BeepBeep: A High-Accuracy Acoustic-Based System for Ranging and Localization Using COTS Devices

CHUNYI PENG, University of California, Los Angeles
GUOBIN SHEN, Microsoft Research Asia
YONGGUANG ZHANG, Microsoft Research Asia

We present the design and implementation of BeepBeep, a high-accuracy acoustic-based system for ranging and localization. It is a pure software-based solution and uses the most basic set of commodity hardware – a speaker, a microphone, and some form of interdevice communication. The ranging scheme works without any infrastructure and is applicable to sensor platforms and commercial-off-the-shelf mobile devices. It achieves high accuracy through three techniques: *two-way sensing*, *self-recording*, and *sample counting*. We further devise a scalable and fast localization scheme. Our experiments show that up to one-centimeter ranging accuracy and three-centimeter localization accuracy can be achieved.

Categories and Subject Descriptors: C.3.3 [**Special-Purpose and Application-Based Systems**]: Real-Time and Embedded Systems; C.5.3 [**Computer System Implementation**]: Microcomputers—*Portable devices*

General Terms: Algorithms, Design, Performance

Additional Key Words and Phrases: Ranging, acoustic ranging, time-of-arrival, localization, ranging and localization systems

ACM Reference Format:
Peng, C., Shen, G., and Zhang Y. 2012. BeepBeep: A high-accuracy acoustic-based system for ranging and localization using COTS devices. ACM Trans. Embedd. Comput. Syst. 11, 1, Article 4 (March 2012), 29 pages.
DOI = 10.1145/2146417.2146421 http://doi.acm.org/10.1145/2146417.2146421

1. INTRODUCTION

In this article, we present the design and implementation of BeepBeep, an acoustic-based system that offers high-accuracy ranging and localization services. Our system works with only the most basic set of commodity hardware: a speaker, a microphone, and some form of interdevice communication. Such a system, if feasible, will have many desirable features and be widely applicable in many sensing and mobile applications. This is because the hardware requirements can be considered as a common dominator of many sensor platforms and mobile devices, including many commercial off-the-shelf (COTS) devices like cell phones, PDAs, MP3 players, etc. Take ranging service as an example. Compared to alternatives that require special-purpose hardware (such as Girod et al. [2006] and Youssef et al. [2006]) or pre-existence of location infrastructure [Harter et al. 1999], a commodity-based solution will obviously have wider applications

An early version of this article appeared in *Proceedings of the 5th International Conference on Embedded Networked Sensor systems (SenSys'07)*.
Authors' addresses: C. Peng, WiNG Lab, UCLA Computer Science Department, 4681 Boelter Hall, Los Angeles, CA 90095-1596; G. Shen and Y. Zhang, Microsoft Research Asia, Building 2, No. 5 Danling Street, Haidan District, Beijing, 100080, China; email: chunyip@cs.ucla.edu.

DOI 10.1145/2146417.2146421 http://doi.acm.org/10.1145/2146417.2146421

and cost less. For the same reason, we further desire a solution that can be implemented in software, preferably entirely in the user space. Once high-accuracy ranging is available, solutions to localization can be devised thereafter. Many localization systems use dependable ranging as the basic building block [Borriello et al. 2005].

High-accuracy ranging, as well as localization, is typically achieved through measuring time-of-arrival (TOA) information of acoustic or radio signals [Harter et al. 1999; Priyantha et al. 2000; Whitehouse and Culler 2002; Sallai et al. 2004; Girod et al. 2006; Youssef et al. 2006]. The distance is thus the product of the signal speed and the time of flight of the signal traveling between two devices. Consequently, the ranging accuracy depends on the signal speed and the precision of the TOA measurement. In practice, acoustic signals are typically chosen because of their relative slow speed. TOA measurement is often taken with both sides recording a timestamp of their respective local clock at the moment the signal is emitted or received. The main roadblock to achieve high accuracy for ranging is the TOA measurement scheme.

There are three intrinsic uncertainty factors in the TOA measurement scheme that leads to the ranging inaccuracy: the possible clock skew and drift between devices, the possible misalignment between the sender timestamp and the actual signal emission, and the possible delay of a sound signal arrival being recognized at receiver. In general, many factors can cause the latter two uncertainties in a real system, such as the lack of real-time control, software delay, interrupt handling delay, system loads, etc. These uncertainties, if not controlled, can seriously affect the ranging accuracy. For example, our tests on two COTS mobile devices reveal that these two delays can easily add up to several milliseconds on average, which translates to several feet of ranging error (Section 2).

It is therefore challenging to provide high-accuracy ranging in a software-only solution using only the minimum commodity hardware set we specified earlier. For the solution to be applicable to COTS mobile devices, there are additional constraints. We cannot assume we have a real-time operating system or we are able to change kernel or driver. In fact, many COTS devices such as cell phones are built on closed platforms, and many often have operator-imposed locks that prevent changing OS. We will have to implement the entire ranging system in user-space. It is clear that this timestamp-based approach will not be able to provide the high accuracy we desire.

In this article, we have developed a novel high-accuracy acoustic ranging mechanism and further implemented it in a pure software-based ranging system on COTS mobile devices. We also use the ranging mechanism as the building block to offer localization functionality. The key idea for achieving high accuracy is our innovative use of three techniques: *two-way sensing*, *self-recording*, and *sample counting*. First, the two devices will each in turn emit a specially designed sound signal, called a "Beep", within one second. Meanwhile, each device will also record a few seconds of continuous sound from its microphone. Each recording should then contain exactly two Beep signals picked up by its microphone, one emitted from the other device and one from itself. Next, each device will count the number of sound samples between these two Beeps, and divide the number by the sampling rate to get the elapsed time between the two TOA events. The devices further exchange the elapsed time information with each other. The differential of these two elapsed times represents the sum of the time of flight of the two Beeps and hence the two-way distance between the two devices. We called our system "BeepBeep" because of the signature double Beep sounds during a ranging session.

By using sample counting instead of timestamping, our mechanism mitigates all the uncertainties listed earlier, and avoids the source of inaccuracies found in traditional timestamp approaches. In fact, our mechanism has no notion of local clock or timestamp at all. The granularity of our TOA measurement is limited only to the sound sampling rate. Under today's prevailing hardware standard of 44.1KHz, our mechanism can have

a ranging accuracy of 0.8cm. As far as we know, this is the best ever achieved using only commodity hardware (speaker and microphone) on COTS mobile devices. It is also comparable to, if no better than, the best results ever reported in the literature that use special hardware design or complex signal processing.

Once ranging is available, we devise a new solution to localization. Our localization scheme is scalable to the number of anchors because the acoustic signaling overhead remains constant. Instead, a naive solution based on pairwise ranging between the sensor and each of the multiple anchors incurs a signaling overhead in the order of the anchor population. We also use linear matrix processing techniques to speed up the localization computation.

In summary, the described BeepBeep system makes three contributions. It proposes novel schemes for ranging and localization that overcome several technical roadblocks of existing TOA-based solutions. To this end, we identified the three major uncertainties common to any time-of-arrival based ranging system and evaluated them on COTS mobile devices. We then proposed the BeepBeep ranging mechanism that overcomes all these uncertainties. Our localization is scalable and fast since the signaling overhead does not grow with the anchor population and linear matrix processing techniques are used. Moreover, we designed and implemented the BeepBeep system, purely in software. We have systematically evaluated the system and our design choices under typical indoor and outdoor environments using COTS mobile devices. We have achieved centimeter accuracy in ranging, the best ever reported in the literature.

The rest of the article is organized as follows: we identify the challenges for TOA-based systems in Section 2. The base technology of BeepBeep and the solution to scalable localization are presented in Sections 3 and 4. We describe the system implementation and detailed design in Section 5. Section 6 provides experimental evaluation. Section 7 reviews the related work, and Section 8 shows two application examples using BeepBeep, followed by conclusions in Section 9.

2. CHALLENGES OF TOA-BASED SYSTEMS

High-accuracy ranging typically serves as the foundation to further achieve localization. We hence start our system with solutions to ranging. The popular approach to high accuracy ranging is based on measuring time-of-arrival (TOA) information of acoustic or radio signals [Harter et al. 1999; Priyantha et al. 2000; Sallai et al. 2004; Girod et al. 2006; Youssef et al. 2006]. In this section, we elaborate on the technical challenges of such TOA-based systems and critique on the various solutions proposed in the literature.

The TOA-based approach to high-accuracy ranging works as follows. Given the pair of devices, the system estimates the distance D between the sender and the receiver to be the product of the time-of-flight, that is, the time (Δt) it takes a signal such as sound, radio wave, or light to reach the receiver, and the propagation speed c of the signal, which is usually assumed to be a constant known a priori.

$$D = c \cdot \Delta t \tag{1}$$

Given the requirement on the desired precision, acoustic signal is usually chosen because the speed of radio or light signal is so fast that a small timing error would lead to an unacceptably large ranging error. Even though the relatively slower acoustic signal is chosen, the precision requirement on TOA estimation is still very stringent. For example, one millisecond error in TOA estimation will translate to more than 30-centimeter error in the ranging result.

Traditionally, TOA measurement is done with both sides taking a timestamp of their respective local clock at the moment the signal is emitted or received. There are several intrinsic uncertainties in this process that will contribute to the TOA measurement

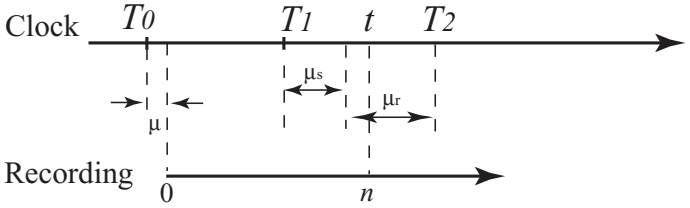


Fig. 1. Event time line in the uncertainty experiment. The upper axis represents system clock (T_0: start sound recording, T_1: start sound playing, T_2: realize a incoming signal), and the bottom one represents the number of recording samples(0: the first sample, n: the sample number atg TOA).

error. The first one is clock synchronization uncertainty (μ_c): the possible clock skew and drifting between the two devices. To address this problem, many solutions have been proposed in the literature. Some relied on GPS [Enge and Misra 1999] for time synchronization and some others chose to work around by using round-trip time measurement (assuming symmetric propagation path) so that all time readings refer to the same clock [Webr and Lanzl 1998]. Yet most solutions have resorted to dedicated mechanisms [Priyantha et al. 2000; Kushwaha et al. 2005; Youssef et al. 2006; Girod et al. 2006].

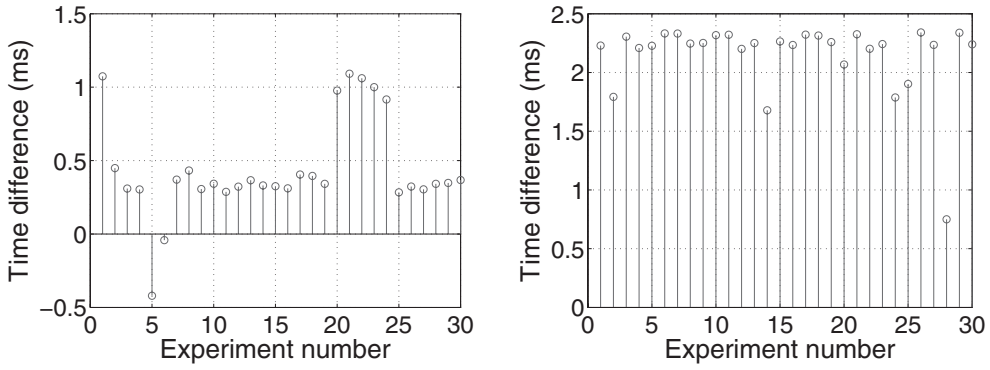
The second uncertainties is the sending uncertainty (μ_s): the possible misalignment between the timestamp and the actual signal emission. For example, there is often a small yet arbitrary delay after an output command is issued till the sound actually comes out from the speaker. Similarly, the third uncertainty is the receiving one (μ_r): the possible delay of a sound signal arrival being recognized. In general, many factors can cause these two uncertainties in a real system, such as the lack of real-time control, software delay, interrupt handling delay, system loads, etc. There has been little work in addressing the sending and receiving uncertainties in software. Most previous work managed to minimize them by resorting to customized hardware design so that the system can precisely control and obtain the exact instant when a signal is sent or received [Harter et al. 1999; Priyantha et al. 2000]. This is clearly inapplicable if we desire a software solution and only use commodity hardware.

To understand how large these two uncertainties can be in a general purpose mobile device, we conduct an experiment using a COTS mobile phone, the HP iPAQ rw6828. The experiment is designed to find out a lower bound for $\mu_s + \mu_r$ if a TOA measurement is done in software. To make the signal time of flight negligible, we put the speaker and microphone together. We wrote a program to do the following during an experiment run. The program first takes a timestamp at time T_0 and starts sound recording. It then takes another timestamp at time T_1 and immediate sends out a sound signal. When the sound comes out of its speaker, the recording should be able to pick it up from its microphone. Momentarily the program examines the recording and finds the index n of the recorded signal. Figure 1 illustrates the time line of these events.

From the figure, we can see that we assume the sound sampling in the recording actually started after an unknown delay μ. Immediately after T_1, there is a sending delay μ_s when the sound actually emits from the speaker and arrives at the microphone. Then there is a receiving delay μ_r before a normal TOA measurement can realize an incoming signal, that is, T_2, which can be no earlier than time t, the time when the sound sampling picks up the signal. From the time relationship, we can derive the following equation (f_s is the sampling rate):

$$\mu_s + \mu_r > T_0 + \mu + n/f_s - T_1 > T_0 + n/f_s - T_1. \tag{2}$$

Here, $T_0 + n/f_s - T_1$ is a lower bound estimation of $\mu_s + \mu_r$. We repeated the experiment many times and plot the values of this estimation in Figure 2.



(a) CPU idle (no load) (b) CPU heavily loaded

Fig. 2. A lower bound estimation of the sending and receiving uncertainties.

The results indicate that $\mu_s + \mu_r$ appears to be very random and affected heavily by the CPU load. Both the average and deviation increases when the load becomes heavy, such as playing a video, even if we give the test program the highest priority. In any case, this study shows that the uncertainties easily add up to several milliseconds and translate to several feet of ranging error when the TOA measurement is done in software.

Many existing solutions based on the TOA approach handle the above uncertainties by using special hardware (such as ENSBox [Girod et al. 2006] and Cricket [Priyantha et al. 2000]) or requiring infrastructure support (such as BAT [BatSystem 2003]). Other solutions working with COTS hardware can only offer low-accuracy ranging (such as WALRUS [Borriello et al. 2005], RADAR [Bahl and Padmanabhan 2000] and HORUS [Youssef and Agrawala 2005]). Our BeepBeep solution aims to provide a software solution that can eliminate these uncertainties and achieve high-accuracy ranging and localization on COTS devices.

3. ACOUSTIC RANGING

Unlike existing ranging systems, the BeepBeep system is purely a software solution that does not require specialized hardware design or any modification to the commercial operating system. Our system can well handle all the three uncertainties mentioned previously, that is, clock synchronization, sending and receiving uncertainty. In this section, we explain how the basic BeepBeep scheme works and elaborate on the reasons that lead to the high-precision results.

3.1. Basic scheme

We start from the basic ranging procedures with only two devices, say A and B, and will extend this to multiple devices in the next section.

The basic ranging scheme takes three steps, as shown in Figure 3. In the first step, a two-way sensing is performed. Assume both devices are in recording state. Device A first emits a sound signal through its speaker S_A and then device B emits another sound signal back. Both signals will be recorded by its own microphone (self-recording) as well as the other peer. In the second step, both devices examine their recorded data and locate the sample points when previously emitted two signals arrived. We denote the time difference between these two signals as *elapsed time between the two*

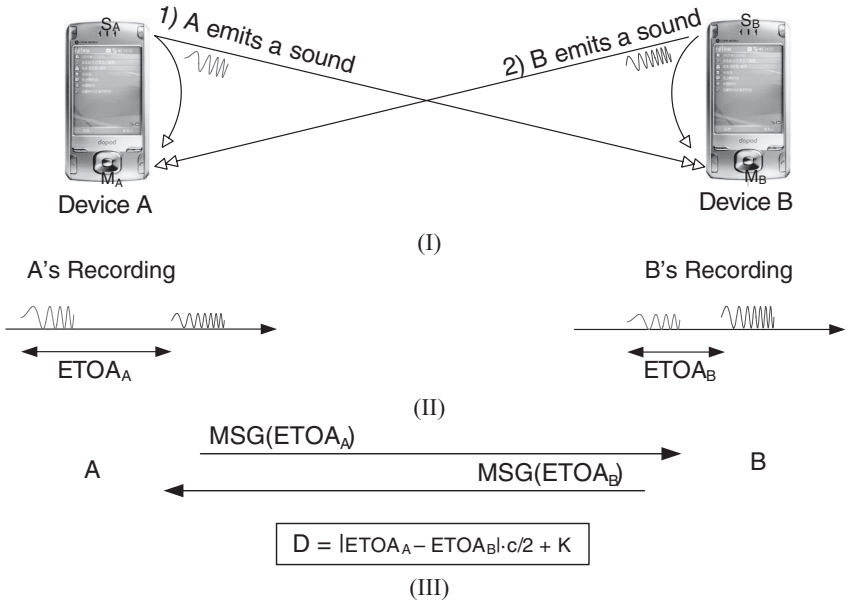


Fig. 3. Illustration of the BeepBeep ranging procedure: (I)two-way sensing stage: both devices are in the recording state; A emits a sound and then B emits another sound; (II)ETOA calculation stage: both detect TOA in recording data and calculate local ETOA; (III) ETOA exchange stage: both exchange ETOA and calculate distance between them. c and K are constants in our derivation.

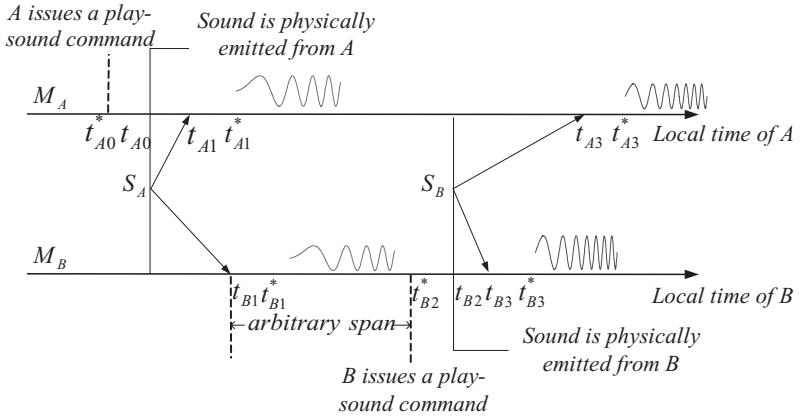


Fig. 4. Illustration of event sequences in the BeepBeep ranging procedure. No timestamping is required in our mechanism.

time-of-arrivals (ETOA). [1] The two devices will exchange their locally measured ETOA and in the final step, the distance between the two devices can be simply computed based on these two values.

Figure 4 illustrates the timing relation among events when doing two-way sensing in the first stage. Two time lines are drawn in the figure with the upper one presenting

[1]We use the term ETOA here in order to differentiate from the well-defined term DTOA (differential times of arrival) or TDOA (time differences of arrival) that usually refers to the differential between two TOAs measured at two different receivers using the same sound source.

the local time of device A and the bottom one the local time of device B. We denote t^*_{A0} the time when device A instructs its speaker to emit the sound signal. However, due to the sending uncertainty (here, $\mu_s = t_{A0} - t^*_{A0}$), the actual time when the speaker physically emits may be t_{A0}. The time of the signal arriving at microphones of device A and B are marked as t_{A1} and t_{B1}, respectively. Again, due to the receiving uncertainty, applications on device A and B may obtain these signal data only at time t^*_{A1} and t^*_{B1}. Similarly, we denote t^*_{B2} and t_{B2} as the time when device B instructs to send out a sound signal and when the signal is physically out; t_{A3} and t_{B3} as the time when the signal from device B arrives at microphones of device A and B; and t^*_{A3} and t^*_{B3} as the time when the applications on device A and B conclude the arrival of the signal data.

We denote $d_{x,y}$ as the distance between the device x's speaker to device y's microphone. It is clear that we have

$$d_{A,A} = c \cdot (t_{A1} - t_{A0}) \tag{3}$$

$$d_{A,B} = c \cdot (t_{B1} - t_{A0}) \tag{4}$$

$$d_{B,A} = c \cdot (t_{A3} - t_{B2}) \tag{5}$$

$$d_{B,B} = c \cdot (t_{B3} - t_{B2}) \tag{6}$$

where c is the speed of sound. Then, the distance between the two devices D can be approximated as

$$\begin{aligned} D &= \frac{1}{2} \cdot (d_{A,B} + d_{B,A}) \\ &= \frac{c}{2} \cdot \big((t_{B1} - t_{A0}) + (t_{A3} - t_{B2})\big) \\ &= \frac{c}{2} \cdot (t_{B1} - t_{B2} + t_{B3} - t_{B3} + t_{A3} - t_{A0} + t_{A1} - t_{A1}) \\ &= \frac{c}{2} \cdot ((t_{A3} - t_{A1}) - (t_{B3} - t_{B1}) + (t_{B3} - t_{B2}) + (t_{A1} - t_{A0})) \\ &= \frac{c}{2} \cdot ((t_{A3} - t_{A1}) - (t_{B3} - t_{B1})) + \frac{1}{2}(d_{B,B} + d_{A,A}) \end{aligned} \tag{7}$$

In Eq. (7), the two terms of $d_{A,A}$ and $d_{B,B}$ are the distances between the speaker and the microphone of device A and device B, respectively. They only depend on the speaker/microphone placement and remain constant to a given device and can be measured a priori. Therefore, the distance between two devices is determined solely by the first two terms, which are actually the ETOA values measured on device A and B, respectively. Note that ETOA is calculated by each individual device independently, that is, without referring any timing information on the other device, so that no clock synchronization between devices is needed. Moreover, due to the self-recording strategy, all time measurements are associated with the arrival instants of the sound signals, and therefore, the sending uncertainty is also removed. In the next subsection, we show how precise ETOA can be obtained.

3.2. ETOA Determination

In a typical computer system, obtaining the exact time instance when the signal arrives is difficult due to the indeterministic latency introduced by hardware and software (receiving uncertainty). In our design, this issue is resolved by not referring to any local clock while inferring timing information directly from recorded sound samples.

Realizing that the received sound signal is always sampled at a fixed frequency (represented by f_s) by the A/D converter, we can therefore directly obtain ETOA by counting the sample number between the two TOAs of signals from recorded data, without dealing with the local clock of the end system. In other words, we do not rely

on the end system to tell the timestamp that it "thinks" the signal has arrived. Rather, we depend on the fidelity of the recording module. Since all the sound signals are recorded, we only need to check the recorded data and identify the first sample point of each signal. Then, ETOA is obtained by counting the number of samples between the two corresponding first samples.

Note that this strategy has another preferable property to eliminate the necessity of instantaneous signal detection and may shift the signal detection task out of the sensing stage. Indeed, what we need to do is to simply sample and record the received signal and conduct signal detection at a later time or even offline. As a consequence, more complex signal processing techniques can be applied in our case while not requiring special hardware support or critical speed optimization.

With sample counting, Eq. (7) can be rewritten as

$$D = \frac{c}{2} \cdot \left(\frac{n_{A3} - n_{A1}}{f_{sA}} - \frac{n_{B3} - n_{B1}}{f_{sB}} \right) + K, \tag{8}$$

where n_x denotes the index of the sample point at instant t_x, f_{sA} and f_{sB} are the sampling frequency of device A and B, respectively, and $K = \frac{1}{2} \cdot (d_{B,B} + d_{A,A})$ is a constant. In the rest of the article, we will assume the the sampling frequency to be 44.1 kHz unless explicitly noted since the 44.1 kHz sampling frequency is the basic, de facto standard that almost every sound card supports. In this case, we have $f_{sA} = f_{sB}$, and Eq. (8) can be simplified to

$$D = \frac{c}{2 \cdot f_s} \cdot \big((n_{A3} - n_{A1}) - (n_{B3} - n_{B1}) \big) + K \tag{9}$$

By using sample counting instead of timestamping, our mechanism avoids intrinsic receiving uncertainties due to the lack of real-time control, software delay, interrupt handling delay, system loads in a real system, found in traditional timestamp approaches. In fact, our mechanism has no notion of local clock or timestamp at all. This solution leverages the fact the independent A/D convertor works at a fixed and stable sampling frequency, thus generating high-accuracy time information without unknown delay factors from the phone's operating system. From Eq. (9), the measurement granularity is positively proportional to the sound speed c and inversely proportional to the sampling frequency f_s. Take a typical setting of $c = 340$ meters per second and $f_s = 44.1$ kHz, the distance granularity is then about 0.77 centimeters. The granularity will be further improved if higher sampling frequencies can be afforded.

Therefore, BeepBeep can well handle the three uncertainties to achieve a high accuracy ranging. In addition, the impact of other factors is also quite small. From Eq. (9), we see that there are three possible sources of errors relating to the three parameters, that is, sound speed c, sampling frequency f_s, and various sample indices n_t (i.e., TOA detection). For instance, the propagation speed of sound in the air varies with temperature and humidity, and the sampling frequency may also drift. Fortunately, their impacts are usually negligible in practice and can be mitigated by taking temperature and humidity into consideration using well-established sound speed models and by shortening the sensing interval, respectively. To achieve accurate TOA results, we need handle ambient noise, multipath effects, signal distortion caused by COTS hardware limitation (e.g., microphone and the speaker usually have a very limited spectrum band), which will be described in the algorithm for signal design and detection in Section 5. Consequently, BeepBeep can achieve high-accuracy ranging through novel software techniques.

4. FROM RANGING TO LOCALIZATION

In this section, we describe the BeepBeep localization scheme. Given a set of N anchors (say, $N \geq 4$) with known positions, localization seeks to determine the exact location for a target of interest. In the literature, most localization methods are ranging based. Such methods measure the distance of the target to each of the anchors via ranging. Once all such distance measurements are available, the location of the target can be computed through standard techniques such as trianglization.

BeepBeep localization is designed using its ranging scheme as the basic building block. Given a target of interest, BeepBeep first determines its distances to the multiple anchors via the novel and scalable *Multiple BeepBeep Ranging* (MBR) scheme. Trilateration like techniques are then used to compute the exact position of the target. In the rest of this section, we describe our MBR and trilateration-based localization in details.

4.1. Multiple BeepBeep Ranging Scheme

MBR seeks to measure the distance from the target sensor to each of N anchors. Our goal is to design a multiranging solution that is scalable to the number of anchors. In the BeepBeep context, the signaling overhead when exchanging the Beep signals should not grow linearly with the number of anchor devices.

4.1.1. Naive MBR. A naive MBR scheme applies the BeepBeep ranging technique between the sensor and all the anchor devices. In this solution, pairwise ranging is performed between the sensor and each anchor. It is a sequential procedure in multiple rounds. In each round, both the sensor and the selected anchor will each emit a beep signal once so that their distance can be measured by the two beep signals. This interactive process proceeds until all the N anchors have completed the ranging measurement at least once.

This straightforward MBR proposal works but does not scale. Note that the measurement of the distance between two sensors requires both participate in the process. Therefore, $2N$ or at least $N+1$ beeps are required to sequentially measure the target's distances to N anchors. In addition, the time interval between two consecutive chirps has to be long enough to avoid confusing the different emitters. In summary, the sequential procedure to measure the N ranging distances takes more time and is more vulnerable to ranging errors.

4.1.2. Novel Scalable MBR. We propose a novel MBR that eliminates the scaling issue incurred by the sequential ranging scheme. Our solution simultaneously estimates the target sensor's distances to all anchors. No matter how many anchors we use for localization, only two chirps are required. These two chirps will be used by both the sensor and multiple anchors to estimate the ranging distances in parallel, thus speeding up the localization process.

Figure 5 illustrates our MBR scheme for two anchors. The scheme also works with multiple anchors. Assume that anchor positions (i.e., A and B) are known a priori and all nodes turn on their recording functionality. Anchor A first emits a beep signal and then node M emits another beep. Both beeps will be captured by all listening nodes, for example, A, B, and M. These nodes independently detect the elapsed time of arrivals (ETOA) between the two beeps. Figure 6 shows the time sequence of the whole procedure: a beep is emitted from A at T_0, and arrives at nodes A, B and M at T_{A1}, T_{B1} and T_{M1}, respectively. Similarly, M emits the second beep at T_1, which reaches A, B and M at T_{A2}, T_{B2} and T_{M2}, respectively. Since it is a typical BeepBeep procedure between nodes A and M, we have

$$D_{AM} + D_{MA} = D_{AA} + D_{MM} + c \cdot (T_{A2} - T_{A1}) - c \cdot (T_{M2} - T_{M1}), \tag{10}$$

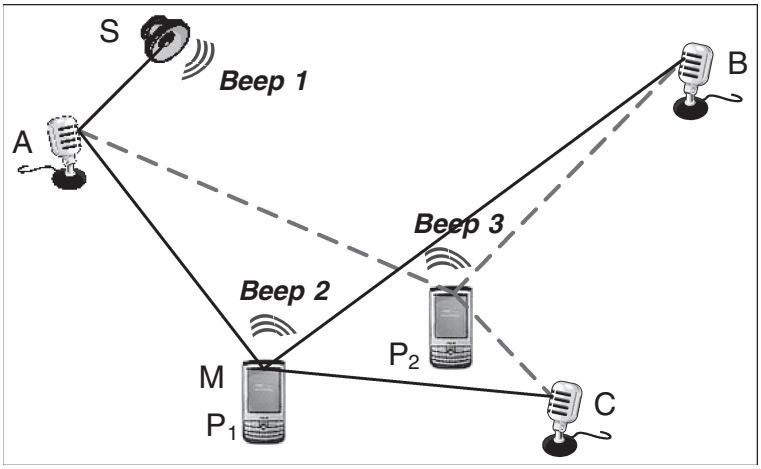


Fig. 5. Illustration of MBR Scheme.

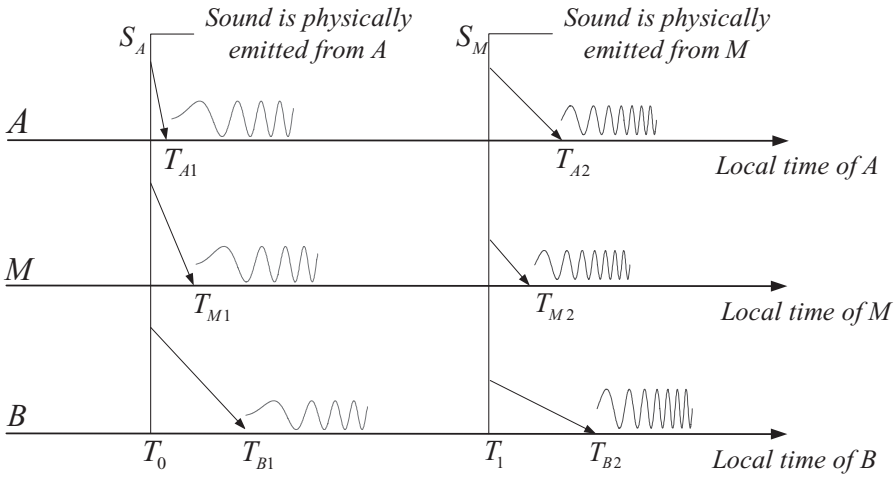


Fig. 6. Illustration of Event Sequences.

that is, the distance between anchor A and the target sensor M can be calculated by ETOA at nodes A and M.

Our MBR scheme is able to measure the distances from M to any other anchor (say, B) via passive listening of the two beeps, without extra Beepbeep signals. We now explain how it works. It is easy to see that

$$T_{B1} = T_0 + D_{AB}/c, \tag{11}$$
$$T_{B2} = T_1 + D_{MB}/c, \tag{12}$$
$$T_{M1} = T_0 + D_{AM}/c, \tag{13}$$
$$T_{M2} = T_1 + D_{MM}/c, \tag{14}$$

where c is the constant sound speed. Combining (11) and (12), we then have

$$T_1 - T_0 = T_{B2} - T_{B1} + D_{AB}/c - D_{MB}/c. \tag{15}$$

Similarly, the following also holds:

$$T_1 - T_0 = T_{M2} - T_{M1} + D_{AM}/c - D_{MM}/c. \tag{16}$$

Then, we can obtain

$$D_{AM} + D_{MB} = D_{AB} + D_{MM} + c \cdot (T_{B2} - T_{B1}) - c \cdot (T_{M2} - T_{M1}). \tag{17}$$

In general, given any anchor X, we can obtain the following equation

$$D_{AM} + D_{MX} = D_{AX} + D_{MM} + c \cdot (T_{X2} - T_{X1}) - c \cdot (T_{M2} - T_{M1}). \tag{18}$$

Since D_{AX}, D_{MM}, D_{MA} are all known a priori, it is easy to compute the distances from M to all anchors. We can then compute the position for M using the technique described next. Note that the BeepBeep localization system only requires that the target node be equipped with a microphone and a speaker and all anchors need a microphone per device. Moreover, one anchor should have a speaker, and other anchors may not have a speaker per device.

4.2. Trilateration-Based Localization

Once multiple D_{MX} are obtained through this scalable ranging procedure, we can use trilateration-based localization to localize node M. To this end, the common method is to calculate the optimal position via *least squared minimization (LSM)*, which minimizes the sum of the squared errors of each distance. Then, the target location (x, y, z) should minimize

$$S_0 = \sum_{i=1}^{n} \left(\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} - D_i \right)^2, \tag{19}$$

where (x_i, y_i, z_i) is the ith anchor's position, and D_i is its distance to the target sensor.

This optimization problem can be solved by a recursive algorithm, for example, the Newton-Raphson algorithm or the gradient descent technique. However, such algorithms have two downsides. The initial guess is vital and the algorithm performance is rather sensitive to the initial guess in many cases. Moreover, the iterative optimization is computationally intensive, and does not fit for devices such as sensors.

Therefore, we propose a novel, fast algorithm based on linear matrix operations. To this end, we slightly modify the optimization criterion (19) as follows:

$$S_1 = \sum_{i=1}^{n} \left((x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 - D_i^2 \right)^2. \tag{20}$$

The optimal position is the least-squared solution to the following matrix:

$$S_1 = \| \mathcal{B}\mathcal{X} - \nabla \|_2, \tag{21}$$

where we have

$$\mathcal{X} = (x^2 + y^2 + z^2, x, y, z)^T$$

$$\mathcal{B} = \begin{pmatrix} 1 & -2x_1 & -2y_1 & -2z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & -2x_n & -2y_n & -2z_n \end{pmatrix}$$

$$\nabla = \begin{pmatrix} D_1^2 - x_1^2 - y_1^2 - z_1^2 \\ \vdots \\ D_n^2 - x_n^2 - y_n^2 - z_n^2 \end{pmatrix}$$

Therefore, we can obtain the node's position via the following matrix processing:

$$\mathcal{X} = (\mathcal{B}^T\mathcal{B})^{-1}\mathcal{B}^T\nabla \tag{22}$$

Note that $\mathcal{B}$ remains constant during the localization process as it relates only to the anchors' positions. We can calculate $(\mathcal{B}^T\mathcal{B})^{-1}\mathcal{B}$ once the anchors are deployed and the system is calibrated. At hence, our method uses the above linear matrix processing to obtain the location for node M. Note that the first element in $\mathcal{X}$ is dependent on, thus redundant to the other three elements x, y, z. We ignore this item when computing the node position. It is also easy to see that Eq. (22) needs to have full rank. This necessary condition implies that we need at least four anchors to localize a 3D node using our fast computation method.

5. DETAILED DESIGN AND IMPLEMENTATION

In the section, we elaborate on the design and implementation details of BeepBeep ranging and localization. We show how various components in BeepBeep work in concert to ensure high-accuracy ranging and localization.

5.1. Designing the Acoustic Signal

To achieve high precision in ranging, it is critical to accurately locating time-of-arrival, that is, the first signal sample in the recorded sound. To suppress the effect of ambient noise and interference, the overall TOA scheme takes a correlation-based detection approach. That is, each device records the received signal, which is correlated with the reference signal in the time domain. Two design factors play crucial roles in the accuracy of TOA detection: one is the design of the acoustic signal and the other is the correlation processing. In this section, we first discuss the signal design, and then describe our signal processing scheme.

The main goal for signal design is to specify a reference signal with good cross-correlation property, and ensure that it only incurs small signal distortion even using commodity hardware. However, this is particularly challenging for COTS mobile devices. In general, speakers and microphones in such devices are designed with the mindset that the primary application is voice conversation. Therefore, it is natural that they possess good frequency response only around the narrow spectrum for human voice. Figure 7 shows the frequency responses of the HP iPAQ rw6828 and the Dopod 838 smartphones when we playback and record a chirp signal from 1kHz to 20kHz. The sound signal has been greatly attenuated when the frequency is higher than 8kHz, which is the upper bound of human voice. Therefore, we select the frequency range of the sound signal to be within 2–6kHz.

Another vital issue is the sound waveform design. The signal should have good autocorrelation property and allow for accurate signal detection even in the presence of ambient noise. A typical signal pattern that meets our requirement is the linear chip signal. However, we do not claim this choice to be optimal and more discussions on waveform design can be found in Girod and Estrin [2001].

The third issue when designing the signal is to choose a proper length for the sound signal. In particular, when working in an indoor environment, sound signal could reach the destination over multiple paths with different delays. This *multipath effect* may incur ambiguous ETOA detection, thus significantly reducing the detection accuracy. The signal length provides tradeoffs between the multipath effect and SNR. On one hand, to achieve high SNR, we prefer to large signal length. On the other hand, long sound signals may suffer more from the multipath effect because the signals from secondary paths overlap more with the primary-path signal. In our implementation, we choose the fixed signal length to be 50 milliseconds to balance between multipath effect suppression and noise resistance. A more elegant, yet also more complex,

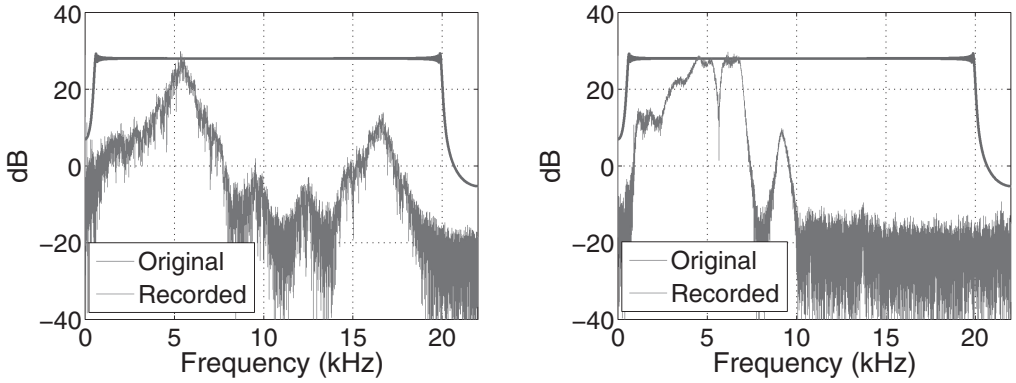


Fig. 7. Frequency response of the HP iPAQ rw6828 (left) and Dopod 838 smartphones (right).

solution is an adaptive signal length. For example, we can use shorter signals when the environment remains relatively quiet but subjects to multipath interference (indoor environments); while we use longer sound signal in the outdoor environment where noise is the dominating factor that affects the ranging precision. For simplicity, our prototype system does not take the adaptive approach.

Another concrete problem we have uncovered is that the sound waveform being played out has very large distortion during the initial few milliseconds. We believe that this is caused by the speaker diaphragm inertia. To address this issue, we precede the chirp signal with a 5-millisecond, 2kHz cosine waveform to warm up the speaker.

In summary, we have designed a linear chirp signal for our ranging system, but its spectrum is limited due to the constraints coming from the speaker and the microphone in COTS devices (2K–6K Hz). The signal length is fixed as 50 milliseconds.

5.2. Detecting the Signal

The signal is detected by correlating the reference chirp signals in the time domain. In our implementation, since the same chirp signal is used by all ranging parties, it is required to associate each signal to an individual device in order to calculate ETOAs. To differentiate these signals, we employ a schedule-based protocol that allocates a specific time window for each party in a ranging process to emit the sound signal.[2] As aforementioned, all devices are not tightly synchronized. Therefore, the scheduled time window should be large enough to reliably separate sound signals from different parties. We denote N as the number of samples for the chirp signal. For example, when the signal length is 50ms and the sound sampling rate is 44.1kHz, N equals to 2205 sample points.

To detect the signal, the recorded data are correlated with the reference signal and the maximum "peak" is located. This maximum peak is inferred as the location of a signal if its cross-correlation value is significantly larger than that with background noise. We calculate the L_2-norm of the cross-correlation values within a small window of w_0 samples around the peak, $L_2(S)$. Then, we calculate L_2-norm of correlation values in a w_0 window that is at least N samples before the peak, $L_2(N)$, where it is considered to contain only noise. A signal is detected only when $L_2(S)/L_2(N) > T H_{SD}$. If no such quantified point is found, we conclude that the detection failed, which could be because

[2]Pseudonoise signal can be used to exempt the schedule-based protocol, at the cost of significantly increased signal detection complexity.

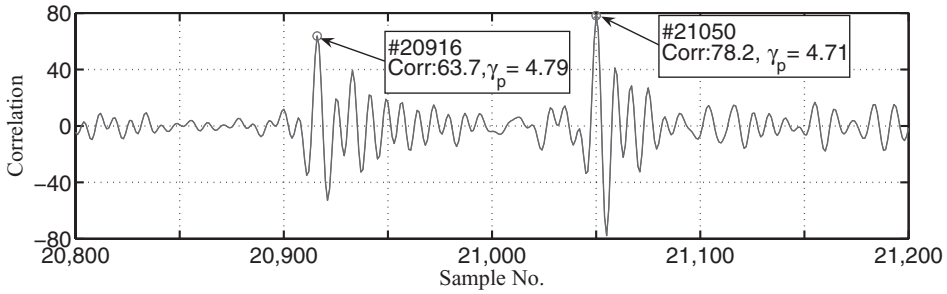


Fig. 8. One example case of multipath effect.

the signal energy is too weak or the noise level is high. In our implementation, we set $TH_{SD} = 2$ (i.e., 3dB) and $w_0 = 100$.

We also design techniques to address potential multipath effect. In an indoor environment, reflection from a secondary path may overlap with the signal from the line-of-sight (LOS) path. Such signal combination may cause the maximum peak to appear at the secondary path, which is slightly lagged regarding to the signal that travels in the primary path. One example is shown in Figure 8. It is clear that the peak corresponding to the primary path occurs at sample 20916, while the maximum peak occurs at sample 21050, which is about 0.3ms later. In our design, we handle the multipath effects by locating the earliest "sharp" peak in the shadow window. Intuitively, sharpness characterizes the the level of peak regarding to its surrounding side-lobes. Since cross-correlation values of signal from different paths should have similar sharpness, we determine the first peak that has comparable sharpness as the maximum peak as the TOA of the signal. The detailed procedures is as follows: we calculate the sharpness of a peak as the ratio of the peak value to the average absolute cross-correlation values in its w_1 vicinity. Then, we compute all peaks in the shadow window before the maximum peak and find the first one whose sharpness γ_p is larger than $\gamma_{max} \times TH_{MP}$, where TH_{MP} is a threshold. In our implementation, we empirically set $TH_{MP} = 85\%$ and $w_1 = 100$.

5.3. Handling Measurement Inaccuracy for Localization

We design further techniques to handle measurement inaccuracy in localization. In our BeepBeep system, measurement errors come from the noisy samples taken at each anchor, and can be reflected by the standard deviation of the samples. Our proposed combinational ranging scheme inherits relatively high accuracy from the BeepBeep ranging scheme. However, as shown in Section 6, the standard deviation of the ranging results can still be as large as four centimeters and the confidence level in the indoor environment is only about 94%. Therefore, the localization scheme in BeepBeep still needs to address this issue carefully.

Our solution technique is to incorporate a weighting factor, which reflects the confidence level of the ranging results. When low-quality ranging samples are assigned small weights, their impact on the localization results will also be reduced. To this end, we examine the following weighted form of the objective function of (20):

$$\mathcal{J}_2 = \sum_{i=1}^{n} \omega_i \big((x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 - D_i^2\big)^2, \tag{23}$$

where ω_i denotes the weight of the ith anchor.

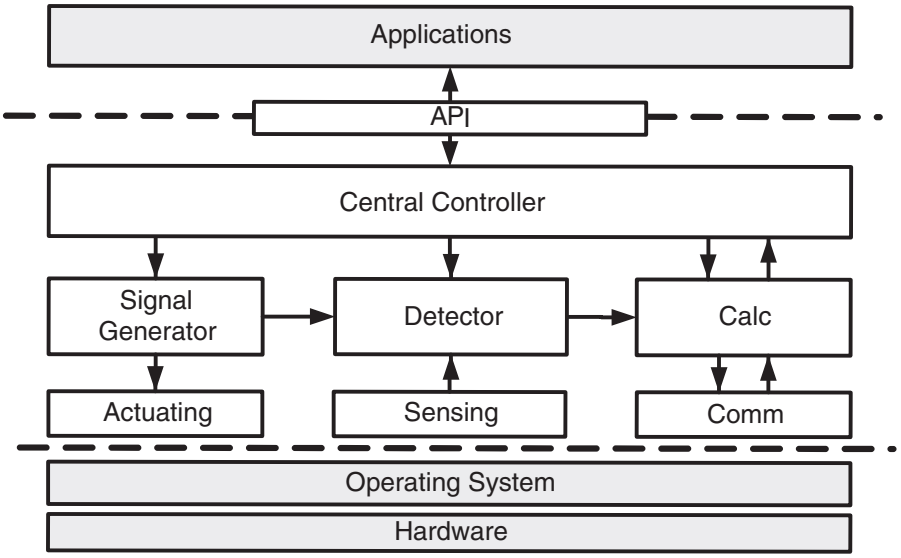


Fig. 9. Software architecture of BeepBeep system.

Then, our solution becomes

$$\mathcal{X} = (\mathcal{B}^T \mathcal{W}^T \mathcal{W} \mathcal{B})^{-1} \mathcal{B}^T \mathcal{W}^T \mathcal{W} \nabla, \tag{24}$$

where $\mathcal{W}$ is the diagonal weight matrix consists of ω_i.

In BeepBeep, we also come up with a nice way to assign the weight factor ω_i. Ideally, ω_i should reflect the confidence level of the ranging result against the ith anchor. In chirp signal detection, the recorded data are correlated with the reference signal, and the maximum "peak" is located. This maximum peak is inferred as the location of a signal if its cross-correlation value is significantly larger than that with background noise. We use their ratio to set the weight ω_i. In the extreme case, we may simply set $\omega_i = 0$ and the result from the noisy, ith anchor is consequently discarded.

5.4. System Architecture

The BeepBeep system is implemented purely at the application-layer. It is ready for deployment on most COTS mobile devices.

Figure 9 shows the overall software architecture. It has fours major parts: the underlying physical device related function modules, the core logic part, the interface to other applications (API), and applications.

The physical-device-related function modules include the actuating module that sends out the sound signal generated by the signal generator; the sensing module which continuously records all the sound into a local buffer and feed the buffered data to the signal detector; and the communication module that allows for lightweight information exchange such as the ETOA data and scheme-specific parameters.

The core logic part consists of the central controller, the signal generator, the signal detector and the distance calculation module. The central controller implements the overall BeepBeep ranging protocol and controls all other modules' actions, as well as signal design and calibration module. It also interacts with other applications by receiving requests and sending back responses through API. A state machine is maintained in the central controller for ranging scheduling. The signal generator generates the waveform of ranging signals based on given parameters and feeds the signal to the

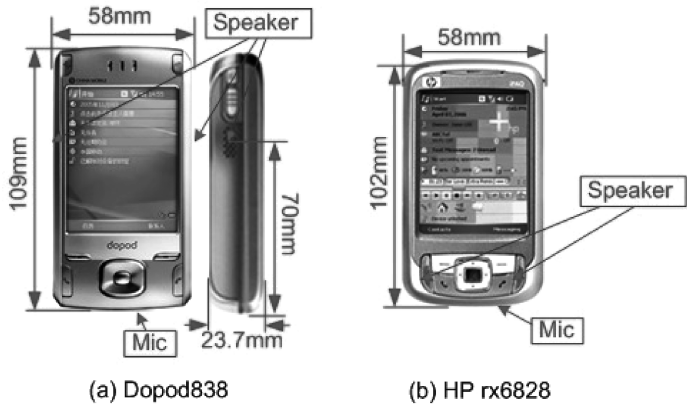


Fig. 10. Two COTS mobile devices used in the evaluation of BeepBeep ranging system.

actuating module. The generated signals are also stored as reference signals for signal detection. The signal detector implements the signal detection algorithms and determines the indices of the first samples (i.e., TOAs) of other parties' signals as well as its own. Ranging signals are detected by matching the recorded data from the sensing module against their respective reference signal templates. The distance calculation module simply calculates the distance to other parties after receiving all respective ETOAs according to Eq. (9).

The API module defines the interface to the BeepBeep ranging and localization. Various applications can be constructed via this API and the basic ranging scheme.

We have implemented the BeepBeep system in Windows Mobile 5.0. It works as a user-mode dynamic linkable library that other applications can load and use. We use multimedia services (*WaveXXX* series APIs) embedded in Windows Mobile to control microphones and speakers and rely on WinSock for data communications over WiFi.

6. BEEPBEEP SYSTEM EVALUATION

In this section, we assess the performance of the BeepBeep system. We use various experiments to gauge the effectiveness of the proposed ranging and localization schemes.

In our experiments, we use two models of commercial off-the-shelf PocketPC phones, HP iPAQ rw6828 and Dopod 838, as shown in Figure 10. Both devices run Microsoft Windows Mobile Version 5.0 (Phone Edition), with WiFi and Bluetooth radios, 64 MB RAM, two built-in speakers and a microphone that supports 16-bit 44.1 kHz sampling rate. The HP iPAQ rw6828 features a more powerful Intel XScale 416 MHz processor while Dopod 838 is equipped a 195 MHz TI OMAP850 processor. The speakers are laid out at the bottom on the front face for the HP phone and at the two sides on the Dopod phone.

Due to space limit, we only report the results for Dopod 838 cell phone. The experimental results on the HP iPAQ rw6828 phone are similar. Unless explicitly stated, all experiments are performed using the left speaker and the microphone on the device. Recall that, we need certain calibration to remove the impact of K, the constant length between the speaker and the microphone for each device. The specific calibration values for the HP iPAQ rw6828 phone and the Dopod 838 phone are 3cm and 8cm, respectively. No calibration is needed for the experiments using earphone because we let $K = 0$ by placing the earphone very close to the microphone.

6.1. Ranging Experiments

We use a series of experiments to evaluate the accuracy and operation range of the BeepBeep ranging scheme.

6.1.1. Experiment Settings. In our experiments, the sound speed used in the distance calculation is set according to the following model [Wikipedia]: $c_{air} = 331.3 + 0.6 \cdot \theta$ where 331.3 m/s is the benchmark speed at 0°C and θ represents the air temperature in Celsius (°C).

We use the following three metrics in our evaluation: (1) *Accuracy*: Accuracy is defined as the difference between the ranging results and the real distance. It may be expressed by the maximum, minimum, median and mean ranging error and the standard deviation of the ranging error; (2) *Confidence*: confidence is defined as the percentage of the times a known level of accuracy is reached. We denote α-confident as the proportion of ranging experiments that achieved a ranging error no larger than the threshold α, where α can be an absolute error (e.g., 5cm) or a relative one (e.g., 1%); (3) *Operational range*: operational range is defined as the maximum range that the ranging system can still achieve a known-level of accuracy β with a certain confidence α. In our evaluation, β is set to 5cm and α to 90%.

Our experiments are performed in the following four scenarios:

—*Case-A – Indoor, quiet*: A meeting room that is approximately 5m×11m with a big table in the center where different test locations are also illustrated. The environmental temperature is 25°C.
—*Case-B – Indoor, noisy*: The same room as Case-A but with background noise consisting of the air conditioner noise, background pop music (at various volume levels and various relative positions between the sound box and the smartphones), and people chatting around the table. The temperature is 25°C.
—*Case-C – Outdoor, car park*: The open area in front of a medium size car park at the side of a large building and nearby the main driving course. The weather is windy and wind speed is about 10 miles per hour. The temperature is 10°C.
—*Case-D – Outdoor, subway station*: The open area at the entrance of a subway station, just after the rush hour and with medium traffic. The temperature is 15°C.

In all the experiments, the devices are placed parallel to each other and face up, as depicted in Figure 3. In each setting, the experiments were repeated 50 times. In the rest of this section, we use the following convention when presenting the accuracy results: In each figure, the left and the right vertical axes are the reference scales for the ranging error (max/min, median and mean) and the standard deviation, respectively. The dash line denotes the mean ranging error, and the solid line shows the standard deviation.

6.1.2. Indoor Cases. Figure 11(a) plots the accuracy measurements at different distances for the Case-A setting. From the figure, we can see that our system yields highly accurate and stable ranging results in a quiet, indoor environment. Both the median and mean ranging errors are within ±1cm and the standard deviations are within 2cm. Figure 11(b) shows the corresponding α-confidence plots where α is set to 1cm, 2cm, 3cm and 5cm, respectively. High percentage of experiments lead to less than 1cm accuracy. The performance is very robust within the range of 4 meters. However, the performance starts to deteriorate when the distance is even larger. For instance, about 95% experiments have absolute ranging errors within 5cm when the distance is 4 meters, but the percentage drops to zero at the distance of 5 meters.

The accuracy and confidence measurements at different distances for the Case-B setting are shown in Figure 12. From the figure, we see that the BeepBeep ranging

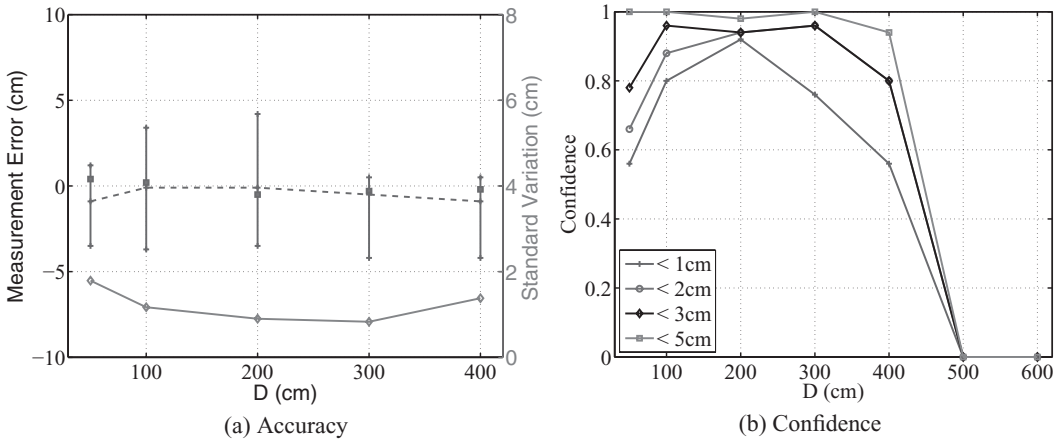


Fig. 11. Accuracy and confidence measurement results at different distances in the Case-A setting.

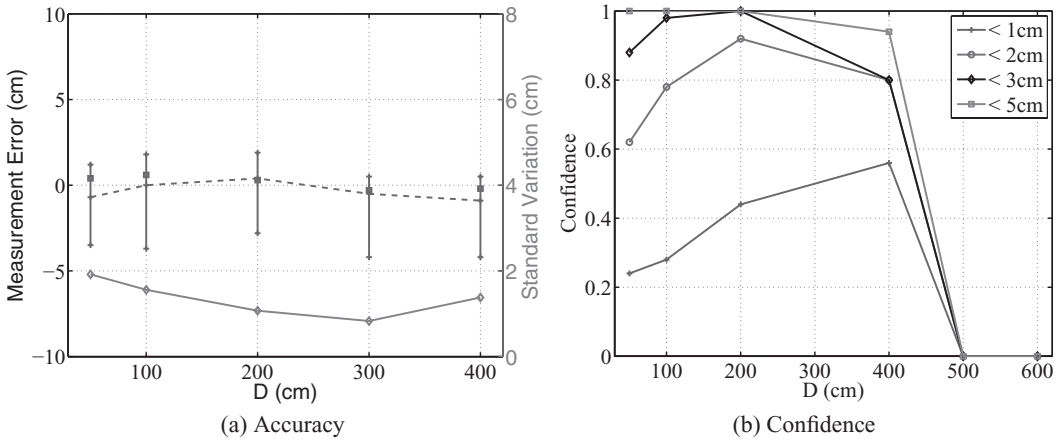
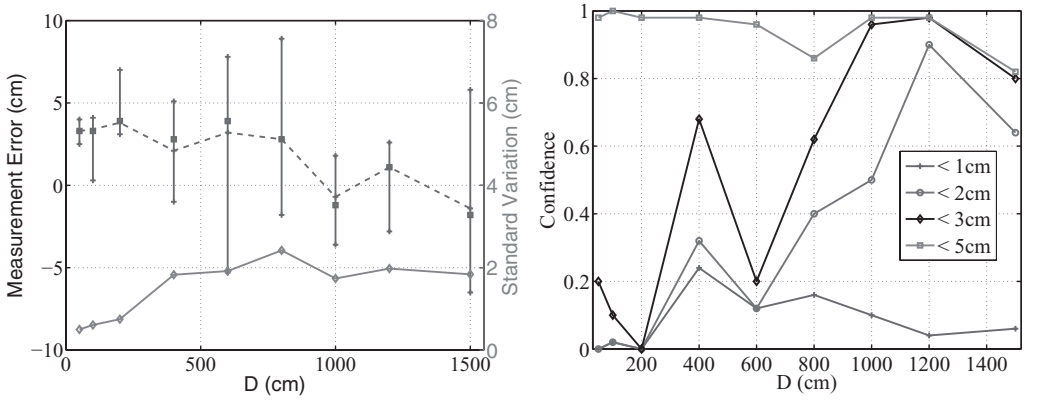


Fig. 12. Accuracy and confidence measurement results at different distances in the Case-B setting.

system still performs fairly well. The overall ranging accuracy is comparable to that of Case-A. These results demonstrate the excellent noise-resistance property of the chirp signal. The comparison between Figure 11 and 12 further reveals that the background noise only leads to some decreases in the 1cm-confidence case, but has no visible impact on the 2cm-confidence and these cases.

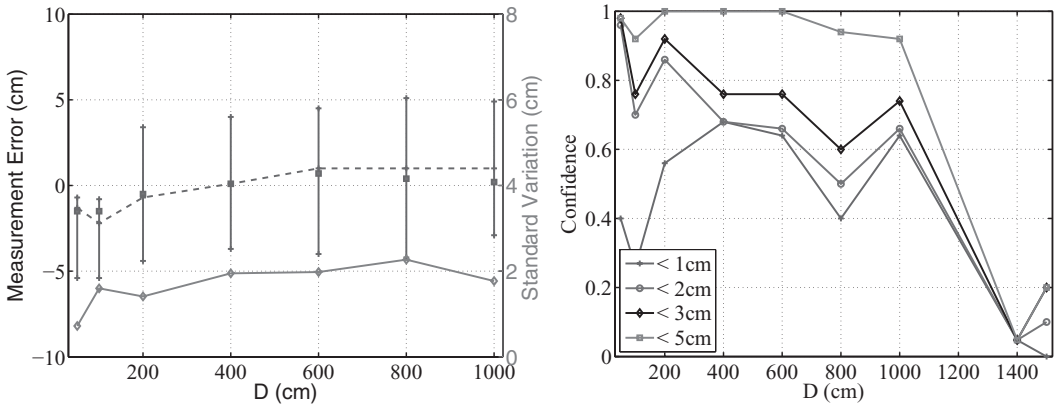
These experiments show that, the operational ranges are both 4 meters for the given settings of Case-A and Case-B. Such results are mainly due to the multipath effects induced by the small size of the meeting room [Peng et al. 2007]. Nonetheless, as long as the devices are within the operational range, our system works reliably and the ranging results are highly confident.

6.1.3. Outdoor Cases. The accuracy and confidence measurements for Case-C and Case-D environments are shown in Figures 13 and 14, respectively. The BeepBeep ranging system still works well. The median and mean ranging errors for Case-C are within [-2cm, 4cm], and the errors for Case-D are within [-1.5cm, 1cm]. Compared with the indoor cases, the ranging error increases and exhibits a larger dynamic range.



(a) Accuracy (b) Confidence

Fig. 13. Accuracy and confidence measurement results at different distances in the Case-C setting.



(a) Accuracy (b) Confidence

Fig. 14. Accuracy and confidence measurement results at different distances for in the Case-D setting.

However, the standard deviation is still very small (less than 2cm in most cases); this again demonstrates the robustness of our system.

The figures also show that, the confidence plots of the two cases differ significantly. While Case-D seems to be normal as we have expected, the α-confidence plots (when α is set as 1cm, 2cm, and 3cm) for Case-C seem quite abnormal. We suspect that this is due to a time-of-the-day effect and impact by the car traffic. We started the experiment at 6:30pm (i.e., off hours) and the venue is at the entrance of the parking lot and close to the main drive way. Another observation is that in Case-D, when the distance exceeds 10 meters, the performance starts to drop, and most experiments failed when the distance is larger than 14 meters, as shown in Figure 14(b). The reason is still the multipath effect since one device is close to the wall of a high building when the distance is large. In contrast, in Case-C, the performance is still good when the distance is larger than 14 meters. In fact, a significant portion of experiments succeeded when the distance is 20 meters, but the time is already deep into night. Therefore, we exclude those results. It is not hard to understand that the sound volume is an dominating factor to the operational range. In our experiments, the volume from COTS phone are too small to be heard when we are more than 10 meters from the phone in a noisy

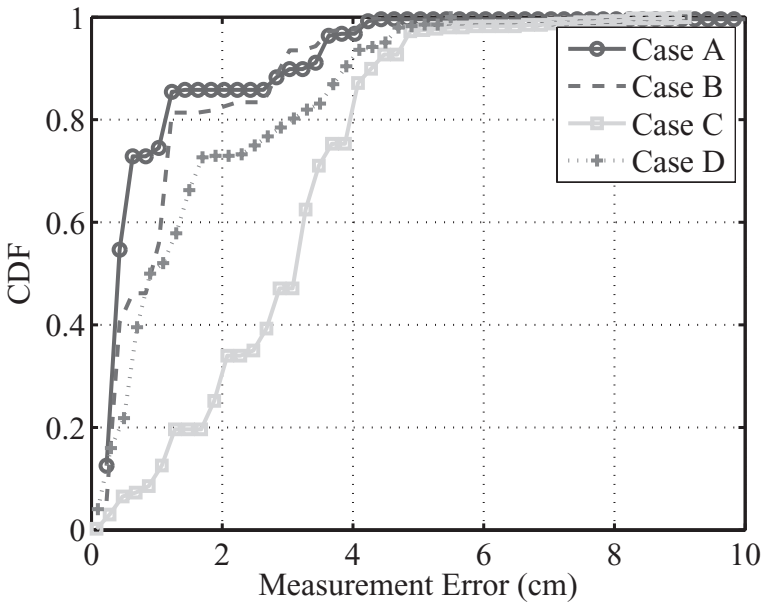


Fig. 15. Cumulative distribution function of the α-confidence for the four test cases.

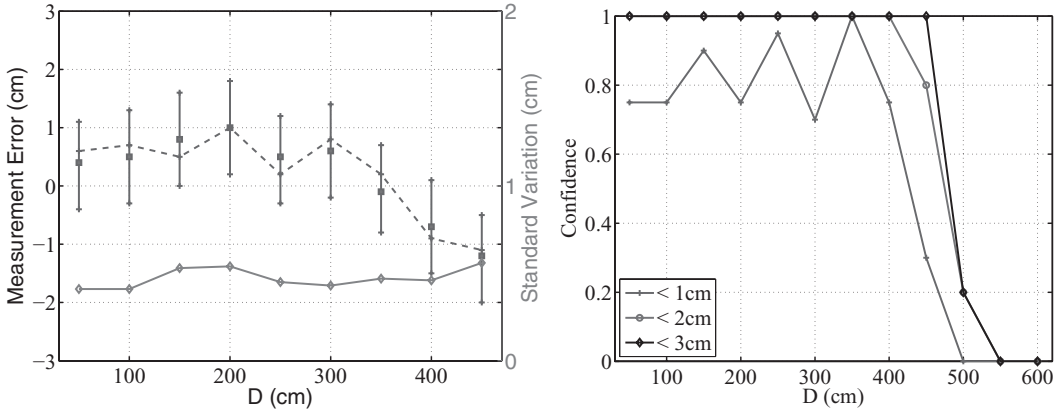
outdoor environment. It finally leads to a low Signal-Noise-Ratio(SNR) and restricts the operational range. The phone with a larger volume or a more powerful speaker can extend the operational range easily, though our experiments has validated that the operational range on COTS phones is quite large.

In this measurement results, we have shown the accuracy and confidence metric at different distances under the four test cases. In practice, since such distance information is not known a priori, we therefore plot the cumulative distribution function of the ranging accuracy for all the experiments so as to provide a holistic view. From Figure 15, we can see that for all the test cases, our ranging results are highly reliable within their operational ranges. The probability that our system leads to less than 4cm ranging accuracy is higher than 95% for Case-A, Case-B and Case-D and is about 86% for Case-C.

6.1.4. Impact of Signal Distortion. Note that one source of TOA detection error is the signal distortion. As an earphone generally has better frequency responses than the speaker and better preserves the signal waveform, we conducted another set of experiments in the Case-A environment using the earphones. The accuracy and confidence measurements are shown in Figure 16.

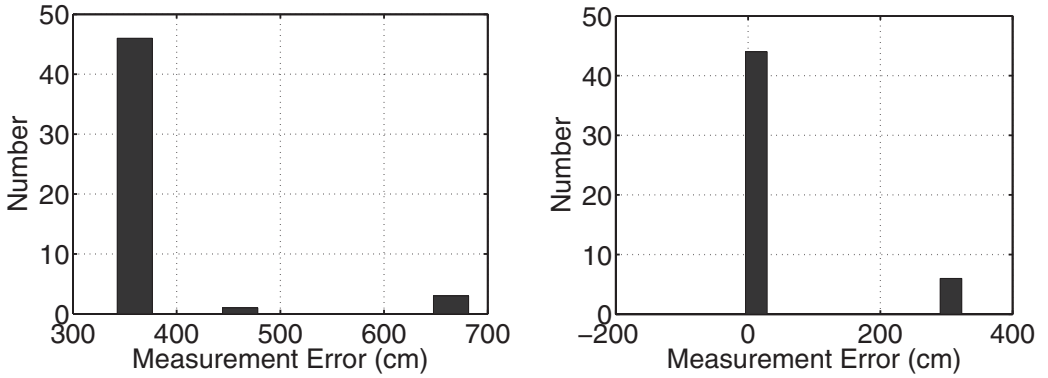
From the figure, we see that using earphone leads to excellent accuracy with less than 1cm-mean and median ranging errors and the maximum ranging error is within $\pm$2cm. The ranging results are also quite stable. The average and maximum standard variations among all the test distances are 0.47cm and 0.6cm, respectively. The α-confidence plots state that 100% 2cm-confidence is achieved and with over 70% chances we have the ranging error smaller than 1cm.

Comparing Figures 16 and 11, we see that the earphone outperforms the speaker in all aspects. These experiments confirm that signal distortion indeed impairs the ranging accuracy. Therefore, a better speaker will improve the ranging performance. Contrary to our expectation, although the earphone's signal power is much weaker than that of the speaker, the earphone actually has a larger operational range. This



(a) Accuracy (b) Confidence

Fig. 16. Accuracy and confidence measurement at different distances in the Case-A setting, using earphone.



(a) No mitigation (b) With mitigation

Fig. 17. Multipath effects and mitigation in Case-A.

observation suggests that the impact of signal distortion should be even larger than that of the signal to noise ratio.

6.1.5. Multipath Effect Mitigation. The multipath effect typically becomes evident when the signals from non-LOS paths have strengths comparable to that of the LOS signal. The effect will be aggravated when the lengths of non-LOS paths are close to that of the LOS path, thus causing interferences. This is the primary reason why our experiments failed when the distance is large for both indoor cases. In those experiments, due to the space constraint, we placed one device close to walls and heavy multipath effects showed up consequently. We did not observe such effects in the Case-C setting.

Figure 17(a) plots the distribution of measurement errors at a 5-meter distance in the Case-A setting if we simply determine the TOA sample according to the maximum cross-correlation value, that is, without our mitigation effort in multipath effect. Clearly, all the 50 experiments failed and the errors are larger than 3 meters. Figure 17(b) shows the measurement error distribution when our heuristic multipath effect mitigation algorithm is applied. All the cases with more than 4 meters errors and most cases with 3 meters errors are correctly handled. The percentage of successful experiments becomes 88%. However, there are still a few error cases remaining, which

Table I. Summary of all experimental results

Environment Setting	Operational Range	Confidence Level	$\text{Avg}_d\left(\text{Avg}_n(\lvert Err\rvert)\right)$ (cm)	$\text{Max}_d\left(\text{Avg}_n(\lvert Err\rvert)\right)$ (cm)	$\text{Avg}_d(Std)$ (cm)	$\text{Max}_d(Std)$ (cm)
Earphone	4.5m	100%	0.6	1.4	0.4	0.6
Case-A	4.0m	94%	0.9	1.4	1.2	1.9
Case-B	4.0m	94%	1.1	1.7	1.0	1.3
Case-C	12m	98%	2.7	3.8	1.0	2.1
Case-D	10m	92%	1.0	2.2	1.4	1.6

Note: Subscript d and n denote that the operations (i.e., avg, max) are conducted over different measured distances in each test case and over n ($n = 50$) runs of experiments, respectively.

suggests more future research work. Note that, in the previous evaluation results, the multipath mitigation algorithm has not been applied. If it were applied, the operational range would be significantly improved, at least relaxing requirements for accuracy or confidence.

6.1.6. Experiments Summary. Table I summarizes the statistics for all the experiments. In the table, the overall average and maximum ranging errors are computed over all the measured distances for different test cases. Here, the *absolute* values of ranging errors are used for the mean and standard deviation calculation. This table shows that the BeepBeep ranging system indeed leads to high-accuracy (about 1cm) ranging results and works reliably in all test cases. The operational range for the indoor cases is around 4 meters and that for outdoor cases is generally larger than 10 meters. As long as the devices are within the operational range, our system works reliably and the ranging results yield high confidence factors.

6.2. Localization Experiments

We also perform experiments to assess the effectiveness of our localization scheme, as well as the impact of anchor placement. All the experiments have been conducted in an office room with furniture such as sofa, desks, cabinets, etc, and eight anchors are deployed around the center of the room, spanning about $1.8\text{m}\times1.6\text{m}\times0.85\text{m}$ space $\mathcal{C}$.

We randomly select six positions in $\mathcal{C}$ that are roughly uniformly distributed, and perform localization in a quiet environment.[3] Two techniques are assessed: one is the original linear matrix solution (Eq. (20)), and the other is the weighted one (Eq. (23)). Figure 18 plots the statistical results over more than 80 independent trials. From the figure, we see that BeepBeep-based localization indeed delivers very high accuracy. The mean square localization error (the upper figure) is typically less than 3, with the maximum difference being 7. Based on the mean error and the standard deviation, we infer that such maximum-difference cases happen rarely. The localization performance is also very stable: the standard deviation (the bottom figure) is less than 1cm for most positions. Our weighted linear matrix solution clearly outperforms its original counterpart. In the rest of experiments, we always use the weighted linear matrix solution.

We next gauge the impact of anchor placements. By further examining Figure 18, we find that Positions 4 to 6 suffer from larger errors than other positions. This can be explained by their distances to the anchor, which emits the first beep signal. We observe that Nodes 4 to 6 are closer to that anchor than others.

We verify this hypothesis using simulations in a 2D setting, where 4 anchors are placed on the vertexes of a 100cm by 100cm square. The beeping anchor is at (0,0). The

[3]The noisy case has also been tested but the results are not shown here due to space limit. The localization performance in the noisy case is similar but shows larger variances, and matches with aforementioned ranging results.

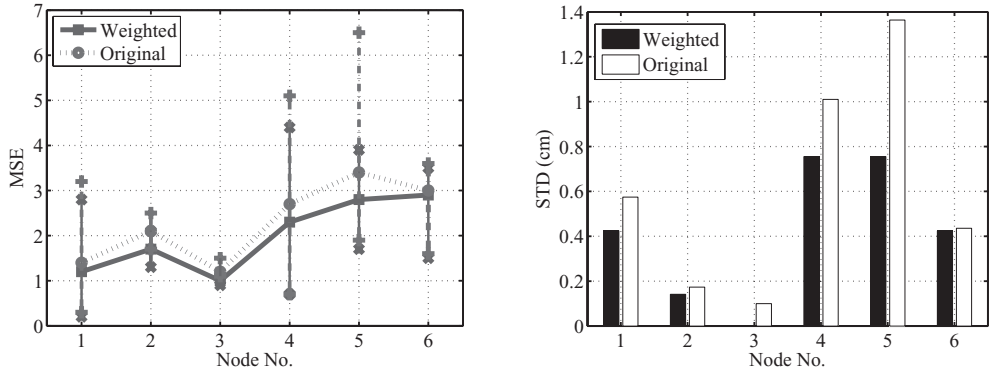


Fig. 18. Localization of points at different positions, using three methods. Left: mean square error. Right: standard deviation.

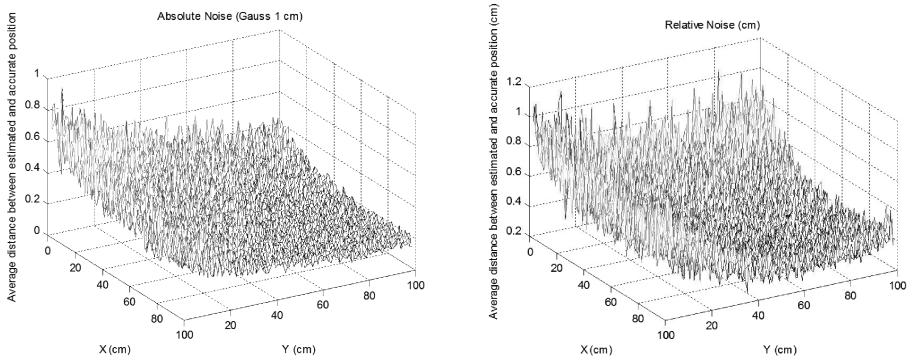


Fig. 19. Impact of beeping anchor placement on localization. Left: Absolute ranging noises; Right: Relative ranging noises.

node to be localized is placed at position [1:99, 1:99]. Assume that the measurement noises are Gaussian. Two sets of simulations are conducted. In the first set, the noise is independent of the range and its standard deviation is set as 1 cm. In the second set, the noise is (1%) in proportion to its distance from the beeping anchor. The results are shown in Figure 19. As expected, we find that, as the node moves closer to the boundary, the accuracy decreases. It is more sensitive to its distance from the beeping anchor. This observation suggests that, to perform more accurate localization, we should choose the faraway anchor as the beeping node.

7. RELATED WORK

Ranging and localization have been an active research area in recent years, spanning both theoretical studies [He et al. 2003; Aspnes et al. 2006; Whitehouse et al. 2005] and practical systems [Harter et al. 1999; Priyantha et al. 2000; Niculescu and Badrinath 2003; Bahl and Padmanabhan 2000; Borriello et al. 2005; Stoleru et al. 2006]. The various system prototypes use diverse techniques of angulation or lateration [Harter et al. 1999; Priyantha et al. 2000; Niculescu and Badrinath 2003], proximity sensing [Bahl and Padmanabhan 2000; Borriello et al. 2005], or even scene analysis [Stoleru et al. 2006]. We now review the proposals that are most relevant to BeepBeep.

Most existing high-accuracy ranging schemes and lateration-based localization systems rely on time-of-arrival (TOA) of acoustic signals. We focus on such solutions and

compare these approaches with BeepBeep in Table II. The table shows that, many TOA-based systems require special hardware or infrastructure support, or even OS modification to support time synchronization. They include Bat [Harter et al. 1999], Cricket [Priyantha et al. 2000], AHLoS [Savvides et al. 2001], ENSBox [Girod et al. 2006], and microphone-array-based solutions [Aarabi 2003]. In contrast, BeepBeep works with COTS and does not need special hardware support. In addition, it also uses ranging techniques different from the existing solutions.

The Bat system [Harter et al. 1999; BatSystem 2003] is a high-accuracy indoor localization system, based on time-of-flight of ultrasonic signal. It achieves accuracy up to 3 centimeters in most tested cases. However, it relies on an infrastructure consisting of an irregular matrix of networked, ultrasonic receivers. Similar to the microphone-array-based solutions [Aarabi 2003], all anchors are synchronized in clock. BeepBeep differs from Bat in two aspects: Its anchors do not need explicit clock synchronization and complex infrastructure support. BeepBeep also uses audible acoustic signals and works well with standard mobile phones.

Cricket [Priyantha et al. 2000] and AHLoS [Savvides et al. 2001] perform ranging through the difference in the propagation speeds of radio and ultrasonic signals. They use the radio signal to synchronize anchors. Cricket depends on pre-installed anchors while AHLoS iteratively turns recently localized nodes into new anchors. AHLoS achieves ranging accuracy up to a few centimeters through advanced multi-lateration techniques, much better than Cricket (which accuracy is about 20cm). BeepBeep differs from Cricket and AHLos in two aspects: Its anchors do not need radio signal for clock synchronization and therefore it is not necessary to emit radio and acoustic signals simultaneously, which requires real time control in operating systems. BeepBeep works well with commodity hardware and provides higher accuracy.

Aarabi's work [Aarabi 2003] and ENSBox [Girod et al. 2006] further incorporates information on the microphone array into position estimation. Compared with the TOA-based solution, microphone-array based methods are using differential time-of-arrival (DTOA). In DTOA, the source emits a sound signal, all the receivers (with known positions) record the sound. Through combining DTOAs, the position of the sound source can be determined. Aarabi's work utilizes spatial likelihood functions (SLFs) produced by each microphone array to enhance sound localization to the accuracy of 8cm 90cm. ENSBox [Girod et al. 2006] designs a box equipped with three microphones at specific positions achieves an accuracy up to a few centimeters in outdoor environment. Compared with them, BeepBeep does not need any of such special hardware. Moreover, BeepBeep performs combinational ranging, that is, the result is always the sum of two ranges while they use direct ranging based on beamforming technology.

There are other acoustic-based ranging systems that yields lower accuracy. For example, Frampton [2006] is another DTOA-based systems but it reversely uses the DTOA idea. In Frampton [2006], Clock-synchronized sound sources at different locations in turn send out sound signals at controlled times and sensors report their detected DTOA to a centralized server. Then, localization is done through nonlinear optimization. Their idea on the innovative use of DTOA is similar to our use of ETOA. The most distinctive feature that sets apart the BeepBeep ranging system is the complete avoidance of the clock synchronization among devices. In most DTOA based schemes described above, synchronization must be ensured among all senders or all receivers. Their performance is also different. [Frampton 2006] provides meter-level accuracy in outdoor scenarios while BeepBeep can achieve 1$\sim$3cm accuaracy.

There are also other systems that do not use acoustic signals but provide high accuracy ranging and localization, such as RIPS [Maroti et al. 2005], PinPoint [Youssef et al. 2006]. RIPS [Maroti et al. 2005] is based on radio interferometric and achieves high precision up to a few centimeters. PinPoint [Youssef et al. 2006] is another

Table II. Comparison of acoustic-based ranging and localization techniques

	Bat [Harter et al. 1999]	Cricket [Priyantha et al. 2000]	AHLos [Savvides et al. 2001]	Aarabi's work [Aarabi 2003]	EnsBox [Girod et al. 2006]	BeepBeep
Category	TOA	TOA	TOA	DTOA (Beamforming)	DTOA (Beamforming)	TOA
Medium	Ultrasound	Ultrasound+RF	Ultrasound + RF	Audible sounds	Ultrasound	Audible sounds
Hardware Requirement	Special HW (Ultrasound transceiver)	Special HW (Ultrasound+RF transceiver)	Special HW (Ultrasound+RF transceiver)	Special HW (mic array)	Special HW (3 mics at specific positions)	Commodity HW (COTS mic/speaker)
Synchronization	Yes	Yes (RF sync)	Yes	Yes	No	No
Ease of deployment	Difficult, requires matrix of sensors	Easy	Medium, (a large amounts of sensors)	Difficult, requires matrix of microphones	Easy	Easy
Applicable	Indoor	Indoor	Not specified	Not specified	Outdoor	Indoor/Outdoor
Independent Ranging	No	Yes	Yes	No	Yes (only ranging)	Yes
Ranging accuracy	—	10~20cm	A few centimeters	—	~5cm	~1cm
Localization accuracy	~3cm	10~20cm	A few centimeters	8~90cm	—	~3cm

TOA-based ranging system and achieves impressive accuracy (up to several feet) using radio signals. Their key idea is a mathematical technique for clock difference compensation which leads to very precise timestamp recovery while allowing the nodes' local clocks to run asynchronously. BeepBeep differs from PinPoint in that PinPoint uses the radio signal for ranging and requires special hardware with a high-frequency clock, while we do not need any reference clock at all. However, the two systems have in common the concept of two-way sensing. In fact, PinPoint performs two-way sensing twice in order to recover the timestamps. BeepBeep has several salient features different from the above systems. It does not need special hardware for ranging, no dependency on the radio signal, no requirement on clock synchronization among devices and no infrastructure needed for ranging. Moreover, BeepBeep uses novel solution techniques.

There are also a few ranging systems that do not need special hardware and can work on COTS devices. They include WALRUS [Borriello et al. 2005], Radar [Bahl and Padmanabhan 2000], Horus [Youssef and Agrawala 2005], ARIADNE [Ji et al. 2006], SurroundSense [Azizyan et al. 2009]. Such systems typically provide rough proximity information, rather than precise ranging information. Specifically, WALRUS reaches room-level precision regarding the mobile device's location via the ultrasound signal, because the ultrasound signal does not penetrate walls. Radar uses received radio signal strength from multiple APs and determines the location using the radio propagation model and the radio strength map obtained offline. Radar reaches precision up to a few meters and leverages the predeployed APs as its infrastructure. Using the same concept, Horus system identifies and combats several causes of wireless signal variation, proposes a location-clustering technique with less than one meter accuracy most of the time. ARIADNE further simplifies the construction of the signal strength map by using a floor plan and a single signal strength measurement, and proposes a clustering algorithm for localization. ARIADNE achieves about two-meter accuracy. SurroundSense uses ambient sound, light, and color in a place and convey a photo-acoustic signature for localization. All these ranging and localization systems are different from BeepBeep. Constraints such as their dependency on the infrastructure, indoor operation environment setting, and low precision, make them unsuitable for many of our target scenarios that are mobile and ad-hoc in nature and require high precision.

8. APPLICATIONS

This section describes how applications can exploit the functionality of ranging and localization offered by BeepBeep to enhance human-to-device communication. We have implemented two applications using the localization facility provided by BeepBeep or similar acoustic-ranging technique to bootstrap intuitive device-to-device communication.

8.1. Draw-in-the-air

Draw-in-the-air (DITA) is a gesture-based human-device interaction system based on the localization technique of BeepBeep. Its premise is that free-space human gestures are arguably the most natural and intuitive way for human-space interaction. With such a system, a customer uses simple and intuitive human gestures to interact with the smart home devices. For example, a tickle or cross gesture will turn the device on or off, and a left-leafing or right-leafing gesture will switch to the previous or next channel.

DITA provides a solution to capture the hand movement of a user with centimeter accuracy. It offers a simple, intuitive and robust solution to gesture-based user-device interaction. The key enabling technology for DITA is BeepBeep-based gesture capture scheme. DITA leverages a standard mobile phone and acoustic signals to track the

user's movement. When the user makes a hand gesture while holding the mobile phone, the phone will produce acoustic signals to be recorded by "anchors". DITA can subsequently establish the phone's trajectory from these acoustic traces and further extract users' intended gesture. Finally, the central server invokes the corresponding action at the intended device. Using DITA, a handheld mobile phone can turn into a gesture-based universal controller. More technical details can be found in Shen et al. [2009].

8.2. Point&Connect

Point&Connect (P&C) offers an intuitive and resilient device pairing solution on standard mobile phones in a multiparty setting – when there are many mobile devices within the communication range. It can be readily applied to a number of real life scenarios, such as file swapping, music sharing, and collaborative gaming on mobile phones in a close proximity. To do this, the mobile phones have to get connected first. With P&C, a user can follow the simple sequence of point-and-connect: when a user plans to pair her mobile phone with another device nearby, she makes a simple hand gesture that points her phone towards the intended target. The system will capture pointing action, understand the target selection intention, and complete the device pairing.

The key technique of P&C is to apply a novel collaborative scheme which enables a user to express, and the system to capture, an intention of device selection via a simple pointing action. The pointing action is detected by an acoustic-based maximum distance change. When a user moves her phone towards the target device, the selecting device emits two sequential "beep" sound signals and each "candidate" nearby will independently record them and compute the elapsed time of arrival (ETOA) of the two beeps heard. P&C subsequently exploits the fact that the candidate along the point direction should report the largest relative distance change. Using P&C, a user can bootstrap an effective device-to-device communication for any mobile phones or small devices without relying on infrastructure or special hardware. More technical details can be found in Point&Connect [Peng et al. 2009].

9. CONCLUSION

BeepBeep is a high-accuracy ranging and localization system designed for COTS mobile devices. The ranging accuracy is up to a centimeter. It is achieved through a pure, software-based solution and with only the most basic set of commodity hardware – a speaker, a microphone, and certain form of device-to-device communication. Moreover, the ranging solution operates in a spontaneous, ad-hoc, and device-to-device context without leveraging any pre-planned infrastructure. Once high-accuracy ranging is complete, localization service can be built upon the ranging module. BeepBeep offers scalable localization because its acoustic signaling overhead remains constant with respect to the anchor population during the localization process.

The solution approach of BeepBeep still follows the paradigm of measuring time-of-arrival (TOA) information of acoustic signals. But it devises several novel techniques to address various uncertainties associated with the TOA-based system. In the ranging solution, for example, the two-way sensing strategy avoids clock synchronization uncertainty, the self-recording strategy eliminates the sending uncertainty, and the sample counting method avoids the receiving uncertainty. In localization, rather than using pairwise ranging between the target sensor and each anchor, BeepBeep uses two signals and passive but simultaneous listening to keep the signaling overhead constant. Our extensive experiments have confirmed the accuracy and effectiveness of the BeepBeep solution. It achieves about 1cm and 2cm average ranging accuracy with less than 2cm standard deviations for typical indoor and noisy outdoor

environments, respectively. The localization error is typically less than 3cm, and the standard deviation is less than 1cm in most positions.

Because of the minimum hardware requirements, the solution techniques proposed in BeepBeep can be directly incorporated into the design of other customized sensor platforms and will lead to significant cost savings. In addition to ranging and localization, we also believe that BeepBeep will stimulate more applications in low-cost sensor networks. It may also spur a compelling set of social related mobile applications that request for proximity awareness and fine-grained control over the spatial relationship [Shen et al. 2007].

REFERENCES

AARABI, P. 2003. The fusion of distributed microphone arrays for sound localization. *EURASIP J. Appl. Sig. Process. 2003,* 1, 338–347.

ASPNES, J., EREN, T., GOLDENBERG, D. K., MORSE, A. S., WHITELEY, W., YANG, Y. R., ANDERSON, B. D. O., AND BELHUMEUR, P. N. 2006. A theory of network localization. *IEEE Trans. Mobile Comput. 5,* 12, 1663–1678.

AZIZYAN, M., CONSTANDACHE, I., AND CHOUDHURY, R. R. 2009. Surroundsense: Mobile phone localization via ambience fingerprinting. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MobiCom '09)*. 261–272.

BAHL, P. AND PADMANABHAN, V. N. 2000. Radar: An in-building rf-based user location and tracking system. In *Proceedings of the 19th Annual IEEE Conference on. Computer Communications (INFOCOM '00)*.

BatSystem 2003. The bat system. http://www.cl.cam.ac.uk/research/dtg/research/wiki/BatSystem.

BORRIELLO, G., LIU, A., OFFER, T., PALISTRANT, C., AND SHARP, R. 2005. Walrus: Wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys'05)*. 191–203.

ENGE, P. AND MISRA, P. 1999. Special issue on GPS: The global positioning system. *Proc. IEEE*. 3–172.

FRAMPTON, K. D. 2006. Acoustic self-localization in a distributed sensor network. *IEEE Sensors J. 6,* 1, 166 – 172.

GIROD, L. AND ESTRIN, D. 2001. Robust range estimation using acoustic and multimodal sensing. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '01)*. Vol. 3. 1312–1320.

GIROD, L., LUKAC, M., TRIFA, V., AND ESTRIN, D. 2006. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys'06)*. 71–84.

HARTER, A., HOPPER, A., STEGGLES, P., WARD, A., AND WEBSTER, P. 1999. The anatomy of a context-aware application. In *Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99)*. 59–68.

HE, T., HUANG, C., BLUM, B. M., STANKOVIC, J. A., AND ABDELZAHER, T. 2003. Range-free localization schemes for large scale sensor networks. In *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom '03)*. 81–95.

JI, Y., BIAZ, S., PANDEY, S., AND AGRAWAL, P. 2006. Ariadne: A dynamic indoor signal map construction and localization system. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services (MobiSys'06)*. 151–164.

KUSHWAHA, M., MOLNAR, K., SALLAI, J., VOLGYESI, P., MAROTI, M., AND LEDECZI, A. 2005. Sensor node localization using mobile acoustic beacons. In *Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems Conference (MASS'05)*.

MAROTI, M., VOLGYESI, P., DORA, S., KUSY, B., NADAS, A., LEDECZI, A., BALOGH, G., AND MOLNAR, K. 2005. Radio interferometric geolocation. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (SenSys'05)*. 1–12.

NICULESCU, D. AND BADRINATH, B. R. 2003. Ad hoc positioning system (APS) using AOA. In *Proceedings of the 22st Annual IEEE Conference on. Computer Communications (INFOCOM '03)*.

PENG, C., SHEN, G., ZHANG, Y., LI, Y., AND TAN, K. 2007. Beepbeep: A high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems (SenSys'07)*. 1–14.

PENG, C., SHEN, G., ZHANG, Y., AND LU, S. 2009. Point&Connect: Intention-based device pairing for mobile phone users. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys'09)*. 137–150.

PRIYANTHA, N. B., CHAKRABORTY, A., AND BALAKRISHNAN, H. 2000. The cricket location-support system. In *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*. 32–43.

SALLAI, J., BALOGH, G., MAROTI, M., LEDECZI, A., AND KUSY, B. 2004. Acoustic ranging in resource-constrained sensor networks. In *Proceedings of the International Conference on Wireless Networks (ICWN'04)*.

SAVVIDES, A., HAN, C.-C., AND STRIVASTAVA, M. B. 2001. Dynamic fine-grained localization in ad-hoc networks of sensors. In *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom'01)*. 166–179.

SHEN, G., LI, Y., AND ZHANG, Y. 2007. Mobius: Enable together-viewing video experience across two mobile devices. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services (MobiSys '07)*. 30–42.

SHEN, G., PENG, C., LI, Y., SHI, C., ZHANG, Y., AND LU, S. 2009. Dita: Enabling gesture-based human-device interaction using mobile phone. http://research.microsoft.com/en-us/people/jackysh/dita.pdf.

STOLERU, R., VICAIRE, P., HE, T., AND STANKOVIC, J. A. 2006. Stardust: A flexible architecture for passive localization in wireless sensor networks. In *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys'06)*. 57–70.

WEBR, J. AND LANZL, C. 1998. Designing a positioning systems for finding things and people indoors. *IEEE Spectrum 35,* 9, 71–78.

WHITEHOUSE, K. AND CULLER, D. 2002. Calibration as parameter estimation in sensor networks. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*. 59–67.

WHITEHOUSE, K., KARLOF, C., WOO, A., JIANG, F., AND CULLER, D. E. 2005. The effects of ranging noise on multihop localization: an empirical study. In *Proceedings of the 4th International Conference on Information Processing in Sensor Networks (IPSN '05)*. 73–80.

WIKIPEDIA. Speed of sound. http://en.wikipedia.org/wiki/Speed_of_sound.

YOUSSEF, M. AND AGRAWALA, A. 2005. The horus wlan location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys'05)*. 205–218.

YOUSSEF, M., YOUSSEF, A., RIEGER, C., SHANKAR, U., AND AGRAWALA, A. 2006. Pinpoint: An asynchronous time-based location determination system. In *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services (MobiSys'06)*. 165–176.

Received September 2008; revised October 2010; accepted December 2010