Towards Accurate Acoustic Localization on a Smartphone

Kaikai Liu, Xinxin Liu, Lulu Xie, Xiaolin Li
University of Florida, Gainesville, FL 32611, USA
kaikailiu@ufl.edu, xinxin@cise.ufl.edu, llxie@ufl.edu, andyli@ece.ufl.edu

Abstract—Since our daily activities are dominantly indoor, as smartphones emerge as the most popular personal computing companions, major IT companies recently launched aggressive investment on mobile indoor location services and positioning systems, e.g., on iOS or Android mobile devices. However, one major hurdle has not been conquered yet: smartphone-based high-resolution indoor localization. In this paper, we propose a practical solution for accurate ranging and localization based on acoustic communication between anchor nodes with speakers and the microphone on a smartphone. To identify different anchor nodes and enable time-of-arrival (TOA) ranging, we propose approaches for signal modulation, symbol detection and demodulation, synchronization and ranging. Experimental results show that the communication bit-error-rate and ranging accuracy is sufficient for our target applications. The preliminary results of localization demonstrate that our algorithm could achieve high-accuracy of 23cm in the offline mode with a promising potential for realtime smartphone-based indoor localization.

I. INTRODUCTION

Due to the blockage of the GPS signals in indoor environments, alternative approaches over the years have been proposed to address the problem of indoor automatic location sensing. These approaches vary in many aspects, such as the positioning method, signal type, cost of infrastructure, power requirement, and resolution in time and space.

Ranging-based method is preferred for its low complexity and high scalability [1]. Time-of-Arrival (TOA) and received-signal-strength (RSS) estimation are two typical ranging techniques. RSS based approach is less accurate and need the radio attenuation model as a prior. TOA estimation scheme is often preferred in some systems with high-accuracy requirement.

The ranging accuracy directly depends on the bandwidth and transmission speed of the operating signal ($\sim c/B$) [1]. Two types of signals are suitable for the ranging purpose, (1) Impulse-Radio Ultra-wideband (IR-UWB) signal with its sharp pulse and wideband properties (large B); (2) Ultrasound or acoustic signal due to its lower transmission speed (small c) [2]. However, UWB devices are very expensive and not available in consumer market. Using acoustic signal for localization, the limited bandwidth and strong attenuation drawback need to be compensated by using radio signal for assistance, e.g., Cricket [2]. Further, the requirement on dedicated ultrasound receiver impedes its wide adoption.

In this paper, we propose a fine-grained indoor localization prototype system using the standard microphone in a mobile phone. Using the audible-band acoustic signal to perform passive sensing without radio assistance and sound disturbance is promising for large deployment. However, the diverse noise sources in indoor environments, difficulty in synchronization and strong interference of the audible-band, as well as the low-complexity of the microphone receiver, pose significant challenges in the design of the smartphone-based accurate indoor location algorithm and system. Our contributions are manifold:

- We propose a TOA estimation scheme to obtain accurate ranging results; along with a symbol demodulation method with dynamic decision threshold and *transmit reference*. The beacon signal was designed to be unnoticeable.
- By identifying the most reliable ranging measurement as the reference node, we introduce a reference-based localization method to achieve reliable position results.

We implement the passive acoustic localization system and achieve preliminary results in demonstrating the feasibility of using acoustic signal for smartphone-based localization.

II. SYSTEM ARCHITECTURE AND TRANSCEIVER DESIGN

A. System Architecture

Our proposed indoor localization system makes the high-accurate indoor localization possible by using smartphones without additional hardware burdens on the user side. With several low-complexity anchor nodes preconfigured indoor, the mobile phone can localize itself by receiving beacon signals from more than three anchor nodes. By recording and demodulating the acoustic beacon signals sent by different anchor nodes, the receiver can calculate the relative TOA and perform real-time localization. Moreover, the anchor nodes modulate its unique information in the beacon signal; the mobile phone performs demodulation and decision to access the information bits, e.g., ID, position or temperature. With the ID information demodulated from the transmitted signal, the receiver can obtain the position of the anchor nodes by matching the ID to its database.

B. Signal Design and Modeling

Driven by the specific design goal of only using the microphone on the receiver side, we need to choose an appropriate transmitter signal band to match the capabilities of a user's

smartphone. The acoustic signal band of the microphone is very limited; the typical band of microphone is in the audible range, i.e., 200Hz-20KHz. To reduce interference between beacon signals and daily environmental noises, we choose the high frequency side of 17KHz-20KHz as the operating band. By keeping the power spectrum density (PSD) lower than the perception level of human's ears, and selecting an appropriate bandpass speaker to minimize the frequency leakage, the acoustic beacon can be unnoticeable.

III. SYMBOL DETECTION AND RANGING

A. Symbol Demodulation

The symbol detection and demodulation are the prerequisite for ranging with the duty of detecting the signal start region in the symbol level and differentiating anchor nodes.

Assume the received multi-path frequency-modulated signal is $r_{i,j}(k) = \sum_{l=0}^{\xi_{i,j}-1} \sqrt{\varepsilon} \cos(2\pi f_b k + \phi + \varphi_l(f_b)), b = 0, 1$; $\varphi_l(f_b)$ is a function of f_b due to the frequency selective channel that f_0 and f_1 may suffer different fading; f_0 and f_1 represents the symbol '0' and '1'. Construct the local correlation template $v_0(k) = \cos(2\pi f_0 k)$ and $v_1(k) = \cos(2\pi f_1 k)$ to demodulate symbol '0' and '1'. Then, using the constructed local templates to perform correlation and information extraction. For symbol representation, the correlation process can be shown as

$$a_{i,j}^0 = \mathbb{E}_k\{g_{i,j}(k) \cdot v_0(k)\} + w_{i,j}^0, \tag{1}$$
$$a_{i,j}^1 = \mathbb{E}_k\{g_{i,j}(k) \cdot v_1(k)\} + w_{i,j}^1$$

where $w_{i,j}^b = \mathbb{E}\{n_{i,j}(k) \cdot v_b(k)\}$; $b = 0, 1$; $k \in [k_j^{toa} + iN_k, k_j^{toa} + (i+1)N_k]$, $i = 0, \ldots, N_s - 1$. Define $\mathbf{s}_F^b = [\mathbb{E}_k\{g_{i,j}(k) \cdot v_0(k)\}, \mathbb{E}_k\{g_{i,j}(k) \cdot v_1(k)\}]^T$, $\mathbf{a} = [a_{i,j}^0, a_{i,j}^1]^T$ and $\mathbf{w} = [w_{i,j}^0, w_{i,j}^1]$. The process of $\mathbb{E}(\cdot)$ performs filtering and averaging. Using symbol '0' for example. When the signal is presented with two multi-paths ($\xi_{i,j} = 2$), frequency demodulation of $\mathbf{s}_F^0$ can be rewritten as

$$\mathbf{s}_F^0 \tag{2}$$
$$= \sup_\phi \frac{1}{\lceil F_s/B \rceil} \sum_{k=1}^{\lceil F_s/B \rceil} \sum_{l=0}^{\xi_{i,j}-1} \sqrt{\varepsilon} \cos(2\pi \Delta f \cdot k + \phi + \varphi_l(f_b))$$
$$= \sup_\phi \operatorname*{mean}_k [2\sqrt{\varepsilon} \cos(2\pi \Delta f \cdot k + \phi + \varphi_+(f_b)) \cos(\varphi_-(f_b))]$$

where $\varphi_+(f_b) = (\varphi_0(f_b) + \varphi_1(f_b))/2$, $\varphi_-(f_b) = (\varphi_0(f_b) - \varphi_1(f_b))/2$ for $l = 0, 1$; $\Delta f = f_b - f_0$; the bandwidth is $B = (f_0 - f_1)$. $\sup_\phi$ is calculating the super-bound of the function with the parameter of ϕ, i.e., the envelop. When symbol '0' is transmitted ($f_b = f_0, \Delta f = 0$), we have

$$\mathbf{s}_F^0(0) = \sup_\phi 2\sqrt{\varepsilon} \cos(2\pi k0 + \phi + \varphi_+(f_b)) \cos(\varphi_-(f_b)) \tag{3}$$
$$= 2\sqrt{\varepsilon} \cos(\varphi_-(f_b))$$

When the opposite symbol is transmitted ($f_b = f_1, \Delta f = B$), (2) can be shown as

$$\mathbf{s}_F^1(0) = \sup_\phi 2\sqrt{\varepsilon} \cos(2\pi kB + \phi + \varphi_+(f_b)) \cos(\varphi_-(f_b))$$
$$\tag{4}$$

From (3) and (4), we know that decision vector $\mathbf{s}_F^0 = [2\sqrt{\varepsilon} \cos(\varphi_-(f_0)), 0]^T$ when symbol '0' is transmitted; $\mathbf{s}_F^1 = [0, 2\sqrt{\varepsilon} \cos(\varphi_-(f_1))]^T$ when symbol '1' is transmitted. With the equal prior probability $P(b = 0, 1) = 1/2$, the maximum-likelihood (ML) is the optimal decision rule, resulting in the decision process as

$$\mathbf{y} = \mathbf{a}(1) - \mathbf{a}(0) = \mathbf{d}_\delta + \mathbf{w}_F(1) - \mathbf{w}_F(0) \gtrless_0^1 = \eta_d \tag{5}$$

where $\mathbf{d}_\delta = \mathbf{s}_F^1 - \mathbf{s}_F^0$. From (3) and (4), we can calculate $\mathbf{d}_\delta = [2\sqrt{\varepsilon} \cos(\varphi_-(f_1)), -2\sqrt{\varepsilon} \cos(\varphi_-(f_0))]^T$, where η_d is the communication decision threshold. When $\mathbf{y}$ is larger than the threshold, we can declare the information bit '1' is transmitted, and vice versa.

B. Dynamic Decision Threshold and Transmit Reference

The decision threshold η_d can be chosen as 0 with unknown prior information of ε and $\cos(\varphi_-(f_b))$. In real situations, the additional term $2\cos(\varphi_-(f_b))$ may cause significant performance loss, e.g., the demodulated signal vector can achieve the value of '0' when $\varphi_-(f_b) \approx \pi/2$, causing the signal undetectable.

Our proposed solution uses *transmit reference* (TR) to estimate the prior information of the received signal for symbol '1' and '0'. The beacon signal was designed by adding two bits "10" at the beginning of other information bits. On the receive side, we can assume that these TR bits suffer from the same attenuation as other bits in the same beacon period when passing through the aerial channel. By estimating the amplitude of the known symbol '1' and '0', we obtain the value of $2\sqrt{\varepsilon} \cos(\varphi_-(f_b))$. The dynamic communication decision threshold can be set as

$$\eta_d = \begin{cases} 2\sqrt{\varepsilon} \left[\cos(\varphi_-(f_1)) - \cos(\varphi_-(f_0))\right], & d = 1 \\ -2\sqrt{\varepsilon} \left[\cos(\varphi_-(f_1)) - \cos(\varphi_-(f_0))\right], & d = 0 \end{cases} \tag{6}$$

If the amplitude estimated from TR bits is s_{ref}^0 for symbol '0'; s_{ref}^1 for symbol '1', (6) can be re-written as $\eta_d = [\pm s_\delta]$, where $s_\delta = s_{ref}^1 - s_{ref}^0$.

For flat channel, frequency f_1 and f_0 suffer from the same attenuation, and $\varphi_-(f_0) = \varphi_-(f_1)$, (6) is simplified to $\eta_d = 0$. When $\varphi_-(f_0)$ and $\varphi_-(f_1)$ show large differences, setting the decision threshold dynamically according to (6) can ensure robust decision and compensate the performance loss caused by the frequency-selective channel.

C. TOA Estimation

In the previous Subsection, the symbol '1' is detected after the demodulation process (5). The ranging process should be started after the symbol '1' is received, i.e., the signal start region $\hat{i}$ is obtained. TOA estimation can be performed to estimate the first path signal ($l = 0$) from $\mathbf{r}_{j,k}$, i.e., estimate the TOA sample k_j^τ when signal is present.

Using the symbol demodulation results can benefit the TOA estimation by focusing on the obtained signal region.

By performing jump-back and search-forward (JBSF) [1], the TOA decision process can be shown as

$$\hat{\tau}_j^{toa} = \hat{k}_j^{\tau}/F_s = \min_k(k|\mathbf{r}_{j,k} > \hat{\eta}_{toa})/F_s - \frac{1}{2F_s}, \quad (7)$$

$$k \in [\hat{i}N_k - J_b, \cdots, (\hat{i}+1)N_k]$$

where J_b is the sampling points that jumped back, e.g., $J_b = N_k/2$; $\hat{\eta}_{toa}$ is the TOA estimation threshold. Setting the $\hat{\eta}_{toa}$ by maximizing the *TOA detection probability* is proposed on our previous work [3].

With $\hat{\tau}_j^{toa}$ available, the distance can be obtained by multiplying the speed of acoustic signal c as $d_j = c \times \hat{\tau}_j^{toa}$. The practical formula of sound speed in air can be written as $c = 20\sqrt{\vartheta + 273.15}$m/s, ϑ is the temperature in the air and can be measured by the anchor nodes.

IV. LOCALIZATION

A. Relative Distance From Anchor Nodes

To discriminate different beacon signals, every anchor node modulates its unique pseudo-code in the beacon signal as in the symbol demodulation process (5). For the multi-anchor operation, we choose the time division multiplexing (TDM) technique. It shows that every beacon period has some free space in case of inter-beacon interference. Assume there are total M anchor nodes with each index as m, the beacon information will be repeated for every M beacons in j-th symbol, have $m \equiv j(\mod M)$ and $j = m + \lfloor j/M \rfloor \times M$, $g_{i,j}(k) = g_{i,m}(k)$. The TOA value $\hat{\tau}_m^{TOA}$ is obtained from the mth anchor node periodically. For one round beacon, the TOA value from m-th anchor can be written as

$$\hat{\tau}_m^{toa} = \Delta_b + (m-1)T_p + t_m \quad (8)$$

where $m \in [1, \ldots, M]$, Δ_b is the unknown beginning time. T_p is the period time; t_m is the flight time of the beacon signal to the microphone. The real distance r_m can be represented by t_m as

$$\hat{r}_m = ct_m = c(\hat{\tau}_m^{toa} - \Delta_b - (m-1)T_p) \quad (9)$$

where $\Delta_b + (m-1)T_p$, different from beacon to beacon, is the unknown delay. To minimize the effect of Δ_b, we substrate the same distance from every measurement to obtain a relative distance value. By selecting one anchor node with minimum ranging variance as reference, i.e., $m = f$. By setting $\hat{r}_f = 0$, the relative distance of other nodes to this reference point can be written as

$$\hat{r}_m = [\hat{\tau}_m^{toa} - \hat{\tau}_f^{toa} - (m-f)T_p]c + n_m - n_f \quad (10)$$

In (10), n_m and n_f are the measurement noise for $\hat{\tau}_m^{toa}$ and $\hat{\tau}_f^{toa}$. T_p is the preset beacon period with the known initial value. For the distributed system, physical clocks are not synchronized between the anchor nodes and the microphone. To improve the accuracy of ranging estimation in (10), T_p should be updated when new beacon period is received. Define $m_t = \lfloor j/M \rfloor$ as period index parameter to label the total

$\lfloor N_p/M \rfloor$ times beacon round. For m_t-th round, T_p can be estimated by

$$\hat{T}_p(m_t) = \frac{1}{M}\left(\sum_{m=1}^{M} \hat{\tau}_{m+M(m_t+1)}^{toa} - \hat{\tau}_{m+M(m_t)}^{toa}\right) \quad (11)$$

when only consider one round time that the position result is calculated, we can simplify $\hat{T}_p = \hat{T}_p(m_t)$ in the following analysis.

B. Joint Estimation of the Position and Unknown Bias

With the measured distance from M anchor nodes, trilateration can be performed to localize the position of the microphone. The obtained ranging results in passive ranging process are the pseudo-ranges with unknown delay. Using the pseudo-range in localization, the timing delay should be resolved in positioning estimation. Assume the real position of the microphone is $\mathbf{p} = (x, y)$. For 2-D localization, the three unknown parameters are coordinate (x, y) and the timing delay. The real distance from the microphone to the anchor nodes can be assumed as $r_m = \sqrt{(x-x_m)^2 + (y-y_m)^2}$. The observation equation for the pseudorange $\hat{r}_m$ is

$$\hat{r}_m = \sqrt{(x-x_m)^2 + (y-y_m)^2} + \Delta + \tilde{n}_m \quad (12)$$

where $m = 1, \ldots, M$, $\hat{r}_f = 0$; $\hat{r}_m$ is obtained by (10). Δ is the unknown fixed delay for every anchor node. $\tilde{n}_m$ is the error of the pseudorange in (10). In vector notations, (12) can be expressed as

$$\hat{\mathbf{r}} = \mathbf{f}(x, y, \Delta) + \tilde{\mathbf{n}} \quad (13)$$

where $\hat{\mathbf{r}} = [\hat{r}_1, \cdots, \hat{r}_M]^T$, $\mathbf{f}(x, y, \Delta) = [r_1+\Delta, \cdots, r_m+\Delta]^T$ and $\tilde{\mathbf{n}} = [n_1, \cdots, n_m]^T$.

From $\mathbf{f}(x, y, \Delta)$, we define the vector of unknown parameters as $\theta = \begin{bmatrix} x & y & \Delta \end{bmatrix}^T$. The localization purpose is to estimate (x, y) from measurements, different approaches like Bayesian or ML estimation techniques can be applied depending on the prior information about θ. For the case that the distribution of the unknown parameter θ is unknown, uniform distribution can be assumed. Then, the maximum log-likelihood (ML) estimator can be written as

$$\hat{\theta}_{ML} = \arg\max_{\theta} \log P(\hat{\mathbf{r}}|\theta) \quad (14)$$

where $P(\hat{\mathbf{r}}|\theta)$ is the pdf of the measurement vector $\hat{\mathbf{r}}$ conditioned on θ. The conditional pdf of $\hat{\mathbf{r}}$ can be expressed as

$$P(\hat{\mathbf{r}}|\theta) = \frac{1}{\sqrt{2\pi \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{\mathbf{Q}^T\boldsymbol{\Sigma}^{-1}\mathbf{Q}}{2}\right) \quad (15)$$

where $\mathbf{Q} = (\hat{\mathbf{r}} - \mathbf{f}(x, y, \Delta))$. $\boldsymbol{\Sigma}$ is the covariance matrix for $\hat{\mathbf{r}}$ as $\boldsymbol{\Sigma} = \{\text{Cov}(\hat{r}_i, \hat{r}_j)\}$, $i, j = 1, \ldots, M$. $\det(\boldsymbol{\Sigma})$ calculates the determinant of $\boldsymbol{\Sigma}$. For the Gaussian noise component in (13), it has $\tilde{\mathbf{n}} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the error covariance matrix with diagonal entries of $\sigma_m^2 + \sigma_f^2$, other elements of σ_f^2.

For noise distribution with zero mean and a known covariance matrix, (14) can be simplified as

$$\hat{\theta}_{ML} = \arg\min_{\theta} (\hat{\mathbf{r}} - \mathbf{f}(x, y, \Delta))^T\boldsymbol{\Sigma}^{-1}(\hat{\mathbf{r}} - \mathbf{f}(x, y, \Delta)) \quad (16)$$

where $(\hat{\mathbf{r}} - \mathbf{f}(x, y, \Delta))$ represents the estimation error. (16) is also the form of non-linear least-squares (NLS) estimator, that steepest descent, Gauss-Newton and Taylor series based methods can be used to solve the problem [4].

Rather than solving (16) directly, we use the properties of (10) to cancel out the nonlinear term. Then (16) can be written as a Least-Square (LS) problem [4]. It is equivalent to find θ which minimizes the sum squares of M independent error vector of (16). By minimizing this quadratic function, we can obtain the solution of $\hat{\theta}$.

The estimated position of the target can be obtained by selecting the first two parameters (x, y) of $\hat{\theta}$. The obtained Δ in each result of $\hat{\theta}$ can be used as a constraint that $\Delta \approx r_f = \sqrt{(x - x_f)^2 + (y - y_f)^2}$. By calculating r_f and using the obtained (x, y), the difference between the estimated Δ can be shown as $\hat{e} = \sqrt{(\hat{\theta}(1) - x_f)^2 + (\hat{\theta}(2) - y_f)^2} - \hat{\theta}(3)$. Small $\hat{e}$ indicates good position results. Such delay-constraint (DC) can be used as the self-evaluation of the position results and can filter out some incorrect estimated positions.

V. EXPERIMENTAL EVALUATION

A. Experiment Setup

To evaluate the system performance of localization, we conducted the measurements in a typical office environment to test the localization accuracy. We use $M = 4$ anchor nodes along with a control node to synchronize the whole network. The position of the anchor nodes are $(0.2, 1.2)m$, $(2.7, 0)m$, $(6.4, 0.1)m$, $(6.5, 2.3)m$ and $(2.9, 2.2)m$. The tested smartphone (iPhone 4S) is put in the position of $(1.68, 1.02)m$ and $(4.6, 1.03)m$, denoted as "Env1" and "Env2", respectively. By calculating the position of the mobile phone, and comparing with the true position value, the correctness of the result can be validated. The beacon period of system is $T_p = 0.9710s$; symbol duration is $T_s = 0.0205s$. The total symbol number is $N_s = 17$, with 15 information bits as the unique ID of each anchor, and 2 bits ('1' and '0') for transmit reference. The sampling rate is $F_s = 44.1$KHz.

B. Communication and Ranging Results

TABLE I
BER AND RANGING NMSE RESULTS IN DIFFERENT DISTANCES

Distance(m)	2.1844	3.2004	4.2164	5.2324	6.2484	7.2644
BER	0.0015	0	0.0016	0.0202	0.0020	0.0095
Variance(m)	0.021	0.016	0.035	4.637	0.015	0.067

The metric for assessing the performance of communication is bit-error-rate (BER); the metric for ranging accuracy is the estimation variance defined by $\sigma_{range} = c\sqrt{E[(\hat{\tau}^{TOA} - \tau^{TOA})]^2}$, c is the speed of acoustic signal. To test the communication and ranging performance under different operating distances, we conduct experiment for measuring the BER and variance when put the smartphone at different distances from $2.18m \sim 7.26m$. The BER and variance results by detecting the beacon signal for one anchor node are shown in Table I. From Table I, we know that some distance region,

e.g., near 5.2m, has high BER and variance due to the blockage or interference of the beacon signal. When the distance between anchor node and smartphone is reached to $7.26m$, the ranging and communication results are still acceptable. Such result demonstrates that the operating distance of our proposed system is sufficient for indoor localization.

TABLE II
BER AND RANGING VARIANCE RESULTS FOR 4 ANCHOR NODES WHEN SMARTPHONE IS PLACED IN ENV1$(1.68, 1.02)m$ AND ENV2$(4.6, 1.03)m$

Index	Metrics	m=1	m=2	m=3	m=4
Env1	BER	0.0049	0.0040	0.0011	0.0015
	Range Variance (m)	0.0521	0.9597	0.0747	0.5004
Env2	BER	0.0043	0.0029	0.0031	0.0029
	Range Variance (m)	0.9433	0.6917	0.2488	1.2249

To test the BER and variance in real localization scenarios, the communication and ranging results from 4 anchor nodes when smartphone is placed in two positions are shown in Table II. From Table II, we know that the communication performance is sufficient for localization with largest BER less than 0.49%. The ranging variance from different anchor nodes is hard to compare due to the various propagation blockage or interference. The average ranging variance in Env2 is slightly larger than Env1.
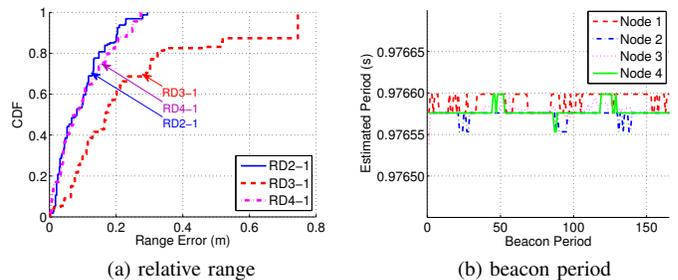

(a) relative range   (b) beacon period

Fig. 1. The relative range results (a) and beacon period estimation results (b) when smartphone is placed in Env1$(1.68, 1.02)m$.

To better evaluate the relative ranging performance of $\hat{r}_m$ with the reference node of $f = 1$, we calculate the CDF of the measured relative distance from node 2 to 1 ("RD2-1"), node 3 to 1 ("RD3-1"), and node 4 to 1 ("RD4-1") as shown in Fig. 1a. The results show that the ranging result of node 4 is less accurate than node 2 and 3. If choosing 80% probability as standard, the accuracy of relative distance of "RD2-1" and "RD3-1" is near 0.17 meter, while "RD4-1" case can achieve the accuracy of 0.32 meter in Env1.

The beacon period T_p in (11) is another important parameter that need to be estimated before localization process. Due to the imperfect clock used in hardware, different anchor nodes may have slightly different offsets. The estimated period from four anchor nodes is shown in Fig. 1b with very small noise variance. Estimating T_p periodically in real-time is a suitable approach to compensate the offset and improve accuracy.

C. Localization Results

The location error can be calculated by $\hat{e} = \sqrt{(\hat{x}-x_a)^2 + (\hat{y}-y_a)^2}$, where (x_a, y_a) is the actual position; $\hat{x} = \hat{\theta}(1)$, $\hat{y} = \hat{\theta}(2)$ is the estimated position. Four anchor nodes (N4) have been used to provide beacon information. The estimation results can be filtered and cleaned using the delay-constraint (DC) and median filter (MF).
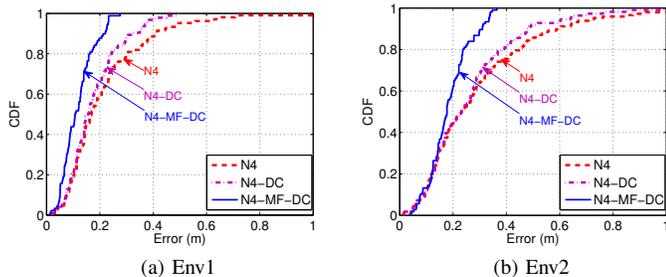

Fig. 2. The CDF of position result in indoor environment when the smartphone is placed near Env1(a) and Env2(b).

Fig. 2a and Fig. 2b show the CDF of the position error $\hat{e}$ using 4 anchor nodes (N4) when the target is near $(1.68, 1.02)m$ and $(4.6, 1.03)m$, respectively. Using the delay-constraint in parameter estimation can improve performance by filtering some erroneous position results that cannot meet the inner constraint, i.e., "N4-DC" is better than "N4" in CDF results. We also use 5 points median filter (MF) to clean the results of "N4-DC". With MF introduced, "N4-MF-DC" achieves improved probability at a given position error bound than other two cases as shown in Fig. 2a and Fig. 2b. The cumulative probability for "N4-MF-DC" is larger than 80% when the position error bound is set to 0.25 meter. The position accuracy in Env1 is better than Env2 due to its better ranging accuracy.
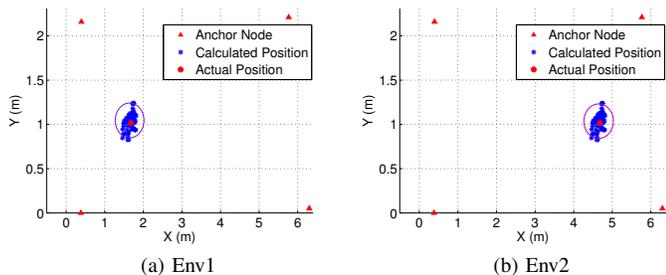

Fig. 3. The position results in indoor environment when the smartphone is placed near Env1(a) and Env2(b) when using the method of 'N4-MF-DC'.

Fig. 3a and Fig. 3b show the position results when the target is near $(1.68, 1.02)m$ and $(4.6, 1.03)m$ when using the method of "N4-MF-DC". The obtained position results are concentrated near the ground truth with very small variance.

VI. RELATED WORK

The three main categories for the position calculation are angle-based, ranging-based [1], [2] and fingerprinting based approaches [5]. The angle-based method relies on the directional antenna. While, most of the narrow beam directional antenna are expensive and difficult to build. The fingerprinting approach performs location sensing by signature matching and do not require on infrastructure. However, the achieved accuracy is only in the room-level and extensive experiments are needed to obtain the prior information for mapping.

Recent solutions using audible-band acoustic signal for localization provides a feasible solution by leveraging the ubiquitous microphone in mobile devices. A. Mandal et al. [6] using PDA to transmit annoying 4KHz acoustic signal to the WiFi-enabled sensor network and perform sound source localization. C. Peng et al. [7] propose solutions of two-way ranging and self recording between per-to-per devices to solve the problem of synchronization with indoor operating range around 4 meters. The active transmit mode easies the synchronization problem without relying on radio signal, however, such mode is very difficult to handle multi-users; e.g., two users transmit beacon simultaneously will cause interference.

VII. CONCLUSION

Using the audible band acoustic signal for localization provides highly accurate results without additional hardware requirement on a smartphone. To facilitate ranging and communication and overcome the limitation of low-cost low-power hardware, a dynamic communication demodulation and a TOA estimation method have been proposed. Experimental results show that the communication BER is less than 0.49% for all cases, and the ranging variance can achieve the accuracy within $10cm$. The preliminary results of localization by using the ranging data shows that our proposed algorithm can achieve about $23cm$ position accuracy with more than 80% probability. Significant improvements could be further achieved in terms of realtime, accuracy, cost and scalability in the future.

REFERENCES

[1] J. Lee and R. Scholtz, "Ranging in a dense multipath environment using an UWB radio link," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1677–1683, 2002.

[2] 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*. ACM, 2000, pp. 32–43.

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

[4] I. Guvenc, C. Chong, and F. Watanabe, "Analysis of a linear least-squares localization technique in los and nlos environments," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*. IEEE, 2007, pp. 1886–1890.

[5] 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*. ACM, 2009, pp. 261–272.

[6] 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*. IEEE, 2005, pp. 348–353.

[7] 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*. ACM, 2007, pp. 1–14.