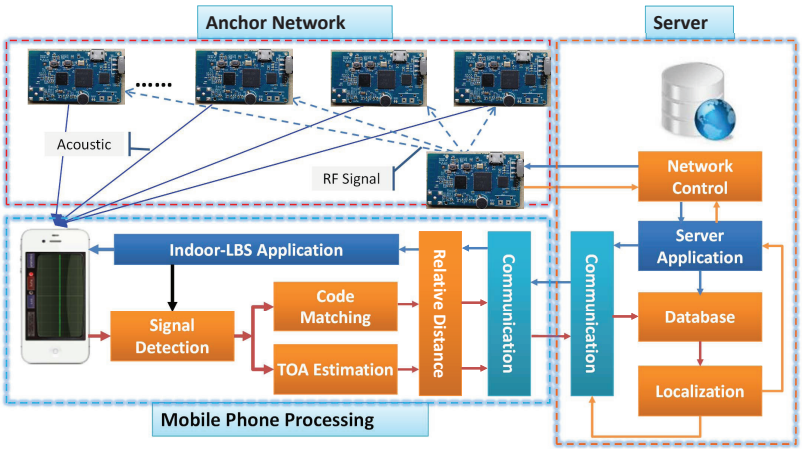


Figure 1: Conceptual Architecture of the Guoguo System.

a new sequence ms', as

$$ms' = \{ms_1 + ms_2 + ms_3\} \gtrless_0^1 0.5 \quad (2)$$

where the length of ms' is 2^L. Reshape ms' into a (L', L) matrix as $\mathbf{m} = \{m_{a,b}\}$, with $L' = 2^{(L-1)}/L$, $a = \{1, \ldots, L'\}$, $b = \{1, \ldots, L\}$. L is the code length, and total L' sequences in $\mathbf{m}_{a,b}$. To maintain the minimum code distance d_{tol} between each sequences, we only select sequences in $\mathbf{m}$ that satisfy the conditions in any phases that

$$p_{m,i} = \arg_{\hat{a}}\{p_{\hat{a},i} | \min ||m_{a,i} - p_{k,i+\triangle_i}||_1 \geq d_{tol}\}, \quad (3)$$
$$\forall k = \{1, \ldots, m-1\}, \forall \triangle_i = \{1, \ldots, L\}$$
$$\hat{a} \in \{1, \ldots, 2^{(L-1)}/L\}, m = m + 1$$

where $||\cdot||_1$ calculates the code distance between $m_{a,i}$ to the selected sequences $p_{k,i+\triangle_i}$, $i + \triangle_i$ performs cyclic phase change for the sequence. (3) filters sequences in $\mathbf{m}$ that meet minimum distance d_{tol} among selected sequence sets and add into $p_{m,i}$. When $L = 16$, among total $2^{(L-1)}/L = 2048$ sequences, only 12 sequences satisfy (3). These sequences can be assigned to M anchors, where $M = 9$ for a micro-cell in our system.

3.4 Anchor Network Protocol

The anchor network structure is shown in Fig. 1, with several anchor nodes and one controller. The controller provides basic timing via radio beacon passing for the whole network; and all the anchor nodes are synchronized to the controller by receiving the radio beacon passively and periodically. Such scheme can guarantee that the transmitted signal from anchor nodes are synchronized to a common timing source. There are two kind of beacons in Guoguo, the first kind of beacon is transmitted by the controller, which synchronizes all the anchor nodes to the controller's clock based on message passing; the second kind of beacon is transmitted by the anchor nodes from the acoustic speaker, which provide ranging and synchronization information to the smartphone.

The controller beacon is realized by using the existing radio chips, e.g., Zigbee. The data domain contains the anchor token, and current beacon index. The anchor node performs its own processing, i.e., transmit its own acoustic beacon, when the anchor token is received. Denote the controller beacon interval is T_p, if every anchor node executes the similar processing, the delay can be viewed as fixed and making the acoustic beacon timing equals to T_p.

Due to the reason that there is no commercial acoustic communication physical layer in mobile devices, we need to design our own processing to realize the communication capability. The ranging and synchronization capabilities of the anchor-smartphone pair are all based on the acoustic communication channel, while a good beacon structure can improve the overall throughput and data rate.

Denote the guard time between each beacon is T_g; the frame duration is $T_f = LT_s + T_g$. One beacon period of anchor can be written as $T_p := T_f = LT_s + T_g$. Using the Time Division Multiple Access (TDMA) to avoid the collision between each beacon, the total round period T_r equals to $T_r = MT_p = M(LT_s + T_g)$. The mobile phone can differentiate all the anchor nodes after T_r, thus the synchronization time between the anchor network and smartphone is $T_{syn} = M(LT_s + T_g)$. The localization process after synchronization also needs M beacons to obtain ranging information from all the anchors, i.e., update rate is $T_{up} = T_{syn} = M(LT_s + T_g)$. When $M = 9$, $L = 16$, $T_s = 0.0781s$, assume $T_g = T_s$, then $T_{up} = T_{syn} = 11.9493s$. Such a long time of synchronization and position update is caused by the low transmission speed of acoustic signal, and it is not sufficient fast enough for tracking the movement of humans. One possible way to increase the update rate is to lower the symbol duration T_s, with faster symbol transmission rate. However, symbol duration is restricted by the delay spread or coherence bandwidth of channel, and could not be further reduced without sacrificing the multipath resistance. To improve the speed of localization without relying on the compression of the symbol interval, a new beacon structure should be developed.

3.5 Symbol-Interleaved Beacon Structure

One possible solution is interleaving the L length symbols into the different beacon period; we call it *symbol-interleaved* acoustic beacon structure. Unlike the conventional TDMA that transmits the whole frame within a beacon period, i.e., $T_p := T_f = LT_s + T_g$, we divide the whole frame into symbols. By transmit one symbol in each beacon period without T_g, the beacon period T_p can be decreased to $T_p := T_s$.

With reduced T_p, the anchor network can achieve higher accuracy by receiving more radio beacons from the controller for every second. In the receiver side, the adjacent received symbols are from different anchors. The receiver maintains L length pipeline, and performs code matching in an iterative way. When the symbols in the pipeline matched with one sequence in $\mathbf{p}$, the anchor node can be identified. Using *symbol-interleaved* beacon structure, the initial synchronization time T_{syn} cannot be lowered, but the following *update time interval* can be reduced to $T_{up} = MT_s = 0.7029s$. The refresh of the location data for every 0.7029 second is sufficient for tracking slow-moving humans.

Using *symbol-interleaved* beacon structure, the transmitted beacon sequence from the anchor network in j-th period can be written as $p_{c(j)}(a(j), b(j))$, where $a(j) \equiv [j \mod M]$ is the index of the anchor node; $b(j) \equiv \lfloor j/L \rfloor$ is the index of the pseudocode in one frame; $c(j) \equiv \lfloor j/(ML) \rfloor$ is one round measurement.

Using *symbol-interleaved* beacon structure, the transmitted beacon from the anchor network can be modeled as

$$g_t(t) = \sqrt{\varepsilon} \sum_{j=0}^{N_s-1} p_{c(j)}(a(j), b(j)) \cdot g(t - jT_s) \tag{4}$$

$g(t)$ is the acoustic sharp pulse designed in (1); ε is the signal energy. The number of total transmitted symbol is N_s.

4. SMARTPHONE LOCALIZATION

4.1 Design Work flow

In the smartphone, the *Signal Detection and Demodulation* module performs audio signal recording and filtering, detect and extract the embedded beacon signal. The detection result is the digital symbols. *Code Matching* module matches the demodulated digital symbols to determine the anchor ID and its predefined location. The *TOA Estimation* module obtains pseudo-ranges by measuring the arrival time of the signal. *Relative Distance* module accumulates all the distance from different anchor nodes until sufficient number of measurements available for localization. *Localization* module performs location calculation by using the measured pseudo-distance pairs and available anchor positions. The technical details of all these modules are elaborated in the following subsections.

4.2 Symbol Detection and Demodulation

In the receiver side, i.e., the smartphone, we need to detect the signal and demodulate the information bits in the received signal $r(k)$. The received signal constitutes ξ_j multi-paths, and these multi-paths can be utilized to extract the symbol. Thus, the detection problem can be written as to detect the signal present or not in the j-th symbol. The received signal waveform in the j-th symbol period could be written as

$$r_j(k) = \sqrt{\varepsilon} \sum_{l=0}^{\xi_j - 1} A_j^l g_j(k - k_j^l) + n_j(k) \tag{5}$$

Assume the sampling rate is F_s, for symbol duration T_s, the total sampling point in one symbol is $M_o = T_s \times F_s$. For the low-duty-cycle pulse used in (1) as the symbol waveform, the actual signal length is shorter than the total symbol length. Assume the multi-path delay spread coefficient is α, the average sampling points for the signal region is $M_p = \alpha \times (3\tau) \times F_s$. Thus, the two conditions of the hypothesis for detecting the signal can be written as

$$\begin{cases} H_0: \ \mathbf{r}_{j,k} = \mathbf{n}_{j,k} \quad k = 1 \cdots M_o \\ H_1: \begin{cases} \mathbf{r}_{j,k} = \sqrt{\varepsilon}\mathbf{s}_{j,k} + \mathbf{n}_{j,k} \quad k = k_j^0 \cdots k_j^0 + M_p - 1 \\ \mathbf{r}_{j,k} = \mathbf{n}_{j,k} \quad k = 1 \cdots k_j^0 - 1, k_j^0 + M_p \cdots M_o \end{cases} \end{cases} \tag{6}$$

where $\mathbf{n}_{j,k}$ is the matrix form of the noise $n_j(k)$, $\mathbf{s}_{j,k}$ is the k-th sampling point for the signal in j-th symbol. The symbol synchronization process is to detect the signal region in the noise background, i.e., detect H_1 condition out of H_0, while the TOA estimation is to detect the first path of signal and its delay k_j^0.

To detect the signal region (H_1 condition), the decision vector $\mathbf{z}_j$ can be obtained by using generalized likelihood ratio test (GLRT) [10]. The decision vector could be derived as the form of $\mathbf{z}_{j,k} > \eta_{syn}$, with j-th symbol declared present if the inequity condition is satisfied and return an estimated value of the signal region k_p. The threshold η_{syn} is chosen to maintain a constant false alarm rate (CFAR) [10], and written as the form of $\beta\sigma$, where σ is the noise variance, β is calculated in experiment by using the given false alarm rate. Then $\hat{p}_{c(j)}(a(j), b(j))$ in the receiver will be set to '1', otherwise $\hat{p}_{c(j)}(a(j), b(j)) = 0$, where $c \equiv \lfloor j/(ML) \rfloor$. $\hat{p}_{c(j)}(a(j), b(j))$ is one estimated version of $p(a(j), b(j))$, c is one round measurement.

4.3 TOA Estimation

After detected the symbols in Section 4.2, more detailed time-of-arrival (TOA) estimation should be performed to estimate the first path sample k_j^0 in the whole symbol duration. The TOA estimation provides ranging information that needed for localization, and its accuracy directly affects the overall position resolution. The TOA estimation problem can be written as to detect k_j^0 in j-th symbol of (5) as

$$\hat{k}_j^0 = \min_k (k | r_j(k) > \eta_{toa}), k \in [k_p - J_p, K_p + M_o] \tag{7}$$

where k_p is the obtained rough signal region during signal detection; J_p is the step length used for Jump-Back-and-Search-Forward (JBSF) [8]; η_{toa} is the TOA estimation threshold; $\hat{k}_j^0$ is the estimated TOA path of k_j^0. The challenges of dynamically change the TOA estimation threshold η_{toa} to balance between the false alarm and miss detection is addressed in [15] by maximizing the *TOA Detection Probability*.

4.4 Synchronization and Code Matching

Perform code matching between predefined pseudo-codes $p_{m,i}$ and the estimated information bits $\hat{p}_{c(j)}(a(j), b(j))$, the m-th anchor node can be identified if these two sequences matched. When symbols with total $M \times L$ length have been received, the code matching process can be utilized to synchronize between the anchor node and the smartphone by

$$[\Delta_a, \Delta_b] = \arg\min_{\Delta_a, \Delta_b} ||\hat{p}_{c(j)}(a(\hat{j}) + \Delta_a, b(\hat{j}) + \Delta_b) - p_{m,i}||_1 < d_{tol} \tag{8}$$

where the beacon period index $j \in [j_0, j_0 + ML]$, $j_0 = c(j) \times (ML)$ is the starting index of the symbol sequence, ML is the total length of symbols that used to perform (8), $\hat{j} =$

$j - j_0$. $c(j)$ can be used to illustrate the index number of code matching. $a(j)+\Delta_a$ and $b(j)+\Delta_b$ is the cyclic shifting in $\hat{p}_{c(j)}(\cdot)$.

With the offsets Δ_a and Δ_b available for the anchor node index and pseudocode sequence index, the mobile phone can aware the the anchor node index ($\hat{m}$) and sequence index ($\hat{i}$) of the current received symbol j as

$$\begin{aligned} \hat{m} &= [(j - j_0) \mod M] + \Delta_a \\ \hat{i} &= [(j - j_0) \mod L] + \Delta_b \end{aligned} \tag{9}$$

where $j_0 = \lfloor j/ML \rfloor \times (ML)$.

4.5 Distance Update

After the synchronization in Section 4.4, and obtained $\hat{m}$ and $\hat{i}$ in (9), then every M symbols can obtain one distance measurement group with the same pseudo sequence index $\hat{i}$. Such group of measurements is the minimum tuple of ranging for one position update, and every measurement in such a group is from different anchor nodes. For j-th symbol, the index of group is $j_g = \lfloor j/M \rfloor$ with each element represents the TOA value obtained in (7). Denote the TOA estimation matrix as $\mathbf{r} = r_{m,j_g}$, $m = [1, M]$. For notation convenience, one group of measurements from all anchor nodes can be represented as $\mathbf{r}_g = r_{m,j_g}$, where j_g is a fixed value, $\mathbf{r}_g$ is a measurement vector.

Due to some symbol-miss in real situations and none distance measure during '0' symbols, the TOA value is not fully available for all M anchor nodes in one round time, i.e., the obtained ranging value $\mathbf{r}_g$ is a sparse vector. To improve the reliability of ranging results, we perform sliding window for the TOA estimation matrix $\mathbf{r}$ with an appropriate length of W, and obtain a new version of $\mathbf{r}_g$. The rational of the sliding window is to use the historical data to represent current sparse measurement. However, different symbol index corresponds to its own starting index of the TOA measurement, e.g., the TOA value in j-th symbol is larger than $(j-1)$-th TOA value by N_k. Thus, for the extracted length of W and width of M matrix, i.e., the latest W column of the matrix $\mathbf{r}$, $\mathbf{r}_g = [r_g(m)]$ can be calculated by

$$r_g(m) = \frac{1}{N_d(m)} \sum_{\Delta_g=0}^{W-1} \left[r_{m,(j_g - \Delta_g)} + \Delta_g \times N_k \right] \tag{10}$$

where $r_{m,(j_g-\Delta_g)}$ only counts when it contains TOA measurement, $N_d(m)$ is the total number of effective measurements available in the W length matrix $\mathbf{r}$, m represents the m-th row. Thus, vector $r_g(m)$ can be used as the current distance measurement, with each TOA value indicates the pseudo-range from the m-th anchor to the mobile device.

The measured pseudo-ranges between the anchor nodes to the mobile phone is $\hat{r}_m = \mathcal{C} \times r_g(m)$, where $\mathcal{C}$ is the speed of the aerial acoustic signal. Denote the true distances are r_m, $m = [1, \ldots, M]$, the unknown starting point is δ_r, $\hat{r}_m$ can be written as

$$\hat{r}_m = r_m + \delta_r + n_m \tag{11}$$

where n_m is the distance measurement noise for $\hat{r}_m$.

4.6 Location Estimation

With the measured distance from M anchor nodes available, multilateration can be performed to localize the smartphone. Assume the real position of the smartphone is $\mathbf{p} = (x, y)$, the position of the anchor node is $\mathbf{p}_m = (x_m, y_m)$, then the real distance from the smartphone to the anchor node m can be written as $r_m = \sqrt{(x - x_m)^2 + (y - y_m)^2}$. We define the vector of unknown parameter as $\theta = [x \quad y \quad \delta_r]^T$. The localization purpose is to obtain estimated value of θ from observations, where $[\hat{x}, \hat{y}]$ in $\hat{\theta}$ is the estimated position.

However, (11) contains the nonlinear term, rather than estimate θ directly, we select one of the M measurements as the reference f, and using this reference to cancel out the nonlinear term. The selection of this reference could be random or based on the code-matching quality (the mis-match residues). To better illustrate this process, we can square the both sides of (11) and minus the reference measurement f as

$$\begin{aligned} &[(x_m^2 + y_m^2 - \hat{r}_m^2) - (x_f^2 + y_f^2 - \hat{r}_f^2)] + 2\hat{r}_m\widetilde{n}_m - 2\delta_r n_m \\ &= 2(x_m - x_f)x + 2(y_m - y_f)y - 2\delta_r(\hat{r}_m - \hat{r}_f) \end{aligned} \tag{12}$$

where $\widetilde{n}_m$ is differential noise term of $n_m - n_f$. Equation (12) can be expressed in matrix form as

$$\mathbf{A}\theta = \nu + \mathbf{p_n} \tag{13}$$

where $\mathbf{A} = \begin{bmatrix} x_m - x_f & y_m - y_f & \hat{r}_f - \hat{r}_m \\ \cdots & & \\ \cdots & & \cdots \end{bmatrix}$, $\theta = \begin{bmatrix} x \\ y \\ \delta_r \end{bmatrix}$ and $\nu = \frac{1}{2}[(x_m^2 + y_m^2 - \hat{r}_m^2) - (x_f^2 + y_f^2 - \hat{r}_f^2)]_{(M-1)\times 3}$, $m = 1, \ldots, M$ when $m \neq f$. $\hat{r}_m$ is the measured distance obtained from (11). $\mathbf{p_n} = 2\hat{r}_m\widetilde{n}_m - 2\delta_r n_m$ is the noise term with its first moment as $\mathrm{E}(\mathbf{p_n}) = 0$ and the covariance matrix of vector $\mathbf{p_n}$ as $\widetilde{\mathbf{\Sigma}} = \mathrm{Cov}(\mathbf{p_n})$. The diagonal element of the covariance matrix are $2\hat{r}_m(\sigma_m^2 + \sigma_f^2) + 2\delta_r\sigma_m^2$, with other elements in the matrix as $2\hat{r}_m\sigma_f^2$.

Then (13) can be formulated as a Least-Square (LS) problem, and we can obtain the solution as

$$\hat{\theta} = (\mathbf{A}^T\widetilde{\mathbf{\Sigma}}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\widetilde{\mathbf{\Sigma}}^{-1}\nu \tag{14}$$

Initially, the value of the covariance matrix $\widetilde{\mathbf{\Sigma}}$ is unknown, we could initialize $\widetilde{\mathbf{\Sigma}}$ with all '1' in its diagonal and '0' for other elements. For the following snapshots, $\widetilde{\mathbf{\Sigma}}$ could be estimated, and we can substitute $\widetilde{\mathbf{\Sigma}}$ into (14) to improve estimation accuracy.

4.7 Error Pruning Techniques

Error pruning techniques are essential for the correct operation of the indoor localization system. The source of potential errors could be the sound noise, blockage, indoor Non-light-of-sight (NLOS) transmission, anchor network failure. In real system, applying too strong criterion in error pruning would lower the location update rate, thus the parameters used in the following techniques should balance between the accuracy and location update rate.

Signal Level Resistance. To deal with the signal noise issues, we perform spread spectrum in transmission and correlation processing in the receiver to obtain noise resistance. We use the the Gaussian second derivative pulse as the waveform, which has the properties of high multipath and timing resolution, low energy consumption, robust to noise, and low spectrum emission.

NLOS Identification and Mitigation. Compared with the noise, indoor harsh environment with blockage and high-dense multi-paths is more challenge. The blocked localiza-

tion beacon could either be attenuated or introduces additional delay that prolonged the obtained ranging distance, we call it NLOS bias effects. Using these problematic ranging results could cause over-fitting in localization calculation, or even make the final results diverge. Thus, how to identify and mitigate these prolonged or problematic ranging measurement is crucial in localization.

One of the superiority of using acoustic signal for ranging lies its high timing resolution and the full access of the acoustic channel information. Extract features from estimated channel condition, the goodness of transmission could be evaluated. For example, if the delay spread of the estimated channel is significant larger than normal conditions, the ranging result from this channel has high probability of blockage, i.e., NLOS condition. By lowering the weighting efficient of these ranging measurements during localization, overall location result could be more stable. We use the combined metrics of RMS delay spread, Kurtosis, and Mean excess delay to identify the NLOS channel condition, and assigns lower weight for the measurements from NLOS channel during localization processing.

Anchor Level Error Mitigation In subsection 4.4, the demodulated pseudocode is matched to identify the anchor node. In real situations, these exists some bit-error for our implemented acoustic communication physical layer. When designing the pseudocode, some extent of redundancy is embedded. The code matching process has some tolerance of d_{tol} for the bit-error. However, if the bit-error for one anchor node is higher than the design redundancy, then the measurements from this anchor node should be mitigated in this round. This process could eliminate some extent of large errors.

5. PERFORMANCE EVALUATION

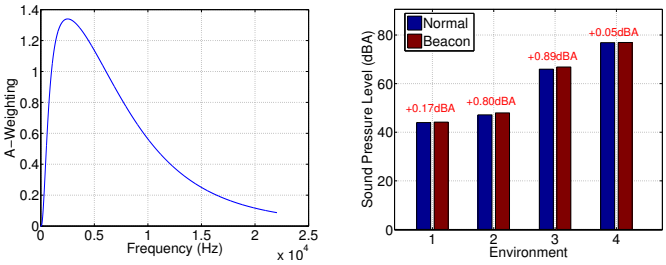
In this section, we perform system performance evaluation by using Apple's iPhone 4S and iPod Touch 5 (iTouch5) without any modification of the hardware or jailbreak of the operating system. First, we provide quantified results of the sound pressure level of our acoustic beacon in subsection 5.1. Second, we evaluate the localization performance of our proposed system in two typical environments: office (Subsection 5.2) and classroom (Subsection 5.3).

The 3dB pulse width of the transmitted signal as in (1) is chosen with $T_w = 1.5ms$ to meet the bandwidth constraint. A sharper pulse results in better multipath robustness and time-domain resolution, but with increased bandwidth occupancy. Restricted by the bandwidth of a microphone, the effective pulse energy and operating distance would be decreased when more frequency components outside the receiver band. The symbol duration is chosen as $T_s = 0.0781s$, resulting in the pulse duty cycle of $R = 1.92\%$. Shorter symbol duration leads to a higher data rate and location refreshing rate, but restricted by the multipath environment. The choice of these parameters is a tradeoff between the achievable resolution and maximum operating distance.

5.1 Sound Pressure Level Measurement

Guoguo system uses audible-band signal as beacon, making it user-friendly and widely applicable for many indoor location services and applications. However, users might be concerned about the noise effect of the acoustic signals in indoor environments.

To make our transmitted acoustic signal unperceptible to users, we tuned the waveform of the beacon signal with appropriate transmission power. To quantify the results, from the perception curve of human ears (*A-weighting* coefficients, we derive the *sound pressure level* (SPL) value as a metric of noisiness of the environment. We place four anchor nodes very close to a smartphone (the most noisy case), and measure the difference of SPL. The normal sound background case is named ("Normal"); the environment that filled with our beacon signal is called ("Beacon"). The SPL values for these two conditions in four different environments are shown in Fig. 2b. From Fig. 2b, we observe that the difference in average SPL level of the "Normal" and "Beacon" is under the level of randomness: different measurement of the SPL in different time has some sort of randomness (near 2dBA). The maximum difference of the SPL caused by the "Beacon" is 0.85dBA, which is below the perception level of human's ear. Therefore, we can conclude that our transmitted acoustic beacon signal is completely ignorable and *disturbance-free* even in extreme cases.



(a) A-weighting curve (b) Sound Pressure Level

Figure 2: (a) The perception level of human ears; (b) comparison of sound pressure level in four scenarios.

5.2 Exp I: Office Environment

Experiment Setup. We deploy 9 anchor nodes in a typical office environment to evaluate the performance of Guoguo as shown in Fig. 3. The maximum distance between anchor nodes is larger than 10 meters.

Total 12 cases have been tested: 6 of them are tested under quite environment for theoretical performance; other 6 cases are tested under simulated realistic environment (simulate human talking by playing video sound near the mobile device). The experiment configuration is summarized in Table 1. The term *Length* represents the measurement lasting time in seconds, which is randomly chosen. The last term is the effective number of accessed anchor nodes (N_{eff}) during localization calculation. The number of real accessed anchor nodes is less than total number of deployed anchors. This is often caused by interference, blockage, and NLOS during localization process.

Localization Accuracy in Quite Environment. The scatters of the localization results for users in different locations in a normal office environment are shown in Fig. 4. In this case study, to support quantitative analysis, we localize a smartphone when its user stands still at different spots. We can get the error surface of localization from Fig. 4. Note that, our system is not limited to localize a standing subject. Localization traces of a moving subject are demonstrated in the next section.

Figure 3: Experiment setup in an office environment

Table 1: Experiment configuration in an office environments

ID	Use Cases	Length (s)	Background	N_{eff}
1	Env1	60.91	Quite	6.86
2	Env2	60.97	Quite	6.15
3	Env3	60.61	Quite	8
4	Env4	60.64	Quite	7.77
5	Env5	60.53	Quite	8.24
6	Env6	60.64	Quite	8.72
7	P4snoise1	522.97	Sound	5.03
8	P4snoise2	300.48	Sound	6.11
9	P4snoise3	346.35	Sound	5.99
10	Touchnoise1	398.79	Sound	5.97
11	Touchnoise2	461.03	Sound	6.13
12	Touchnoise3	582.14	Sound	6.07

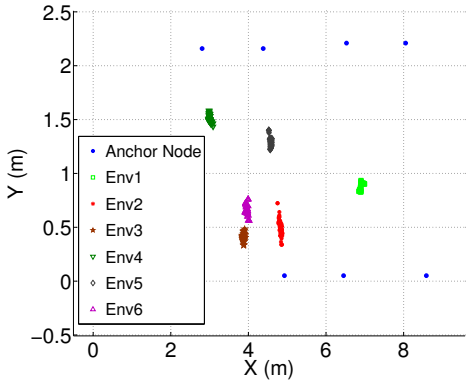


Figure 4: Scatters of the localization results in an office environment.

The cumulative distribution function (CDF) of the localization error (LE) are shown in Fig. 5. If using 80% probability, the localization error for all the cases is in the range between $4cm$ to $8cm$. These localization results are very accurate and sufficient for fine-grained indoor location-based services.

Localization Accuracy under Background Sound. To evaluate the localization performance in realistic environment, we add artificial sound by playing high volume videos. The CDFs of the added sound are shown in Fig. 6. The measured sound pressure level (dBA) is calculated from the received acoustic signal at a smartphone during localization.

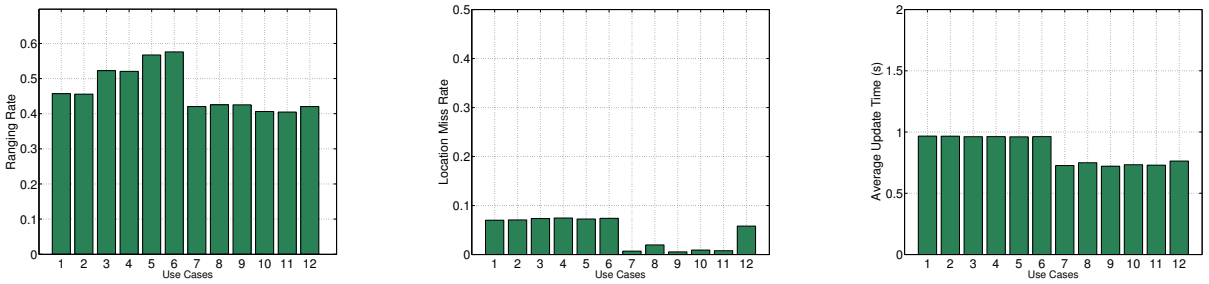
Fig. 7 shows the CDF of the localization error (LE) in an office environment. In these six cases with background sound, the final localization accuracy is still within $10cm$, with less than 10% results slightly affected. From another point of view, the number of effective accessed anchors under background sound is slightly lower than the normal cases from the N_{eff} value in Table 1 because some of the acoustic beacons have been mitigated. Since we use more anchors than the minimum requirement (three anchors), these additional anchors can improve our system's availability under realistic environment.

Localization Metrics. Other metrics that used to evaluate the system performance are the *Ranging Rate*, localization *Miss Rate*, and location *Average Update Time*. The definition of the *Ranging Rate* is N_{TOA}/N_s, where N_{TOA} is the total Number of TOA measurements, N_s is the total number of symbols transmitted. Higher *Ranging Rate* means better ranging detection probability. The maximum theoretical value of *Ranging Rate* depends on the '1' density of the pseudo-code rather than 1. From Fig. 8a, we know that the *Ranging Rate* is larger than 0.5 for most cases, which is contributed by using our derived high '1' density pseudo-codes.

The localization *Miss Rate* evaluates the quality of obtained location values. The definition of *Miss Rate* is N_{loc}/N_{pos}, where N_{loc} is the number of obtained location results; N_{pos} is the number of refined location results after the post-processing module. Lower *Miss Rate* means better localization results. The *Miss Rates* for all the cases in Fig. 8b show very small value, i.e., the localization results have a very good quality.

Another important metric is the *Average Update Time*, representing the refreshing rate of the localization process. If the refreshing rate is too slow, it is hard to keep up with the moving traces of subjects. Due to the low transmission speed of acoustic signal, minimizing *Average Update Time* is nontrivial. From Fig. 8c, we observe that the *Average Update Time* for Guoguo is less than one second, sufficient for capturing the traces of moving subjects at modest speed. This rate is significantly faster than other existing approaches that also use acoustic signal for localization.

Overall Localization Accuracy. Using 50-percentile, 80-percentile, and 90-percentile probability to evaluate the localization accuracy, the results of all 12 cases are shown in Fig. 9. The achieved localization accuracy is in the centimeter-level even for the cases with added interference. Such results could push the current coarse-level LBS into fine-grained level in an effective way.



(a) Ranging Rate (b) Miss Rate (c) Average Update Time

Figure 8: Measured performance metrics of Ranging Rate (a), Miss Rate (b), and Average Update Time (c) for different measurements in an office environment.

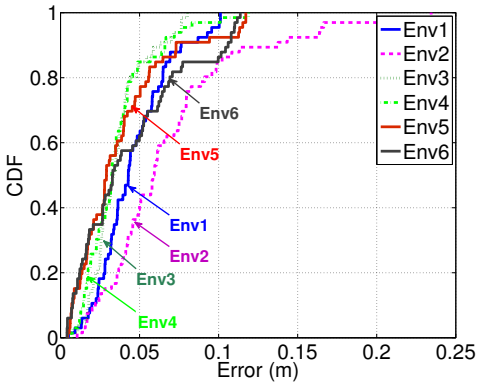


Figure 5: CDF of the localization results in an office environment.

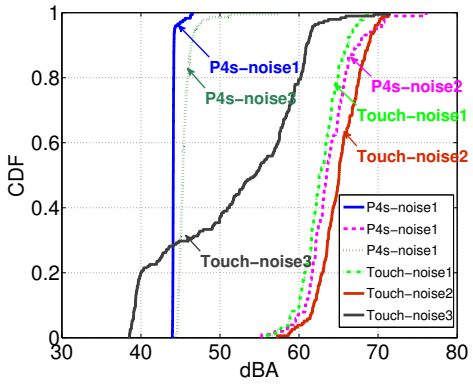


Figure 6: CDF of the added background sound level in an office environment.

5.3 Exp 2: Classroom Environment

Experiment Setup. Similar to subsection 5.2, we deploy 9 anchor nodes in a multimedia classroom environment to evaluate the performance of Guoguo as shown in Fig. 10, with maximum distance larger than 15 meters.

The experiment configuration in this classroom environment is summarized in Table 2.

Table 2: Experiment configuration in a classroom

ID	Use Cases	Length (s)	Background	N_{eff}
1	Env1	590.43	Quite	6.98
2	Env2	408.70	Quite	6.80
3	Env3	459.76	Quite	7.0
4	Env4	676.39	Quite	6.99
5	Env5	270.49	Quite	5.9
6	Env6	599.47	Quite	7.0
7	P4snoise1	279.38	Sound	6.94
8	P4snoise2	641.74	Sound	7.96
9	P4snoise3	240.40	Sound	7.72
10	Touchnoise1	599	Sound	7.0
11	Touchnoise2	88.79	Sound	5.38
12	Touchnoise3	347.5	Sound	5.0

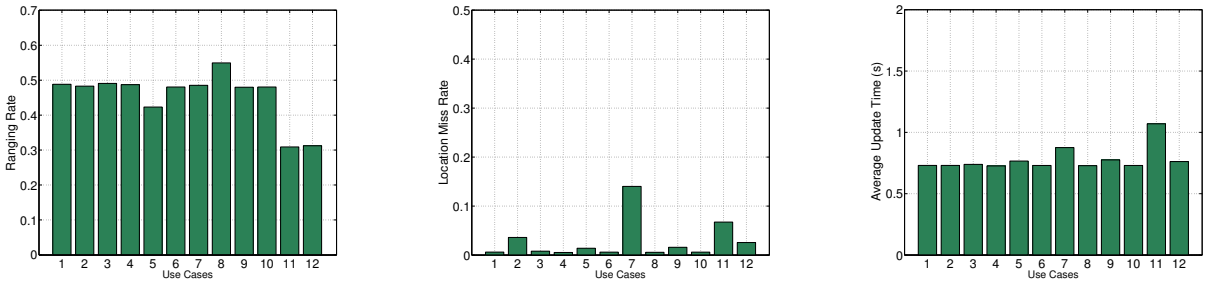
Localization Accuracy under Quite Environment. The scatters of the localization results for users in different locations in a classroom environment are shown in Fig. 11. Similar to Fig. 4, the error surface of achieved location results is very small.

The CDF of the localization results for the measurements in Fig. 11 is shown in Fig. 12. For most of the cases, e.g., using 80% probability, the localization accuracy is very high and in the range from $5cm$ to $10cm$.

Localization Accuracy under Background Sound. To evaluate the localization performance under interference, we added artificial sound in the classroom environment by playing the lecture videos. The CDF of the added different background sound is shown in Fig. 13. The CDF of the localization results under background sound on classroom environment is shown in Fig. 14. One interesting observation from Fig. 14 is that the iPhone4s achieves much better performance than the iTouch5. The reason could be the build-in multi-microphones and noise-canceling mechanisms in iPhone4s; while the iTouch5 is only equipped with one microphone and no noise-canceling hardware.

Localization Metrics. The *Ranging Rate* for the measurements in a classroom environment is shown in Fig. 15a, while two cases of iTouch5 suffers from low *Ranging Rate* under background sound. The *Miss Rates* in Fig. 8b shows very small value. The *Average Update Time* for the environment of classroom is near 0.8 second, and sufficient for mobility cases.

Figure 10: The experiment setup in a classroom environment



(a) Ranging Rate (b) Miss Rate (c) Average Update Time

Figure 15: Measured performance metrics of Ranging Rate (a), Miss Rate (b), and Average Update Time (c) for different measurements in a classroom environment.

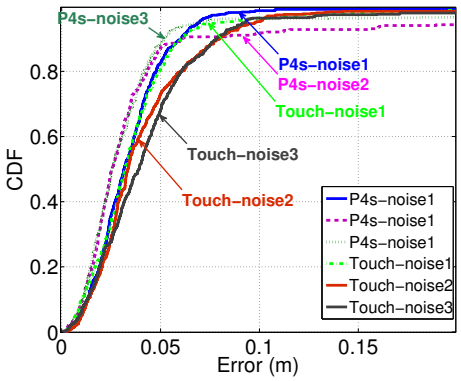


Figure 7: Localization accuracy (LE) in an office.

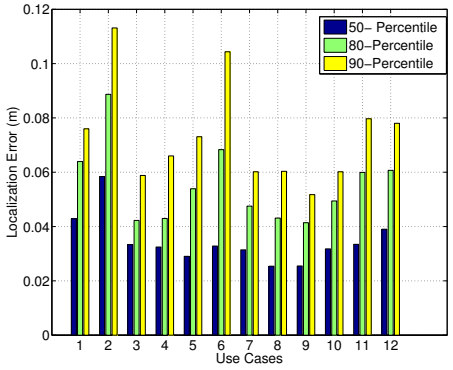


Figure 9: Localization error for different measurements in an office environment

Overall Localization Accuracy. The 50-percentile, 80-percentile, and 90-percentile localization accuracy results are summarized in Fig. 9. Except the iTouch5 under background sound, other cases achieves localization accuracy under 10*cm* for most cases. The localization results of iPod Touch5 are slightly worser than iPhone4s, but still sufficient for indoor location-based services.

6. PUTTING IT ALL TOGETHER

Our proposed smartphone-based indoor localization system achieves centimeter localization accuracy by leveraging the anchor network and processing on smartphone. To make the real system applicable and suitable for real deployment, significant efforts should be made to design the anchor network and implement the algorithm into the smartphone.

Anchor Node Design. Targeting at fine-grained indoor location-based services, e.g., retail store, museum, office and classroom, several hundreds investment on the anchor infrastructure to enable centimeter-level location related services could be profitable. To minimize the infrastructure cost and promote the real system deployment, we designed our own hardware for the anchor node with low-complexity and additional features, e.g., solar powered, battery charging, pluggable sensor board, remote program upload, wireless speaker/microphone. The comparison of our "Guoguo" hardware vs. two quarters is shown in Fig. 17. The total BOM price for the anchor node is less than $10. With one anchor every 10-20 meters deployed in an indoor space (e.g. museum, shopping center) and minimum of four anchors,

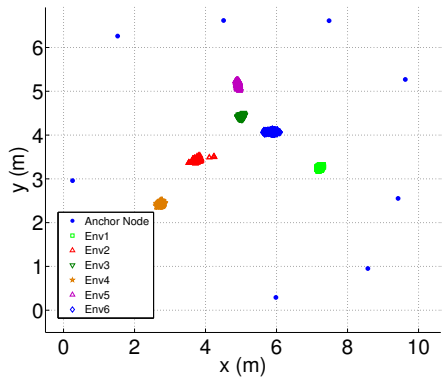


Figure 11: Scatters of the localization results in a classroom environment.

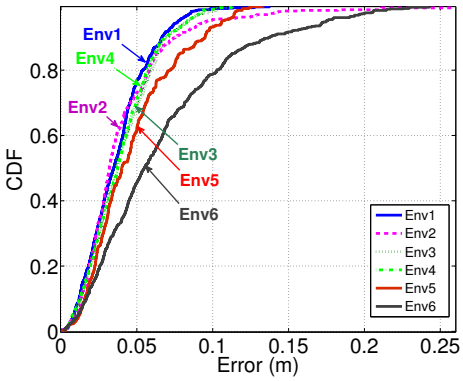


Figure 12: CDF of the localization results in a classroom environment.

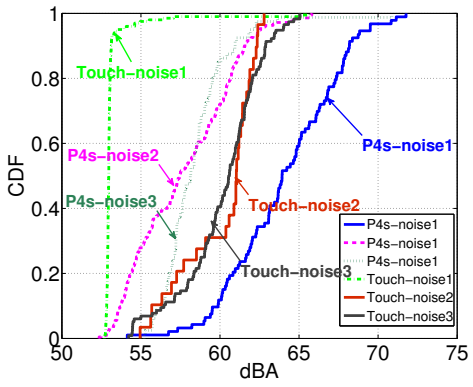


Figure 13: CDF of the background sound in a classroom environment.

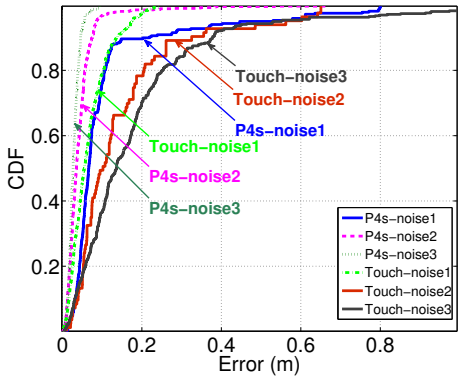


Figure 14: Localization error of the background sound in a classroom environment.

it is feasible and cost effective to deploy such systems in many indoor environments. For example, with marginal investment on the indoor infrastructure, a retail store can enable "smart" shopping experiences for smartphone users with fine-grained shelf-to-shelf navigation, location-aware recommendation and advising, and physical items searching and navigation.

Smartphone Processing. We implemented the algorithms from audio recording and pre-processing, signal detection, TOA estimation, code matching, and NLOS mitigation into the smartphone app. All the processing was put in the iOS Grand Central Dispatch (GCD) queue to enable concurrency with the mobile application without slowing down the smartphone's responsiveness. To ensure efficiency of the smartphone processing, we use the vDSP portion of the Accelerate framework in iOS. Other complex computation was executed in the server to minimize the computation cost in the smartphone. Once the smartphone publishes a ranging result to the server, the process on the server side performs localization and returns the result to the smartphone asynchronously. Such configuration balanced the communication and computation cost: the smartphone extracts ranging information from the audio raw data and greatly reduces communication overheads; the server handles concurrency and serves multiple localization requests from smartphones.

Moving Traces Demo. To demonstrate the performance of "Guoguo" in real situations, three typical moving traces captured by the Guoguo system are shown in Fig. 18. The moving traces demonstrate the localization accuracy and feasible update rate of our system.

7. RELATED WORK

Localization schemes can be classified into three categories: angle-based, fingerprinting-based, and ranging-based. An angle-based approach relies on the directional antenna scan to achieve angle resolution. But a narrow-beam directional antenna is very expensive and unsuitable for consumer applications. Fingerprinting-based approaches [3, 5, 2, 4] or other proximity approaches [20, 9, 6, 1] feature lowest complexity without any requirement on additional infrastructure. However, they achieve only room-level resolution and require the site-survey to build fingerprinting databases at the cost of intensive labors. Ranging-based approach is more straightforward. Measuring the radio signal strength (RSS) and time-of-arrival (TOA) are the two typical ranging solutions. Compared with TOA, RSS information is widely available and lots of work are dedicated to use the RSS of WiFi, bluetooth, or cellular signal, for indoor localization [27, 7, 11, 13]. However, the need of the prior information of the radio attenuation model, and the time varying

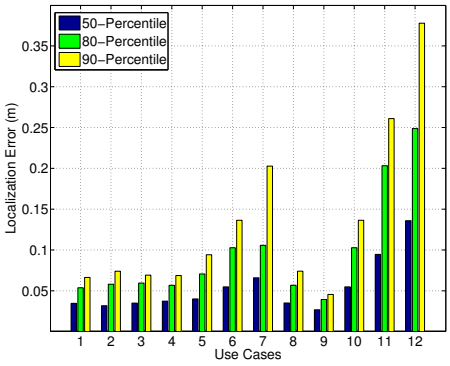


Figure 16: Localization error for different measurements in a classroom environment

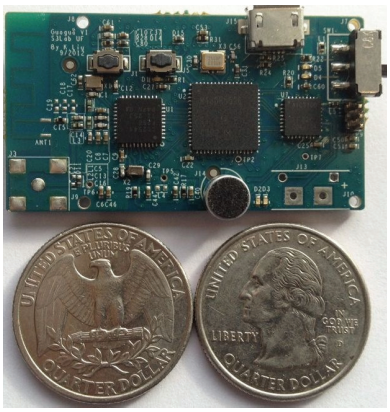


Figure 17: Guoguo anchor node vs. quarters

features of channel make these approaches only suitable for applications with low location accuracy needs.

TOA-based ranging is more accurate and robust. The Cramér-Rao Low Bound (CRLB) of TOA ranging is inversely propositional to the effective bandwidth of signal, it is also why Ultra-wideband (UWB) signal received special attention for its high accuracy on ranging and localization. However, current UWB techniques are still under development and not available on current smartphones. Another option is to use the acoustic signal for TOA ranging due to its low transmission speed. MIT Cricket[23], Active Bat [26] and DOLPHIN [18] are well-known systems that use ultrasound for localization. However, normal smartphones cannot receive ultrasound. In addition, they require radio signal for synchronization, e.g., ZigBee for Cricket. Dependence on special devices and non-applicability on smartphones greatly limits their adoption in daily indoor activities.

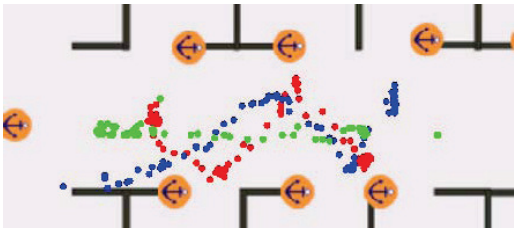


Figure 18: Moving traces obtained by indoor localization

Recent research on leveraging the ubiquitous microphone sensors in a smartphone introduces a convenient and low-complexity approach. A. Mandal et al. [17] used a PDA to transmit noticeable 4kHz acoustic signal. C. Peng et al. [22] proposed to transmit low-attenuation 2-6kHz acoustic signal for better coverage, but their solution causes sound pollution due to the audible signal. H. Liu et al. [14] used a smartphone to transmit high-band audible sound to assist WiFi localization. Authors in [19] performed localization using desktop PCs and laptops, achieving location accuracy of several meters. In addition, these solutions are not scalable and feature low location update rates. Due to the use of two-way ranging and simple acoustic "Beep" signal, only one user can be handled and localized at a time. Adopting time-divided coordination among users could be a partial solution. However, random concurrent access patterns of many users in real situations make their solutions hard to support multiple users.

To compare the performance of Guoguo with other existing indoor localization approaches, we summarize their key features in terms of principle, accuracy, cost, scalability, pros and cons in Table 3. The scalability refers to how many end users can be supported simultaneously. From Table 3, we observe that Guoguo achieves better balance in terms of cost and performance over other acoustic/ultrasound based approaches. Ranging in the passive receiving mode in a Guoguo-enabled smartphone makes it highly scalable while other ranging methods in the active transmitting mode limits the scalability due to interference of signals. Guoguo only requires a smartphone for a user, along with low-cost anchor nodes preconfigured by a service provider. Compared with other low cost approaches (mainly smartphone-based), e.g., RSS and fingerprinting based approach, Guoguo significantly outperforms other approaches in accuracy. Existing smartphone-based approaches achieve only room-level or several-meter-level accuracy. As far as we know, Guoguo outstands as a practical smartphone-based solution with foot-level accuracy without special hardware extensions on users' phones.

8. CONCLUSION

We proposed the Guoguo algorithm and ecosystem to realize the smartphone-based fined-grained indoor localization. For the first time, we can locate a smartphone user at the centimeter-level, which has significant implication for potential indoor location services and applications compared with existing meter-level localization solutions. To address the challenges of utilizing the audible-band acoustic signal in smartphone localization, i.e., strong attenuation, interference-rich, high sound disturbance and difficulty in synchronization, we proposed comprehensive schemes to improve the localization accuracy and extend coverage without sound disturbance. Significant improvements were achieved in terms of accuracy, cost, and scalability, compared with other existing approaches. Experimental results demonstrated that the achieved average localization accuracy is about 6 $\sim$ 25cm in typical office and classroom environments. Guoguo represents a leap progress in smartphone-based indoor localizations, opening enormous new opportunities for indoor location-based services, positioning and navigation systems, and other commercial, educational, or entertainment appli-

Table 3: Comparison of Guoguo with existing localization techniques

System	Signal Type	Technique	Accuracy	Cost	Scalability	Pros	Cons
Guoguo	Modulated acoustic signal (17-20kHz)	Passive-mode ranging	High (6 $\sim$ 25cm)	Low	High	No dedicated device for users, high accuracy and scalability, disturbance-free	Low-complexity anchor nodes
Cricket[23]	Ultrasound, radio-assistant	Ranging based	High (10 cm)	Medium	High	High accuracy	Need dedicated devices
Beep [17]	Acoustic signal (4kHz)	Active-mode ranging	High (60cm)	Medium	Limited	High accuracy	noticeable acoustic signal, complex anchors, limited users
Active Bat [26], DOLPHIN [18]	Ultrasound	Active-mode ranging	High (centimeter)	Medium	Limited	High accuracy	Need dedicated devices, need dense anchor nodes, limited scalability
Ultra-wideband [12, 16]	Impulse-radio Ultra-wideband signal	Ranging based	High	High	High	High penetrating ability, high accuracy, high multi-path resolution	Need dedicated devices, expensive devices
Radio RSS [7, 11, 13, 27]	WiFi, bluetooth, or cellular signal	Ranging based	Low	Lowest	High	No dedicated devices, existing infrastructure, low cost	Need a prior of the radio attenuation model, low accuracy, need calibration
Fingerprinting [2, 3, 4, 5]	WiFi, Cellular, Bluetooth, FM and ambient sound	Fingerprint matching	Low (room-level)	Lowest	High	No need of additional infrastructure, low cost	Need war-driving, time-varying, low accuracy
Proximity [1, 6, 9, 20]	RFID, NFC and acoustic signal	Proximity based	Low	Medium	High	Low cost sensor nodes, support multi-users	Low accuracy, and needs dense sensors

cations. We will build Guoguo-enabled indoor LBS applications, and plan to deploy such ILBS applications in museums and libraries, and further push the envelope of high-resolution indoor localization.

9. FUTURE WORK

Anchor Coverage and Maximum Operation Distance. In this paper, we only demonstrated that the Guoguo works in two typical indoor environments. A more comprehensive study regarding the maximum operation distance of the anchor and the coverage of the anchor network should be conducted. In real environments with blockages, e.g., big shelves or human walking, the effective coverage of the anchor network would be much smaller. Different schemes of anchor node placement, and the guideline for real deployment should be developed for better coverage and easy installation.

Optimized Localization Approach. The localization algorithm used in Guoguo is a normal least squares (LS) based approach. When the pseudorange measurements are non-convex, LS approach would converge to local optimal values. In the future work, we will apply global location optimization approaches, e.g., Semidefinite Programming, to jointly estimate the location and delay (δ_r), and avoid the calculation of the matrix inversion operation in (14).

Interference-robust Scheme. The inherent interference in the crowded audible-band would impede the appropriate operation of Guoguo. Ambient noise causes the in-band interference, and has much large bandwidth and energy than the acoustic beacon of Guoguo. In addition to the approaches proposed in this paper for signal detection, TOA ranging and localization, sophisticated error and interference mitigation schemes should be developed to improve the robustness.

Human and Animal Exposure. In this paper, we made the beacon unnoticeable to human to minimize the potential healthy risk. For the effect caused by long term exposure of the acoustic beacon, related professional studies should be cited. Moreover, we are not sure whether some animal species can actually hear the beacon sound. Further studies should be conducted to analyze the reaction of the animal or pets when exposed to the acoustic beacon. It is very important to make the system pet friendly.

10. ACKNOWLEDGMENTS

We would like to thank our reviewers and our shepherd Anthony LaMarca for all of their great input. The work presented in this paper is supported in part by National Science Foundation under Grant No. CNS-0709329, CNS-0923238, CCF-0953371, CCF-1128805, and CNS-0940805 (BBN subcontract).

11. REFERENCES

[1] Sonitor system website. http://www.sonitor.com/.

[2] M. Azizyan, I. Constandache, and R. Roy Choudhury. Surroundsense: mobile phone localization via ambience fingerprinting. In *Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 261–272. ACM, 2009.

[3] P. Bahl and V. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000.

[4] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. Walrus: wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pages 191–203. ACM, 2005.

[5] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha. Fm-based indoor localization. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 169–182. ACM, 2012.

[6] I. Constandache, X. Bao, M. Azizyan, and R. Choudhury. Did you see bob?: human localization using mobile phones. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, pages 149–160. ACM, 2010.

[7] I. Constandache, S. Gaonkar, M. Sayler, R. Choudhury, and L. Cox. Enloc: Energy-efficient localization for mobile phones. In *INFOCOM 2009, IEEE*, pages 2716–2720. IEEE, 2009.

[8] D. Dardari, C. Chong, and M. Win. Threshold-based time-of-arrival estimators in UWB dense multipath channels. *IEEE Transactions on Communications*, 56(8):1366–1378, 2008.

[9] G. Jin, X. Lu, and M. Park. An indoor localization mechanism using active rfid tag. In *Sensor Networks, Ubiquitous, and Trustworthy Computing, 2006. IEEE International Conference on*, volume 1, pages 4–pp. Ieee, 2006.

[10] S. Kay. *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall PTR, 1998.

[11] M. Kjærgaard, J. Langdal, T. Godsk, and T. Toftkjær. Entracked: energy-efficient robust position tracking for mobile devices. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*, pages 221–234. ACM, 2009.

[12] J. Lee and R. Scholtz. Ranging in a dense multipath environment using an UWB radio link. *IEEE Journal on Selected Areas in Communications*, 20(9):1677–1683, 2002.

[13] K. Lin, A. Kansal, D. Lymberopoulos, and F. Zhao. Energy-accuracy trade-off for continuous mobile device location. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 285–298. ACM, 2010.

[14] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye. Push the limit of wifi based localization for smartphones. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 305–316. ACM, 2012.

[15] K. Liu, X. Liu, and X. Li. Acoustic ranging and communication via microphone channel. In *Proc. IEEE Globecom '12*, Anaheim, California, USA, 2012. IEEE.

[16] K. Liu, H. Yin, and W. Chen. Low complexity tri-level sampling receiver design for uwb time-of-arrival estimation. In *Communications (ICC), 2011 IEEE International Conference on*, pages 1–5. IEEE, 2011.

[17] A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 348–353. IEEE, 2005.

[18] M. Minami, Y. Fukuju, K. Hirasawa, S. Yokoyama, M. Mizumachi, H. Morikawa, and T. Aoyama. Dolphin: a practical approach for implementing a fully distributed indoor ultrasonic positioning system. *UbiComp 2004: Ubiquitous Computing*, pages 347–365, 2004.

[19] R. Nandakumar, K. Chintalapudi, and V. Padmanabhan. Centaur: locating devices in an office environment. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 281–292. ACM, 2012.

[20] L. Ni, Y. Liu, Y. Lau, and A. Patil. Landmarc: indoor location sensing using active rfid. *Wireless networks*, 10(6):701–710, 2004.

[21] N. Patwari, J. Ash, S. Kyperountas, A. Hero III, R. Moses, and N. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, 2005.

[22] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14. ACM, 2007.

[23] N. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43. ACM, 2000.

[24] J. Taylor. *Ultra-wideband radar technology*. CRC, 2001.

[25] J. Valin, F. Michaud, J. Rouat, and D. Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 2, pages 1228–1233. IEEE, 2003.

[26] R. Wand, A. Hopper, V. Falcao, and J. Gibbons. The active bat location system. *ACM Transactions on Information Systems*, pages 91–102, 1992.

[27] J. Yang and Y. Chen. Indoor localization using improved rss-based lateration methods. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6. Ieee, 2009.