Acoustic Ranging and Communication via Microphone Channel

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

***Abstract*—Absolute GPS-coordinates are typically inaccessible indoor. Pervasive smartphones offer new opportunities for relative indoor localization. The low-complexity and high-accuracy ranging under the communication-link is a crucial requirement for enabling location sensing in indoor environments. In this paper, we propose a Time-of-Arrival (TOA) estimation and communication scheme utilizing the ubiquitous speaker/microphone pair to achieve accurate ranging. To compensate for the performance loss caused by this simple device, an optimized TOA estimation method is proposed for enhanced ranging reliability and accuracy. To overcome the strong channel fading in acoustic communication, a dynamic demodulation method that jointly uses frequency and amplitude information has been proposed. Experimental results show that our proposed scheme can achieve near $8cm$ TOA ranging accuracy and 0.55% communication bit-error-rate (BER) with 70% probability.**

I. Introduction

Accurate indoor localization utilizing low-complex devices promises a wide spectrum of applications, especially when the GPS signal is inaccessible due to the blockage of the satellite signal [1]. If sub-meter resolution can be achieved, it will fundamentally change and improve the way that current location-based services are delivered.

TOA estimation based on the communication channel can obtain the signal flight distance as pseudorange [1]. Ranging-based trilateration method utilizes these pseudoranges to calculate location information. For location-aware techniques, communication and accurate TOA estimation capabilities are two important prerequisites. Mainly, two categories of approaches have been proposed to solve this problem. The first type of solutions utilizes Impulse Radio Ultra-Wideband (UWB) technique for indoor TOA-based ranging, because ranging precision directly depends on the bandwidth of the operating signal. Using UWB signal has attracted significant research interests and become a standard as IEEE 802.15.4a [2]. However, the full-digital coherent IR-UWB system requires the bandwidth of several GHz to guarantee sub-meter ranging accuracy, which increases the overall hardware cost and processing power dramatically [3]. Some techniques, e.g., energy receiver and finite-resolution digital receiver, have been proposed to lower the overall complexity [3] [4] [5]. Although much lower complexity can be achieved, it is still very expensive and requires additional special hardware.

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).

The second kind of solutions utilizes the ultrasound signal to perform accurate ranging. Compared with electromagnetic signal used in a UWB device, the aerial acoustic signal is more pervasive and can achieve ranging accuracy with much lower hardware cost. Due to slower transmission speed of acoustic signal, even several KHz signal bandwidth can result in centimeter-level ranging accuracy. Yang et al. [6] uses the acoustic approach to detect the position of a phone using car speakers. However, they only need to detect the relative region in a car. The Cricket localization system developed by MIT [7] [8] [9] using the ultrasound for ranging achieved centimeter-level accuracy. They use the radio signal for synchronization and perform inter-node ranging by using their developed devices. The requirements of the dedicated device impede its wide-spread adoption by ordinary users.

In this paper, we propose to directly utilize the existing hardware of consumers to achieve accurate ranging. We address the problem of ranging and communication by using the aerial acoustic signal over the microphone channel. Our scheme helps users achieve indoor localization by using their smartphones. Even the pre-placed anchor node providing the ranging beacon signal can be implemented with very low cost, i.e., a small speaker with a Microcontroller is sufficient. We design a joint symbol detection and TOA estimation method to achieve robust and accurate ranging results. We derive the TOA threshold by maximizing the TOA right detection probability that can be adaptively tuned in different environments. We also propose a dynamic demodulation method based on direct frequency discrimination and amplitude matching with transmit reference to address the worse communication channel condition than UWB and ultrasound signal. The experimental results show that our proposed TOA and communication scheme achieve very good mean-square-error and bit-error-rate with high probability.

II. System Architecture and Signal Modeling

A. System Architecture

To facilitate the indoor localization, we introduce an acoustic ranging and communication technique that leverages the existing low-complex anchor node infrastructure and microphone sensor in a mobile target (MT), e.g., smartphone. An anchor node sends its unique acoustic beacon signal periodically. The beacon signal contains the ID or position information of its transmitter. On the receiver side, the microphone searches and captures the existing beacons and perform TOA estimation

978-1-4673-0921-9/12/$31.00 ©2012 IEEE

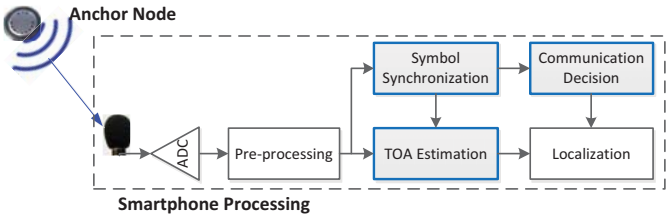


Fig. 1. The System Architecture.

and communication decision. After receiving beacon information from more than 3 anchor nodes, the target position can be determined by combing the information from the TOA estimation process and anchor positions. The simplified architecture of our audible-band acoustic localization system is shown in Fig. 1. In this paper, we focus on the two important prerequisites of localization, i.e., ranging and communication, and present in detail design goals, the signal modeling, symbol synchronization, TOA estimation and communication decision.

B. Transmit and Receive Signal Modeling

The main operating signal in our proposed system is the acoustic signal that should be received and processed by a normal microphone to lower the hardware requirement. However, the acoustic signal band of the microphone is very limited, falling in the audible range (200Hz-20KHz) of humans. To reduce interference between beacon signals and daily environmental noises, we choose the high frequency side of 17KHz-20KHz as the operating band since human's ears are insensitive to such high frequency sound.

The transmitted acoustic beacon signal from the anchor node can be modeled as

$$g_t(t) = \sqrt{\varepsilon} \sum_{j=0}^{N_p-1} \sum_{i=0}^{N_s-1} g_{i,j}(t - jT_p - iT_s) \qquad (1)$$

where $g_{i,j}(t)$ is the transmitted signal with i, j denote the ith symbol in jth beacon period, ε is the signal energy. T_p is the beacon period with the beacon number of N_p; T_s is the symbol durations with N_s symbol numbers. Some free-space are left in each beacon period to avoid the inter-beacon interference, i.e., $N_s \times T_s < T_p$. The value of each information bit is $b_i = \pm 1$ and repeated for every period. b_i is carried by $g_{i,j}(t)$ using binary frequency modulation (BFSK) in f_1 and f_0 frequency point for symbol '1' and '0', respectively.

The acoustic beacon signal is captured by the microphone and converted to the electrical domain after propagation through the free space with distortion. Passing through the analog-to-digital convertor (ADC), the received signal will be digitalized as $r(k)$ with the sampling frequency of F_s, and $k \in [1, \ldots, N_k]$, where $N_k = T_s \times F_s$. The digital version of the transmitted signal is $g_{i,j}(k)$; every symbol contains N_k sampling points. The digital result that we get is

$$r_{i,j}(k) = \sum_{l=0}^{\xi_{i,j}-1} A_{i,j}^l \cdot g_t(k - k_{i,j}^l) + n_{i,j}(k) \qquad (2)$$

where $\xi_{i,j}$ is the total number of propagation paths, with $A_{i,j}^l$, $k_{i,j}^l = \tau_{i,j}^l \times F_s$ represents the digital version of the multi-path delay. The term $n_{i,j}(t)$ is independent white Gaussian noise in the ith symbol of jth period. $N_b = T_p \times F_s$ means the number of sampling points in one beacon period.

III. Symbol Synchronization and TOA Estimation

Due to non-identical clocks in the transmitter/receiver pair, we need to perform symbol synchronization (SS) before communication decision. Measuring the time-of-arrival (TOA) of the transmitted signal is another step that is used to estimate the signal flight distance from the transmitter to the receiver.

The first path of the multipath signal ($\tau_{i,j}^l, i = 0, l = 0$) in one beacon is often called the TOA path that can be used to characterize the line-of-sight distance [10]. To lower the overall system complexity and exploit the similarity feature of the SS and TOA, we use the Neyman Pearson (NP) criterion in SS to detect the signal region with a fixed false-alarm rate. We then use the result of SS to improve the reliability and accuracy of the TOA estimation, and perform TC-based TOA estimation by maximizing our derived right detection probability.

A. Symbol Synchronization

To detect the i-th symbol that the beacon signal starts, we choose to extract continuous M_s points in every N_k interval to speedup the detection; N_k interval equals to the symbol rate T_s. Due to the bandpass properties of the received signal, we need sufficient length of samples to cover the whole period of the high frequency signal, e.g., choosing M_s to cover the whole period for the highest frequency component that $M_s = \lceil (1/f_0) \times F_s \rceil$. The decision process for symbol detection can be expressed as

$$\hat{i}_s = \min_i \left(\frac{1}{M_s} \sum_{k=iN_k}^{(iN_k+M_s)} |r_{i,j}(k)| > \hat{\eta}_{syn} \right) \qquad (3)$$

where $\min(\cdot)$ is the function that selects the first i that the decision vector crossing the threshold; $\hat{\eta}_{syn}$ is the threshold for synchronization; $|\cdot|$ is the absolute function to extract the amplitude information. For simplicity, we define decision vector as $z_i = \frac{1}{M_s} \sum_{k=iN_k}^{(iN_k+M_s)} |r_{i,j}(k)|$.

An important parameter involved in (3) is the synchronization threshold $\hat{\eta}_{syn}$. To determine this parameter, hypothesis tests can be used to minimize error detection probability. The process of (3) is to detect the signal from the noise component $n_{i,j}(k)$ with variance of σ^2. Since each individual sample $r_{i,j}(k)$ is a Gaussian random variable, the first moment of $r_{i,j}(k)$ when signal is present (H_1 condition) can be written as $\mathbf{E}(r_{i,j}(k); H_1) = \sqrt{\varepsilon}\mathbf{E}_l(A_{i,j}^l) \triangleq \sqrt{\varepsilon}\hat{A} = \mu$, $\mathbf{Var}(r_{i,j}(k); H_1) = \sigma^2$; where $\hat{A}$ is the estimated mean value of the multi-path amplitude ($A_{i,j}^l$). In the noise region that

signal is not present (H_0 condition), the statistical parameters of $r_{i,j}(k)$ is $\mathbf{E}(r_{i,j}(k);H_0)=0$, $\mathbf{Var}(r_{i,j}(k);H_0)=\sigma^2$. The decision vector z_i is the mean absolute value of $|r_{i,j}(k)|$, and has a folded normal distribution. The probability density function (PDF) of z_i is given by

$$f(z_i;\hat{A},\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(-z_i-\sqrt{\varepsilon}\hat{A})^2}{2\sigma^2}\right)+\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(z_i-\sqrt{\varepsilon}\hat{A})^2}{2\sigma^2}\right) \quad (4)$$

where $z_i\geq 0$. For the random variable z_i, its first and second moment can be written as shown in [11]. The moment can be simplified under H_0 condition that $\hat{A}=0$, as $\mathbf{E}(z_i)=\sqrt{2/\pi}\sigma$, $\mathbf{Var}(z_i)=(\pi-2)/\pi\sigma^2$.

After performing the absolute function $|\cdot|$ in (3), the probability that the sample crossing the threshold $\hat{\eta}_{syn}$ in the noise region (H_0 condition) can be shown as

$$P_{fa}(|r_{i,j}(k)|)=\Pr(|r_{i,j}(k)|>\hat{\eta}_{syn};H_0)=2Q\left(\hat{\eta}_{syn}/\sigma\right) \quad (5)$$

where σ is an unknown parameter; $Q(x)$ is the Gaussian Q function. $P_{fa}(|r_{i,j}(k)|)$ is the false alarm rate that detected the wrong sampling point. To estimate σ, one method is direct using $\hat{\sigma}=\sqrt{\mathbf{E}(r_{i,j}(k))^2}$ in the noise region. However, using the first moment of the folded normal distribution can lead to a more simplified method as $\hat{\sigma}=\sqrt{\pi/2}\mathbf{E}(z_i;H_0)$, where it does not need to perform complex $(\cdot)^2$ and $\sqrt{(\cdot)}$ in estimation and really suitable for real-time situations.

With the unknown a prior information, using Neyman Pearson (NP) criterion can achieve optimal performance by maintaining a constant false-alarm rate γ_{fa}. Thus, the detection threshold ($\hat{\eta}_{syn}$) can be determined by (5) as

$$\hat{\eta}_{syn}=\sqrt{\pi/2}\mathbf{E}(z_i;H_0)Q^{-1}(P_{fa}(|r_{i,j}(k)|)/2) \quad (6)$$

where $P_{fa}(|r_{i,j}(k)|)$ is the constant false alarm rate that should be pre-set according to application demands; high requirement on $P_{fa}(|r_{i,j}(k)|)$ is often achieved at the sacrifice of detection probability. In the first step detection, $P_{fa}(|r_{i,j}(k)|)$ should be set slightly higher to ensure no information loss, while the second step in TOA estimation can help to keep the overall $P_{fa}(|r_{i,j}(k)|)$ to a lower level.

With the threshold ($\hat{\eta}_{syn}$) available, the probability that the correct signal is detected in the signal region (H_1 condition) is given by

$$P_d(|r_{i,0}(k)|)=\Pr(|r_{i,j}(k)|>\hat{\eta}_{syn};H_1) \quad (7)$$
$$=Q\left(\frac{(\hat{\eta}_{syn}-\sqrt{\varepsilon}\hat{A})}{\hat{\sigma}}\right)+Q\left(\frac{(\hat{\eta}_{syn}+\sqrt{\varepsilon}\hat{A})}{\hat{\sigma}}\right)$$

where $\sqrt{\varepsilon}\hat{A}$ is the signal parameter that can be estimated in the signal region as $\sqrt{\varepsilon}\hat{A}=\mathbf{E}(r_{i,j}(k);H_1)$. By using the detection threshold (6) with a fixed false alarm rate in (3), the signal start region ($\hat{i}_s$) can be detected at the probability of (7) and the communication synchronization can be asserted.

B. TOA Estimation

1) Estimation Procedure: When $\hat{i}_s$ is obtained in (3), we perform precise detection for the TOA path in the symbol region near $\hat{i}_s$. Jump-back and search-forward (JBSF) [12] is a method that suitable for precise TOA searching after coarse detection. We define $z_k=|r_{i_s,j}(i_sN_k-J_b,\ldots,i_sN_k+M_s)|$ as the subtracted decision sequence for TOA estimation, J_b is the number of sampling points that jumped back in JBSF; the length of z_k is $M_k=J_b+M_s+1$. The decision process in this part can be written as

$$\hat{\tau}_j^{TOA}= \quad (8)$$
$$\begin{cases} T_s\cdot[\min(k|z_k>\hat{\eta}_{TOA})+\hat{i}_sN_k-J_b]-\frac{1}{2}T_s & k<M_k \\ \text{Re-estimate } \hat{i}_s & k\geq M_k \end{cases}$$

where $\hat{\eta}_{TOA}$ is the precise TOA estimation threshold; $\hat{i}_s$ is obtained in the previous detection stage indicates the symbol period; J_b can be set the same as the step size in the first step detection; i.e., $J_b=N_k$. The second part of (8) means that we need to re-estimate the signal region when the decision process cannot obtain a TOA value in z_k and $k\geq M_k$, that is no sampling point crossed the threshold. This problem is often due to the false detection in symbol synchronization, a suitable way to solve this problem is to re-estimate $\hat{i}_s$. From such process, we know that the overall false alarm rate in SS (3) can be lowered in TOA estimation.

2) Maximizing the Right Detection Probability: In (8), the TOA threshold $\hat{\eta}_{TOA}$ should be determined before decision process. The decision vector z_k also follows the folded normal distribution in (4) and with the same statistical parameters. Then the probability that sample z_k crossing the threshold in the noise and signal region can be calculated as the way in (5) and (7); shown as $P_{fa}(z_k)$ and $P_d(z_k)$, respectively. The probability that detected the right TOA path can be denoted as P_{rt}, the two kind TOA estimation errors are the early detection P_{ed} and late detection P_{ld}. P_{ed} is caused by selecting the incorrect crossing samples earlier to true TOA path due to noise interference; P_{ld} is the probability that the TOA estimator missed the true TOA path, and detected a wrong sample in the signal region later due to channel fading. Assume the true TOA path k_j^{TOA} is uniformly distributed in the region of the estimation area $(i_sN_k-J_b,\ldots,i_sN_k+M_s)$ with total length of $M_k=J_b+M_s+1$. The probability that the true TOA path in any sampling point is $p_{toa}=1/M_k$. The early detection probability that detected one point in noise region is $P_{ed}(k_j^{TOA})=1-(1-P_{fa}(z_k))^{k_j^{TOA}}$, which means at least one sample crossed the threshold before TOA path. The late detection $P_{ld}(k_j^{TOA})=(1-P_d(z_k))(1-P_{fa}(z_k))^{k_j^{TOA}}$ shows the TOA point is missed. The right detection probability is $P_{rt}(k_j^{TOA})=P_d(z_k)(1-P_{fa}(z_k))^{k_j^{TOA}}$, where $\sum(P_{ed}(k_j^{TOA})+P_{rt}(k_j^{TOA})+P_{fa}(k_j^{TOA}))=1$. Since k_j^{TOA} is uniformly distributed with probability p_{toa}, the

expectation for the error estimation probability over k_j^{TOA} is

$$\begin{aligned} P_{err}^{TOA} &= \mathbf{E}_{k_j^{TOA}}\left(P_{ed}(k_j^{TOA}) + P_{fa}(k_j^{TOA}))\right) \\ &= \sum_{k_j^{TOA}=0}^{M_k}\left(\frac{1}{M_k}[1-(1-P_{fa}(z_k))^{k_j^{TOA}}]\right) \\ &+ \sum_{k_j^{TOA}=0}^{M_k}\left(\frac{1}{M_k}[(1-P_d(z_k))(1-P_{fa}(z_k))^{k_j^{TOA}}]\right) \\ &= 1 - \frac{P_d(z_k)}{M_k P_{fa}(z_k)}[1-(1-P_{fa}(z_k))_k^M] \end{aligned} \tag{9}$$

where the last part of (9) is the correct detection probability $P_{rt}^{TOA} = P_d(z_k)[1-(1-P_{fa}(z_k))_k^M]/M_k P_{fa}(z_k)$. Perform Taylor expansion of $(1-P_{fa}(z_k))_k^M$, we have $(1-P_{fa}(z_k))_k^M \approx 1 - M_k P_{fa}(z_k) + C_{M_k}^2 (P_{fa}(z_k))^2$. Then P_{rt}^{TOA} can be simplified as

$$\begin{aligned} P_{rt}^{TOA} &\approx \frac{P_d(z_k)}{M_k P_{fa}(z_k)}(M_k P_{fa}(z_k) - C_{M_k}^2 (P_{fa}(z_k))^2) \\ &= P_d(z_k)[1 - \frac{1}{2}(M_k - 1)P_{fa}(z_k)] \end{aligned} \tag{10}$$

From (10), we know that increase $P_d(z_k)$ or decrease $P_{fa}(z_k)$ of the single sampling point can contribute a better P_{rt}^{TOA}. Small M_k also helps to improve the performance that less sampling points being detected.

To achieve a better TOA estimation scheme, selecting an appropriate TOA threshold by maximizing P_{rt}^{TOA} provides a feasible way. Since $P_d(z_k)$ and $P_{fa}(z_k)$ is a function of $\hat{\eta}_{TOA}$, P_{rt}^{TOA} can be written as $P_{rt}^{TOA}(\hat{\eta}_{TOA})$. The maximum value of P_{rt}^{TOA} can be achieved when $\hat{\eta}_{TOA} = \hat{\eta}_{TOA}^{oth}$, as

$$\hat{\eta}_{TOA}^{oth} = \arg\max_{\hat{\eta}_{TOA}}(P_{rt}^{TOA}(\hat{\eta}_{TOA})) \tag{11}$$

For simplification, we can use the Maclaurin series of the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\sum_{n=0}^{\infty}[(-1)^n x^{2n+1}]/[n!(2n+1)]$ to express the approximate function of $Q(x)$ as

$$\begin{aligned} Q(x) &\approx \frac{1}{2} - \frac{1}{2}\mathrm{erf}(x/\sqrt{2}) \\ &= \frac{1}{2} - \frac{1}{\sqrt{\pi}}(x/\sqrt{2} - \frac{(x/\sqrt{2})^3}{3} + \ldots) \end{aligned} \tag{12}$$

when x is very small, even the first-order term of (12) can well represent $Q(x)$ as $Q(x) \approx 1/2 - x/\sqrt{2\pi}$. Lets define $z = \hat{\eta}_{syn}/\hat{\sigma}$ and $s = \sqrt{\varepsilon}\hat{A}/\hat{\sigma}$. If only consider $s > 0$, then $P_d(z_k)$ and $P_{fa}(z_k)$ can be simplified as $P_d(z_k) \approx 1/2 - (z-s)/\sqrt{2\pi}$ and $P_{fa}(z_k) \approx 1 - 2z/\sqrt{2\pi}$.

By substituting $P_d(z_k)$ and $P_{fa}(z_k)$ into (10), and define $w = \frac{1}{2}(M_k - 1)$, P_{rt}^{TOA} can be written as

$$\begin{aligned} P_{rt}^{TOA}(z) &= [1/2 - (z-s)/\sqrt{2\pi}][1 - w(1 - 2z/\sqrt{2\pi})] \\ &= -\frac{w}{\pi}z^2 + (\frac{2w-1}{2\sqrt{2\pi}} + \frac{ws}{\pi})z + \frac{1-w}{\sqrt{2\pi}}s + \frac{1-w}{2} \end{aligned} \tag{13}$$

where (13) is a convex function with its maximum achieved when $z = \frac{s}{2} + \frac{\sqrt{\pi}}{2\sqrt{2}}(1 - \frac{1}{2w})$. Then, $\hat{\eta}_{TOA}^{oth}$ can be written as

$$\hat{\eta}_{TOA}^{oth} = \frac{1}{2}\sqrt{\varepsilon}\hat{A} + \frac{\sqrt{\pi}}{2\sqrt{2}}(1 - \frac{1}{M_k - 1})\hat{\sigma} \tag{14}$$

where $\sqrt{\varepsilon}\hat{A}$ and $\hat{\sigma}$ can be estimated as the same in (6). By using (14) in TOA estimation of (8), the optimized TOA performance can be achieved under the criterion of maximum right detection probability as shown in (11).

IV. Communication Decision

After the synchronization and TOA estimation, we can perform symbol demodulation for every N_k points from k_j^{TOA} to obtain the information bit. For the frequency modulated signal of $g_{i,j}(k)$, we can model $g_{i,j}(k) = \sqrt{\varepsilon}\cos(2\pi f_d k + \phi), d = 0, 1$, where ϕ is the fixed unknown phase information between the local template and received carrier wave, $f_d = f_1$ represents the symbol '1'; $f_d = f_0$ represents symbol '0'.

One feasible way to demodulate the information is to construct the local correlation template $v_0(k) = \cos(2\pi f_0 k)$ and $v_1(k) = \cos(2\pi f_1 k)$ for symbol '0' and '1', respectively. The constructed local template is used to perform correlation and information extraction. For symbol representation, the decision vector can be shown as $\mathbf{s}_F = [\mathbb{E}_k\{g_{i,j}(k)\cdot v_0(k)\}, \mathbb{E}_k\{g_{i,j}(k)\cdot v_1(k)\}]^T$, the process of $\mathbb{E}(\cdot)$ performs filtering and averaging. Using symbol '0' for example, and assume the bandwidth is $B = (f_0 - f_1)$, then

$$\begin{aligned} \mathbf{s}_F(0) &= \mathbb{E}_k\{\sqrt{\varepsilon}\cos(2\pi f_d k + \phi)\cos(2\pi f_0 k)\} \\ &= \sup_{\phi}\frac{1}{\lceil F_s/B\rceil}\sum_{k=1}^{\lceil F_s/B\rceil}\sqrt{\varepsilon}\cos(2\pi(f_d - f_0)k + \phi) \end{aligned} \tag{15}$$

where the high frequency part $f_d + f_0$ has been filtered out; $\lceil\cdot\rceil$ is to calculate the minimum integer that larger than the input value; $\sup_\phi$ is used to calculate the super-bound of the function with the parameter of ϕ, i.e., calculate the envelop of the input signal. When $f_d = f_0$, then (15) can be written as

$$\mathbf{s}_F(0) = \sup_{\phi}\sqrt{\varepsilon}\cos(2\pi k 0 + \phi) = \sqrt{\varepsilon} \tag{16}$$

When $f_d \neq f_0$, the result of (15) is

$$\begin{aligned} \mathbf{s}_F(0) &= \sup_{\phi}\frac{1}{\lceil F_s/B\rceil}\sum_{k=1}^{\lceil F_s/B\rceil}\sqrt{\varepsilon}\cos(2\pi B k + \phi) \\ &= \sup_{\phi}\mathop{\mathbb{E}}_{\theta\in[0\sim 2\pi]}[\cos(\theta + \phi)] \approx 0 \end{aligned} \tag{17}$$

where $\mathbb{E}_\theta[\cdot]$ approximates to obtain the mean value of $cos(\cdot)$ in a whole period, which is equal to zero and irrelevant to the phase ϕ. From (16) and (17), we can know that the same local template can obtain distinct values (0 and $\sqrt{\varepsilon}$) when the input signal is different. Using these distinct values will map the input signal to the information domain. Define the decision vector for frequency demodulation is $\mathbf{s}_F$, and $\mathbf{s}_F = [\sqrt{\varepsilon}, 0]^T$ when the information bit $b = 0$; $\mathbf{s}_F = [0, \sqrt{\varepsilon}]^T$ when $b = 1$, ε is the transmitted bit energy.

Only using frequency for demodulation may suffer significant performance loss due to the multi-path fading effect. For microwave signal, the channel can be assumed as flat for several KHz, but not for acoustic signal. Frequency f_1 and f_0 may suffer different attenuation, such channel fading effect should be considered for better demodulation performance. To deal with such problem, the amplitude difference of symbol '1' and '0' can help to compensate the performance loss when we perform joint frequency and amplitude detection.

For the amplitude detection when one symbol shows significant small-scale fading, the difference between the received signal amplitude and the reference value can be used as the decision vector as

$$\bar{\mathbf{s}}_A = \left[|\mathbf{s}_A(b) - s_A^{ref}(0)|, |\mathbf{s}_A(b) - s_A^{ref}(1)|\right]^T \tag{18}$$

where $s_A^{ref}(0)$ and $s_A^{ref}(1)$ are the prior information of amplitude for symbol '0' and '1', $\mathbf{s}_A(b)$ is the amplitude of the received signal. The parameters of $s_A^{ref}(0)$ and $s_A^{ref}(1)$ can be obtained by using transmit reference (TR), e.g., transmit two bit of '1' and '0' at the beginning of the beacon period. At the receive side, we can assume that these TR bits suffer the same attenuation as other bits in the same beacon period. Calculating the amplitude of TR bit provide $s_A^{ref}(0)$ and $s_A^{ref}(1)$ in amplitude demodulation of (18). Such TR scheme is a simplified method for channel estimation, and being used as fingerprint to characterize the channel effect. For the decision vector, and $\bar{\mathbf{s}}_A = [\sqrt{\varepsilon_\delta}, 0]^T$ when the information bit $b = 0$; $\bar{\mathbf{s}}_A = [0, \sqrt{\varepsilon_\delta}]^T$ when $b = 1$, ε_δ is the energy difference.

We can obtain dynamic decision rule as

$$\mathbf{y}_{joint} = \frac{\mathbf{s}_F(1) - \mathbf{s}_F(0)}{\sqrt{\varepsilon}}(1-\mu) + \frac{\bar{\mathbf{s}}_A(1) - \bar{\mathbf{s}}_A(0)}{\sqrt{\varepsilon_\delta}}\mu \gtrless_0^1 0 \tag{19}$$

where μ is the weighting coefficient. The joint demodulation of (19) is only needed when one symbol suffers significant attenuation, otherwise, using the amplitude in demodulation with no envelop distinction may provide negative effects. Thus, we set μ according to

$$\mu = \begin{cases} \frac{\left|s_A^{ref}(0) - s_A^{ref}(1)\right|}{(s_A^{ref}(0) + s_A^{ref}(1))}, & \xi \neq 0 \\ 0, & \xi \approx 0 \end{cases} \tag{20}$$

where $\xi = (s_A^{ref}(0) - s_{noise}) \cdot (s_A^{ref}(1) - s_{noise})$, $s_{noise} = \sqrt{2/\pi}\hat{\sigma}$ is the amplitude value of the folded noise. (20) shows that when one of the symbols being serious attenuated, joint demodulation should be used.

V. Performance Evaluation

A. Experiment Setup

We conduct the measurement in an office environment to test the Signal-to-Noise Ratio (SNR), Bit-Error-Rate (BER) and TOA Normalized Mean Square Error (NMSE) at different communication distances to evaluate the system performance of TOA estimation and communication. We move the anchor node away from the microphone from 0.254m to 7.366m, and conducted measurement for every 0.1m with more than 400s sampling data acquired in each measurement, $N_p = 400$. The beacon period is $T_p = 0.9710s$, symbol duration is $T_s = 0.0205s$. The symbol number in one beacon period 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 $F_s = 44.1KHz$, with the symbols modulated in the frequency of $f_1 = 17.72KHz$ and $f_2 = 19.2KHz$.

To observe the signal attenuation with distance change, we measure the SNR value of symbol '1' and '0' with respect to the communication distance. The empirical SNR calculation equation used is $SNR(b) = 20\log_{10}[(S(b)/s_{noise})]$, where $S(b)$ is the signal mean absolute value for information bit b, $s_{noise} = \sqrt{2/\pi}\hat{\sigma}$ is the mean absolute value of noise.

For the evaluation of the TOA estimation and communication performance, three different methods are evaluated and compared. The first one uses the conventional two-step TOA estimation without threshold optimization and frequency demodulation ('FD') method; the second one uses the optimized threshold in TOA estimation called "OTH-FD"; the last one uses both the TOA optimized threshold and our proposed dynamic demodulation (DD), called "OTH-DD".

The Normalized Mean Square Error (NMSE) is used in this paper to characterize the accuracy of the TOA result by using $NMSE = [(\hat{\tau}^{TOA} - \tau^{TOA})/\tau^{TOA}]^2$. We measure the BER value by comparing the demodulated information bit ($\hat{b}_i$) to the real transmitted data (b_i) by $BER = \sum_{i=0}^{N_s-1} |\hat{b}_i - b_i|/N_s$.

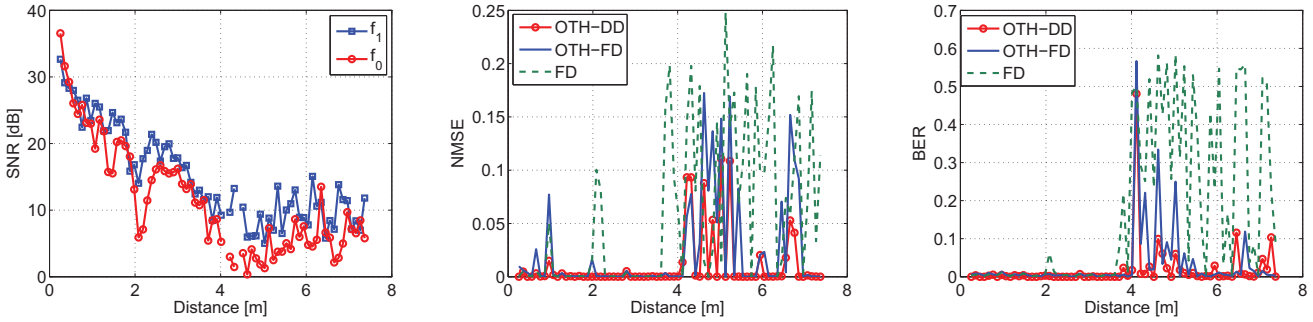
B. Experimental Results

The measurement results of SNR degradation vs. distance are shown in Fig. 2a. The results show that the small-scale fading of the acoustic signal is strong, f_0 and f_1 suffered different attenuation. For most distances, the attenuation of the frequency f_0 is stronger than f_1 due to its slightly higher frequency.

The results of NMSE with respect to distance are shown in Fig. 2b. The x-coordinate is the distance between the transmitter and the receiver, y-coordinate is the measured NMSE. For smaller distance ($< 4.4m$), the NMSE value for the three methods are all very low, showing that the TOA estimation is very accurate when SNR is strong. For larger distance ($> 4.4m$), the estimation error is increased due to the attenuation of the signal. Using our proposed optimized TOA threshold, "OTH-FD" and "OTH-DD" achieved better performance than "FD" case. While using dynamic demodulation, some erroneous TOA estimation results can be identified and filtered by its BER value, thus making the TOA estimation accuracy of "OTH-DD" slightly better than "OTH-FD" due to its better BER performance.

The BER experiment results are shown in Fig. 2c with its y-coordinate as the measured BER. Using our optimized TOA threshold, "OTH-FD" and "OTH-DD" achieve better performance than "FD" case. While using dynamic demodulation can even lower the error rate as shown in "OTH-DD".

For Fig. 2b and Fig. 2c, the most interesting point is near $4.11m$, the performance of TOA estimation and communica-



(a) SNR (b) NMSE (c) BER

Fig. 2. The SNR(a), NMSE(b) and BER(c) measurement results with respect to the distance

TABLE I
PERFORMANCE COMPARISON WITH RESPECT TO DIFFERENT METHODS UNDER SPECIFIC PROBABILITY

Methods	Metrics	70%	80%
OTH-DD	NMSE	0.0005	0.0031
	Range Error (m)	0.0864	0.5270
	BER	0.0055	0.0099
OTH-FD	NMSE	0.0041	0.0259
	Range Error (m)	0.69	4.4
	BER	0.0083	0.0152
FD	NMSE	0.0812	0.1142
	Range Error (m)	13.8107	19.4062
	BER	0.2303	0.4267

tion for all the three methods are really worse at that point. From the echo in that distance, it shows that the waveform has been seriously attenuated both for symbol '1' and '0'. There may be several multi-path signals arrived at the receiver with negative phase and canceling with each other. Such dead zone may affect the final position result, but using more redundant anchor nodes can compensate the performance loss.

To evaluate the performance at specific probability, we write the maximum TOA estimation error and BER under a given probability as shown in Table. I. The metric of Range Error is calculated by NMSE to characterize the ranging accuracy. This table illustrates that 70% or 80% of the total results are less than the given value, e.g., the value 0.0005 means 70% of the NMSE results are less than 0.0005. From Table. I, we know that the maximum range error is less than 0.0864m with 70% probability when using "OTH-DD", such ranging accuracy is much higher than other existing schemes based on pervasive hardware and sufficient to guarantee a precise indoor localization result.

VI. CONCLUSION

We developed a TOA estimation and communication scheme that utilizes pervasive microphone channels and low-complex anchor nodes. Using the acoustic signal for ranging and localization can achieve high accuracy with no additional hardware requirement on users. To facilitate the ranging and communication under the restriction of the speaker/microphone pair, an optimized TOA estimation method and a dynamic demodulation scheme have been proposed. The experiment results show that the TOA ranging accuracy can be achieved within 0.0864 meters with 70% probability; the communication BER can be less than 0.55% for 70% distances in the range of $0.254 \sim 7.366$ meters. Such results of ranging and communication are sufficient for the accurate indoor localization applications.

REFERENCES

[1] 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*, vol. 22, no. 4, pp. 54–69, 2005.

[2] D. Porcino and W. Hirt, "Ultra-wideband radio technology: potential and challenges ahead," *Communications Magazine, IEEE*, vol. 41, no. 7, pp. 66–74, 2003.

[3] H. Yin, Z. Wang, L. Ke, and J. Wang, "Monobit digital receivers: design, performance, and application to impulse radio," *Communications, IEEE Transactions on*, vol. 58, no. 6, pp. 1695–1704, 2010.

[4] I. Guvenc and Z. Sahinoglu, "Multiscale energy products for toa estimation in ir-uwb systems," in *Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE*, vol. 1. IEEE, 2005, pp. 5–pp.

[5] 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*. IEEE, 2011, pp. 1–5.

[6] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. Martin, "Detecting driver phone use leveraging car speakers," in *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM, 2011, pp. 97–108.

[7] H. Balakrishnan, R. Baliga, D. Curtis, M. Goraczko, A. Miu, N. Priyantha, A. Smith, K. Steele, S. Teller, and K. Wang, "Lessons from developing and deploying the cricket indoor location system," *Preprint*, 2003.

[8] 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.

[9] A. Smith, H. Balakrishnan, M. Goraczko, and N. Priyantha, "Tracking moving devices with the cricket location system," in *Proceedings of the 2nd international conference on Mobile systems, applications, and services*. ACM, 2004, pp. 190–202.

[10] I. Guvenc and Z. Sahinoglu, "Threshold-based TOA estimation for impulse radio UWB systems," in *IEEE International Conference on Ultra-Wideband*, 2005, pp. 420–425.

[11] J. Patel and C. Read, *Handbook of the normal distribution*. CRC, 1996, vol. 150.

[12] 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.