BeepBeep: A High Accuracy Acoustic Ranging System using COTS Mobile Devices

Chunyi Peng[1], Guobin Shen[1], Yongguang Zhang[1], Yanlin Li[1,2], Kun Tan[1]
[1]Microsoft Research Asia, Beijing, 100080, China
[2]Tianjin University, Tianjin, 300072, China
{chunyip,jackysh,ygz,v-yanlli,kuntan}@microsoft.com

Abstract

We present the design, implementation, and evaluation of BeepBeep, a high-accuracy acoustic-based ranging system. It operates in a spontaneous, ad-hoc, and device-to-device context without leveraging any pre-planned infrastructure. It is a pure software-based solution and uses only the most basic set of commodity hardware – a speaker, a microphone, and some form of device-to-device communication – so that it is readily applicable to many low-cost sensor platforms and to most commercial-off-the-shelf mobile devices like cell phones and PDAs. It achieves high accuracy through a combination of three techniques: two-way sensing, self-recording, and sample counting. The basic idea is the following. To estimate the range between two devices, each will emit a specially-designed sound signal ("Beep") and collect a simultaneous recording from its microphone. Each recording should contain two such beeps, one from its own speaker and the other from its peer. By counting the number of samples between these two beeps and exchanging the time duration information with its peer, each device can derive the two-way time of flight of the beeps at the granularity of sound sampling rate. This technique cleverly avoids many sources of inaccuracy found in other typical time-of-arrival schemes, such as clock synchronization, non-real-time handling, software delays, etc. Our experiments on two common cell phone models have shown that we can achieve around one or two centimeters accuracy within a range of more than ten meters, despite a series of technical challenges in implementing the idea.

Categories and Subject Descriptors

C.3.3 [**Special-Purpose and Application-based Systems**]: Real-time and embedded systems; C.5.3 [**Microcomputers**]: Portable devices

General Terms

Algorithms, Design, Performance

Keywords

Ranging, Acoustic ranging, Proximity detection, Localization, Ranging systems

1 Introduction

In this paper, we are interested in high accuracy ranging using only the most basic set of commodity hardware capabilities: a speaker, a microphone, and some form of inter-device communication. Such technique, if feasible, will have many desirable features and will be widely applicable in many sensing and mobile applications. This is because the set of capabilities can be considered as a common denominator of many sensor platforms and mobile devices, including many commercial off-the-shelf (COTS) devices like cell phones, PDAs, MP3 players, etc. Compared to alternatives that require special-purpose hardware (such as [1, 2]) or pre-existence of precision location infrastructure [3], a commodity-based solution will obviously have wider applications and cost less. For the same reason, we further desire a solution that can be implemented in software, preferably entirely in user-space.

High accuracy ranging is typically achieved through measuring time-of-arrival (TOA) information of acoustic or radio signals [1–6]. The distance is thus the product of the signal speed and the time of flight of the signal traveling between two devices. Obviously, the ranging accuracy depends on the signal speed and the precision of TOA measurement. To elevate the accuracy, acoustic signals are usually chosen because of their relative slow speed. But the precision of TOA measurement remains a big challenge in any system implementation.

In practice, TOA measurement is often done with both sides taking a timestamp of their respective local clock at the moment the signal is emitted or received. There are three intrinsic uncertainties in this process that can contribute 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

SenSys'07, November 6–9, 2007, Sydney, Australia.

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 2.2).

It is therefore extremely challenging to provide high accuracy ranging in a software-only and device-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 be able to change kernel or driver. In fact, many COTS devices like 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 the above timestamping approach will not be able to provide the high accuracy we desire.

In this research, 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. 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 of each other. 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 it 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 not better than, the best results ever reported in the literature that use special hardware design or complex signal processing.

To summarize, we have made the following contribution. First, we identified the three major uncertainties common to any time-of-arrival based ranging system and evaluated them on COTS mobile devices. Secondly, we proposed the BeepBeep ranging mechanism that cleverly overcomes all these uncertainties. Thirdly, we designed and implemented the BeepBeep ranging system, purely in software. Finally, we systematically evaluated the system and our design choices under several typical indoor and outdoor environments using COTS mobile devices. We have achieved centimeter accuracy, the best ever reported in the literature.

The rest of the paper is organized as follows: we present the motivating scenarios and the challenges with TOA-based systems in Section 2. The detailed BeepBeep ranging mechanism is presented in Section 3. We describe the software system architecture and implementation of the BeepBeep ranging system in Section 4. The performance of the system is evaluated in Section 5. Related work is reviewed in Section 6, followed by in-depth discussions in Section 7. Finally, Section 8 concludes the paper and highlights our future work.

2 Motivation and Challenges

2.1 Motivation

High accuracy ranging and localization systems have been an active research theme in the wireless sensor network research field [1, 7–12]. As we have discussed earlier, if a range technique can achieve similar or better accuracy but use only most basic set of hardware, it will be applicable to more platforms and suitable for more applications. It may further reduce cost to do ranging in sensor networks.

Besides sensor networks, we believe that ranging or proximity information can also be very useful in everyday's mobile applications. For example, multi-device applications like precision asset location [13] and touch-to-connect pairing in Bluetooth 2.1 [14], collocated multi-users applications like spontaneous interaction and collaboration [15], simultaneous photo sharing [16, 17], and "better-together" video viewing [18], can all benefit from high-accuracy ranging. With high-accuracy ranging, fine-grained spatial control can be provided and context-aware systems can be developed. For example, sharing can be automatically terminated once a party goes outside a certain proximity in a co-located sharing scenario. Similarly, the video playback should be dynamically expand to the two screens or shrink to one screen as the other device comes and goes in the together-viewing scenario [18]. The ability to provide high accuracy ranging with commodity software and hardware is even more appealing here, as it can be readily used in a huge volume of COTS mobile devices.

2.2 Challenges of TOA Estimation

Time-of-arrival based system estimates the distance D between the sender and the receiver to be the product of the time-of-flight, i.e., 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. But even if 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 centimeters error in the ranging result.

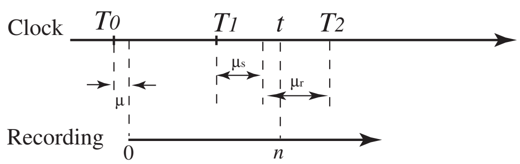


Figure 1. Event time line in the uncertainty experiement

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 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 [19] 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 [20]. Yet most solutions have resorted to dedicated mechanisms [1, 2, 4, 11].

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.

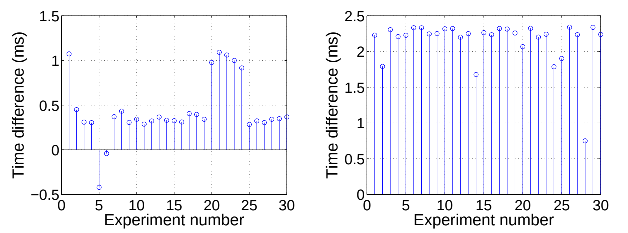
There has been little work in addressing the sending and receiving uncertainties in software. Most previous work managed to minimize them by resorting to customized hardware design so that the system can precisely control and obtain the exact instant when a signal is sent or received [3, 4]. This is 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 send 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. Immediate 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, i.e., 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):

$$\mu_s + \mu_r > T_0 + \mu + n/f_s - T_1 > T_0 + n/f_s - T_1 \quad (2)$$

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

Figure 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.

3 BeepBeep Ranging Mechanism

As briefly mentioned earlier, the BeepBeep system provides high accurate ranging results relying only on the capability of COTS mobile devices. Our system can very well handle all the three aforementioned uncertainties, namely clock synchronization, sending and receiving uncertainty. In this section, we elaborate the design of the BeepBeep ranging mechanism and explain the subtlety that leads to the high precision results.

3.1 Basic ranging scheme

We start from describing the basic ranging procedures with only two devices, say A and B, and we will extend this to multiple devices later.

The basic ranging scheme takes three steps. In the first step, a two-way sensing is performed, as shown in Figure 3. Assume both devices are in recording state. Device A first emits a sound signal through its speaker S_A. This signal will be recorded by its own microphone (self-recording) as well as the other device B. Then, device B emits another sound signal back through its speaker S_B. This signal is also recorded by both microphones on the two devices. 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 time-of-arrivals* (ETOA) [1]. The two devices will exchange their locally measured ETOA and in the final step, the distance between the

[1] We use the term ETOA here in order to differentiate from the

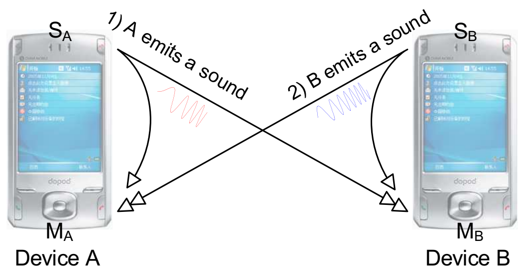


Figure 3. Illustration of the two-way sensing stage in the BeepBeep ranging procedure.

two devices can be simply computed based on these two values.

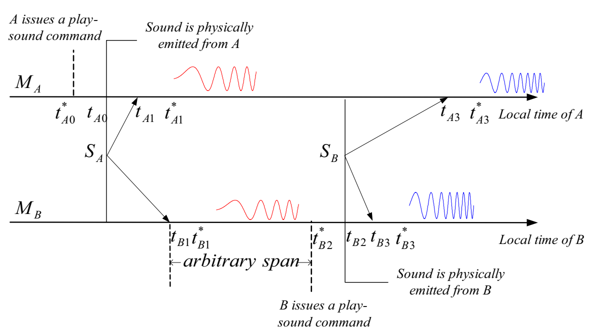


Figure 4. Illustration of event sequences in BeepBeep ranging procedure. The time points are marked for easy explanation and no timestamping is required in the proposed ranging mechanism.

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 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, the actual time when the speaker physically emits may be t_{A0}. The time of the signal arrives 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

$$d_{A,A} = c \cdot (t_{A1} - t_{A0}) \quad (3)$$
$$d_{A,B} = c \cdot (t_{B1} - t_{A0}) \quad (4)$$
$$d_{B,A} = c \cdot (t_{A3} - t_{B2}) \quad (5)$$
$$d_{B,B} = c \cdot (t_{B3} - t_{B2}) \quad (6)$$

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

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

In Equation 7, the latter two terms are the distances between the speaker and microphone of the two devices. This distance is a constant to a certain 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, i.e., 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

well defined term DTOA (differential times of arrival) or TDOA (time differences of arrival) which usually refers to the differential between two TOAs measured at two different receivers using the same sound source.

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, Equation (7) can be rewritten as

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

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 $K = d_{B,B} + d_{A,A}$ is a constant. In the rest of the paper, 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 Equation (8) can be simplified to

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

By using sample counting instead of timestamping, our mechanism avoids the source of inaccuracies found in traditional timestamp approaches. In fact, our mechanism has no notion of local clock or timestamp at all. From Equation (8), 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.

3.3 Signal Design and Detection

To achieve high ranging precision, it is critical to precisely locate the first signal sample in recorded sound samples. This is particular challenging for COTS mobile devices, since in general, the speakers and microphones in such devices have only limited capability, e.g. narrow spectrum support. Furthermore, when working in an indoor environment, sound signal could arrive at the destination through multiple paths with different delay. This *multipath effect* may cause ambiguous ETOA detection and therefore significantly reduce the detection accuracy if not handled well.

We will address the signal design and detection algorithm in detail in Section 4.

3.4 Sources of Errors

We summarize the possible sources of errors in this section. According to Equation (9), there are three possible sources of errors, relating to the three parameters, namely sound speed c, sampling frequency f_s, and TOA detection (i.e., various sample indices n_t). 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.

While ETOA avoids associating the TOA of sound signal to the local clock of the end system, there are still other factors that may influence the TOA detection precision.

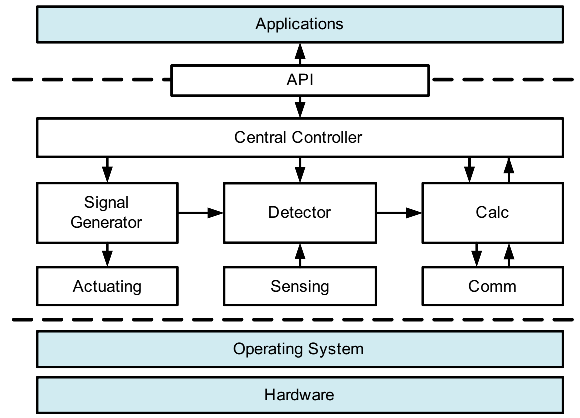


Figure 5. Software architecture of BeepBeep system.

- *Signal to noise ratio (SNR)* – the received sound signal will be attenuated and distorted by the communication channel. Furthermore, the environmental noise may be usually colored. SNR is also affected by energy used when transmitting the signal at the sender.
- *Multipath effects* – the acoustic signal may reach the receiver via different paths, due to reverberation. The received signal is thus the combination of signals from all the possible paths that traverse the position of the microphone.
- *Signal distortion* – The hardware (e.g., microphone and the speaker) of a mobile device usually have good support only for a very limited spectrum band (e.g., around 3kHz) since their primary usage is voice communication. The attenuation differs with different frequency bands. The dynamic range of the speaker's volume is also very limited. It is very easy to get saturated and cause large waveform distortion.

4 System Architecture and Implementation

4.1 Overview

Unlike other ranging or localization systems, the BeepBeep ranging system is purely a software solution that does not require specialized hardware design nor modifications to the commercial operating system. The BeepBeep system can be implemented purely in the application-layer, and is ready for deployment on most ordinary COTS mobile devices.

We further architect the BeepBeep system as a ranging service so that it can be readily used by other applications. Figure 5 shows the overall software architecture. It has three major parts: the interface to other applications (API), the core logic part, and the underlying physical device related function modules. 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 light-weight 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. It also interacts with other applications by receiving requests and sending back responses through API. A local timer is maintained in the central controller for ranging signal scheduling. The signal generator generates the waveform of ranging signals based on given parameters and feeds the signal to the 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 Equation 9.

4.2 Acoustic Signal Design

The acoustic signal should be designed to have good autocorrelation property, which permits accurate signal detection when presenting with ambient noise. One typical signal design that fits our requirement is the linear chirp signal, but the range of its spectrum has to be limited to obey by the constraints of the hardware design of speaker and microphone in COTS devices. Because most of these hardware are designed bearing in mind the primary application being voice conversation, it is natural that they have good frequency response only around the narrow spectrum band of human voice. Figure 6 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 choose the frequency range of the linear chirp signal to be between 2–6kHz. Another issue we meet is that the sound waveform played out has very large distortion in the first few milliseconds. We think it might be caused by the speaker diaphragm inertia. To resolve this, we precede the chirp signal with five milliseconds 2kHz cosine waveform to warm up the speaker. In our implementation, we choose the signal length to be 50 milliseconds which strikes a good compromise between multipath effects suppressing and noise resistance.

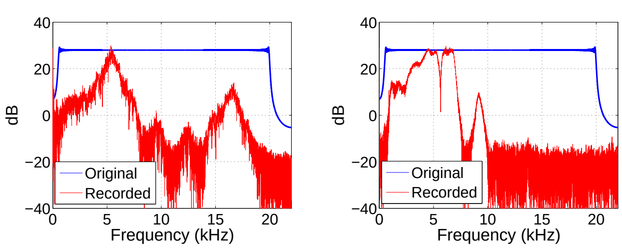


Figure 6. Frequency response of the HP iPAQ rw6828 (left) and Dopod 838 smartphones (right).

4.3 Signal Detection

The signal is detected by correlation with the reference chirp signal 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 recorded data are correlated with the reference signal and the maximum "peak" is located. This maximum peak is concluded 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 $L_2(S)/L_2(N) > TH_{SD}$. If no such quantified point is found, we conclude that the detection failed, which could be because the signal energy is too weak or the noise level is high. In our implementation, we set $TH_{SD} = 2$ (i.e., 3dB) and $w_0 = 100$.

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 7. 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 $w_1 = 100$.

4.4 Ranging Protocol

In this subsection, we describe the ranging protocol used in the BeepBeep system. Here, we assume that each device

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

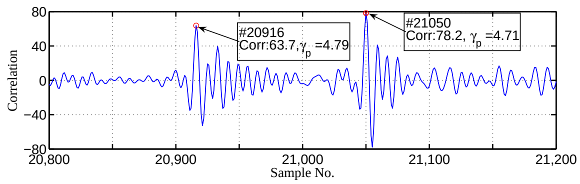


Figure 7. One example case of multipath effect.

has a WiFi radio, and thus, all devices are coordinated with wireless communications. Our protocol can support multiple devices in one ranging process, where each of N $(N \geqslant 2)$ devices is trying to measure distances to all other devices simultaneously. Our protocol will generate only N acoustic signals to obtain all pair-wise distance measurements between any two devices. This property is critical to ensure the scalability of our ranging mechanism when N is large.

Basically, our protocol contains three steps:

1. *Initiation*. A ranging process is started by an initiating device, which calculates and disseminates a schedule in an *initiation message* to all the devices that participate in the ranging process.
2. *Sensing*. Each device calculates a delay according to the schedule and sets a timer. Upon the timer expires, it emits a sound signal.
3. *ETOA Exchanging*. After the last device has emitted the sound signal, each device processes the recorded signals and determine ETOA between its own signal and signals from all other devices. These ETOA values are packed into one packet broadcast to other devices. Upon receiving ETOA information from all other devices, one device can calculate the distance to all other devices using Equation (9).

In the *Initiation* step, the initiating device randomly chooses an order of each device to emit acoustic signal and specifies a time interval between two consecutive transmission of signals. Defining such a schedule has two purposes: 1) it schedules each device to emit acoustic signal at different time to prevent possible collisions; and 2) it also helps identifying the signal of each device. It is because acoustic signals like chirp, are identical for all devices, and it is required for a device to have an one-to-one mapping between the detected signals and the ranging peers in order to calculate ETOA correctly.

After receiving the schedule, each device will start recording with microphone and calculate a proper delay (interval multiplied by its order in the schedule starting from the instance when the initiating message is received) before it transmits the chirp signal. Note that since the delay is calculated by each device based on its local clock, it is possible that the schedules calculated by different devices have slight skews. To accommodate this, the interval between two consecutive sound signals should be large enough to prevent signal overlaps from different devices. In our implementation, we find an interval of one second can reliably separate sound signals from different devices.

After the last device has emitted the sound signal, all devices process their recorded data and search for chirp signals. A chirp signal is related to a device if the signal is detected within the time window of that device according to the pre-defined schedule. Note that it is possible that the signal detection fails. For example, the corresponding device is too far away for sound to reach but still within in the range of WiFi. All measured ETOAs between a device and other devices including detection failures will be exchanged in the last step using broadcast. After receiving the broadcasts from all other devices, a device can calculate its distance to others or re-initiate a new ranging process if failures have occurred.

There can be multiple groups of devices that want to conduct ranging process simultaneously and may contend for the acoustic channel. This contention is resolved by preventing two nearby initiators from starting ranging processes simultaneously. Every device will listen to *initiation messages* from other devices. If an initiator receives an *initiation message* from a nearby device, it should defer the transmission of its own *initiation message* until the end of that ranging process. Note that in some rare case, it is still possible for two ranging processes happening concurrently if one initiator fails to reliably receive the broadcast *initiation messages*. As a consequence, multiple chirp signals may be found in one time window and a collision is detected. Since one device can not differentiate which signal is from the corresponding ranging peer or from a contending device nearby, it should report a failure and the two initiators should restart their ranging process later after a randomly backoff.

4.5 Prototype Implementation

We have implemented the BeepBeep ranging system in Windows Mobile 5.0. We develop it as a user-mode dynamic linkable library that other applications can load and use it for ranging service. We use multimedia services (*WaveXXX* series APIs) embedded in Windows Mobile to control microphones and speakers and rely on WinSock for communication over WiFi. We further develop two demo applications that utilize the BeepBeep system to determine the distance among several mobile devices.

5 BeepBeep System Evaluation

5.1 Hardware Configuration

We have deployed the BeepBeep ranging system onto two models of commercial off-the-shelf PocketPC phones, HP iPAQ rw6828 and Dopod 838, as shown in Figure 8. Both devices are running Microsoft Windows Mobile Version 5.0 (Phone Edition), with WiFi and Bluetooth radios and InfraRed interface, QVGA display, 64 MB RAM, two built-in speakers and one 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 using Dopod 838 cell phone. The experimental results on the HP iPAQ rw6828 phone are similar. Unless explicitly pointed out, all the 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

Figure 8. Two COTS mobile devices used in the evaluation of BeepBeep ranging system.

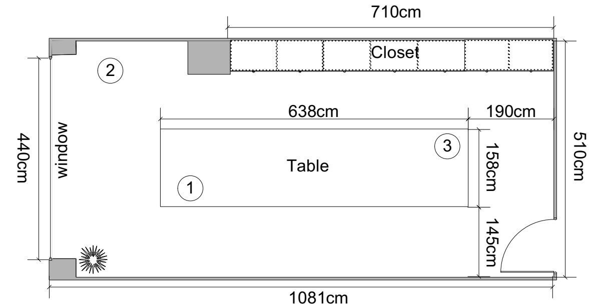


Figure 9. Indoor testbed for ranging experiment.

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 make $K = 0$ by putting the earphone very close to the microphone.

5.2 Performance Metrics

We use the following metrics to evaluate the BeepBeep ranging system:

- *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. Such errors and standard deviation can be represented by the *absolute* values or the percentage to the real distance.
- *Confidence*: confidence is defined as the percentage of the times a known level of accuracy is reached. As confidence depends on the specific accuracy level, in our evaluation, we define 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 error (e.g., 1%).
- *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%. This metric is informative as it heavily depends on the experimental environment such as the room size and the devices' capabilities such as the properties of the speaker and microphone.

5.3 Test Case Design

We evaluate the BeepBeep ranging system under four environments:

- *Case-A – Indoor, quiet*: a meeting room that is approximately 5m$\times$11m with a big table in the center, as shown in Figure 9 where different test locations are also illustrated. The environmental temperature is 25$^\circ C$.
- *Case-B – Indoor, noisy*: the same room as Case-A but with background noise consisting of the air conditioner noise, background pop music (at various volume levels and various relative positions between the sound box and the smartphones), and people chatting around the table. The temperature is 25$^\circ C$.
- *Case-C – Outdoor, car park*: the open area in front of a medium size car park at the side of a large building and nearby the main driving course. The weather is windy and wind speed is about 10 miles per hour. The temperature is 10$^\circ C$.
- *Case-D – Outdoor, subway station*: the open area at the entrance of a subway station, just after the rush hour and with medium traffic. The temperature is 15$^\circ C$.

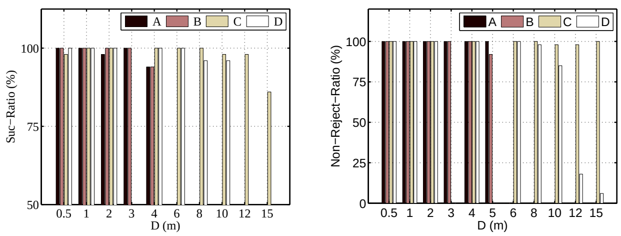
In all the experiments, we have placed the devices parallel to each other and facing up, as depicted in Figure 3, and ensured line-of-sight between the two devices. For every different setting (i.e., different distances under different test cases), the experiments were repeated 50 times.

5.4 Experimental Results

In our experiments, the sound speed used in the distance calculation is set according to the following model [21]: $c_{air} = 331.3 + 0.6 \cdot \theta$ where 331.3 m/s is the benchmark speed at 0$^\circ C$ and θ represents the air temperature in Celsius ($^\circ C$).

From the experimental results, we find there are some obviously failed experiments yielding negative or meters-in-error results. Including these data would make the results less informative because a single such datum will significantly alter the whole accuracy metric. Such failed experiments can be easily detected by speculating the results. Therefore we have excluded them from the accuracy evaluation using a simple criterion: only results with less than 20cm ranging error are included for accuracy evaluation. Figure 10-(a) shows the percentage of such successful experiments, i.e., those used in accuracy evaluation, at different distances under the four test cases. Clearly, only very few experiment results are excluded. Note that in our design, we adopted a threshold TH_{SD} to indicate possible failure of the signal detection. To evaluate its effectiveness, we plot the ratio of experiments that pass the threshold for all the measurement data in Figure 10-(b) where the TH_{SD} is set to 3dB. Comparing the two figures, we can see that the differences are very small which confirms the validity of our design.

Note but *all* the experiments, successful or failed, are included in the α-confidence evaluation so that we can tell how confident our ranging results are in practice. In other words, for a field ranging application run, we can have α-confidence (e.g., 95%) that the result is successful and the possible error for this specific result is less than α centimeters.



(a) 20cm criterion (b) $TH_{SD} = 3dB$ criterion

Figure 10. Percentage of successful experiments at different distances for the four test cases. Left: less than 20cm ranging error criterion. Right: TH_{SD}=3dB criterion.

Furthermore, in the rest of this section, we have followed this convention when presenting the accuracy evaluation results: in the 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 represents the mean ranging error and the solid line shows the standard deviation.

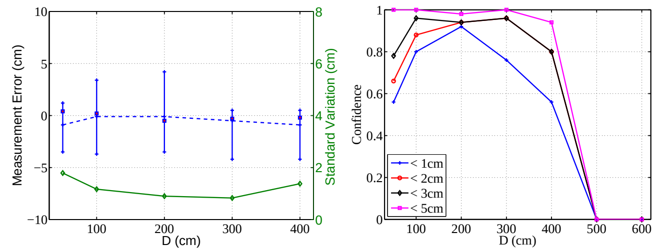
5.4.1 *Indoor Cases*

Figure 11-(a) shows 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 $\pm$1cm and the standard deviations are within 2cm.[3] Figure 11-(b) shows the corresponding α-confidence plots where α equals to 1cm, 2cm, 3cm and 5cm. High percentage of experiments lead to less than 1cm accuracy and is very robust in 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 error within 5cm when the distance is 4 meters, but the number 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. As can be seen from the figures, the BeepBeep ranging system still performs very well. The overall resulting ranging accuracy is comparable to Case-A. These results reflect the excellent noise resistent property of the chirp signal. The comparison between Figure 11 and 12 further reveals that background noises only leads to some decreases in the 1cm-confidence, but has no much impact on the 2cm-confidence and above.

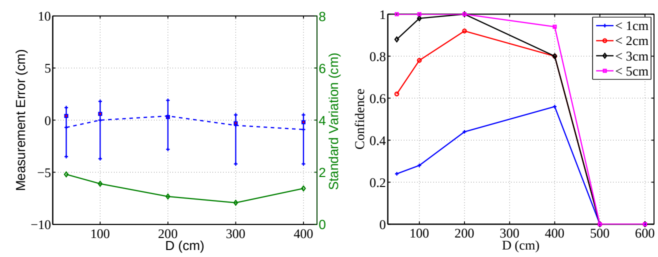
From above experimental results, we can conclude that the operational ranges are both 4 meters for the specific settings of test cases Case-A and Case-B. As will be explained later, the primary reason is the multipath effects caused by the small size of meeting room. In fact, when the distance

[3]Note that we plot the mean ranging error without taking absolute values here to show the accuracy we can achieve if multiple experiments are allowed and to provide some hints of the ranging error distribution. The median ranging error and the standard deviation are more informative to the accuracy of a single measurement. In Table 1, we use absolute value of ranging errors to give the overall performance of our BeepBeep system.



(a) Accuracy (b) Confidence

Figure 11. Accuracy and confidence measurement results at different distances in the Case-A setting. Left and right vertical axes are the reference scales for the ranging error and the standard deviation, respectively. The dash line represents the mean ranging error, the square point is the median ranging error, the cross points are the max/min ranging error, and the solid line shows the standard deviation.



(a) Accuracy (b) Confidence

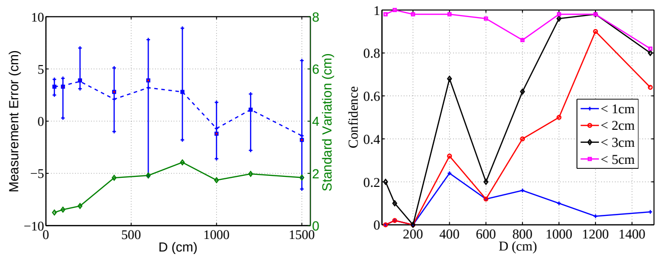
Figure 12. Accuracy and confidence measurement results at different distances in the Case-B setting.

is 5 meters or above, one device is already very close to a side wall, which led to strong reflection of the sound signal.. Nonetheless, as long as within the operational range, our system works reliably and the ranging results are highly confident.

5.4.2 *Outdoor Cases*

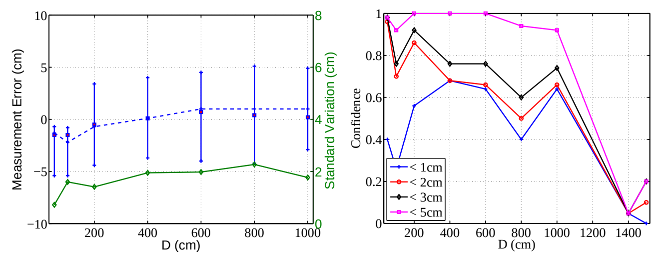
The accuracy and confidence measurements for Case-C and Case-D environments are shown in Figure 13 and Figure 14, respectively. The BeepBeep ranging system still works well. The resulting median and mean ranging errors for Case-C are within [-2cm, 4cm] and those for Case-D are within [-1.5cm, 1cm]. Compared with the indoor cases, the ranging error increases and shows a larger dynamic range. However, the standard deviation is still very small (mostly less than 2cm), which demonstrates the robustness of our system.

The confidence plots of the two cases differ significantly. While that of the Case-D is normal and as expected, the α-confidence plots (for α equals to 1cm, 2cm, and 3cm) for Case-C seems quite abnormal. We suspect it is a time-of-the-day effect and affected by the car traffic, because we started the experiment at 6:30pm (i.e., off the work time) and the venue is at the entrance of the car park and close to the main driving course. 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 evidenced in Figure 10-(b). In contrast,



(a) Accuracy (b) Confidence

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



(a) Accuracy (b) Confidence

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

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.

In above 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 the figure, we can see that for all the test cases, our ranging results are highly reliable within their operational ranges. For instance, the probability that our system will lead 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.

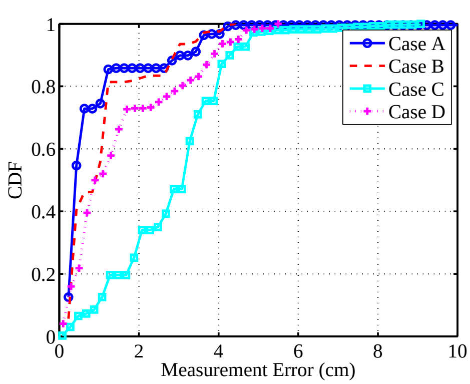
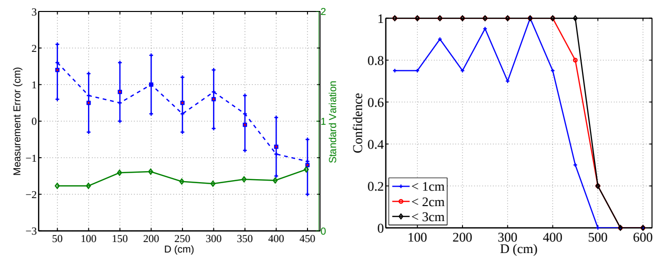


Figure 15. Cumulative distribution function of the α-confidence for the four test cases.

5.4.3 *Impact of Signal Distortion*

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



(a) Accuracy (b) Confidence

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

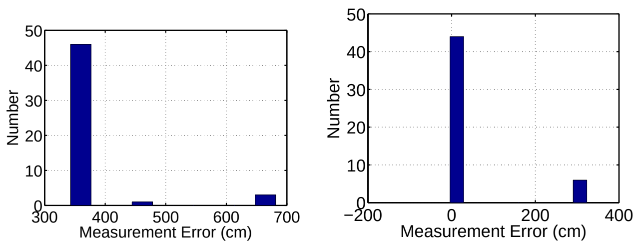
From the figure, we see that using earphone leads to excellent accuracy with less than 1cm mean and median ranging error and the maximum ranging error is within ±2cm. The ranging results are also super stable. The average and maximum standard variations among all the test distances are 0.47cm and 0.6cm,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

Comparing Figure 16 against Figure 11, we can observe 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 observation suggests that the impacts of signal distortion is even larger than the signal to noise ratio.

5.4.4 *Multipath Effect Mitigation*

The multipath effect usually becomes evident when the signals from non-LOS paths have comparable strengths with the LOS signal. The effect will be exaggerated when the lengths of non-LOS paths are close to that of the LOS path as it will cause interferences. This is the primary reason why experiments fails when the distance is large for both the two indoor cases. In those experiments, due to the space limit, we have put one device close to some walls and stronger multipath effects appear in consequence. We did not notice such effect in the Case-C setting.

Figure 17-(a) shows the distribution of measurement error at a 5-meter distance in the Case-A setting if we simply determine the TOA sample according to the maximum cross-correlation value, i.e., without multipath effect mitigation effort. Clearly, all the 50 experiments are failed and the errors are larger than 3 meters. Figure 17-(b) shows the resulting measurement error distribution where our heuristic multipath effect mitigation algorithm is applied. All the



(a) No mitigation (b) With mitigation

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

cases with more than 4 meters errors and most cases with 3 meters errors are correctly handled. There percentage of successful experiments becomes 88%. However, there are still a few error cases left over, which suggests more research work. Note that, in previous evaluation results, the multipath mitigation algorithm has not been applied. If it were applied, the operational range would be significantly improved, at least for some relaxed accuracy or confidence requirements.

5.4.5 Evaluation Summary

Table 1 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* value 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 the test cases. The operational range for the indoor cases is around 4 meters (in our specific experiment room, but in general constrained by the room size) and that for outdoor cases is in general larger than 10 meters. As long as the devices are within the operational range, our system works reliably and the ranging results are highly confident.

6 Related Work

There have been tremendous efforts on the localization problem on theoretical studies [7, 22, 23] as well as building practical systems using angulation or lateration [3, 4, 24], proximity sensing [25, 26], or even scene analysis [10]. While range information can certainly be calculated from the locations, in this section, we mainly review those work that involves ranging and whose ranging scheme is closely related to ours such as using acoustic signal and (possibly) achieves high ranging precision. For a general review of localization systems, we refer readers to [26] where an excellent review was provided. Since one of our primary target (or rather, constraint) is the COTS devices, we also summarize the available systems that are built out of COTS devices.

6.1 TOA-based Systems

Most existing high accuracy ranging schemes and lateration-based localization systems rely on time-of-arrival (TOA) of acoustic signal, with the few exceptions being the RIPS system [9] that is based on radio interferometric yet still achieves high precision up to several centimeters using only radio signal.

The Bat system [3, 27] is a high accuracy indoor localization system, based on time-of-flight of ultrasonic signal. It can achieves an accuracy up to 3 centimeters for most cases. The system relies on an infrastructure consisting of an irregular matrix of networked, ultrasonic receivers daisy-chained together above the ceiling of the room. The receivers are synchronized and orchestrated using a radio channel. The Cricket location system [4] infers distance using the concurrent radio and ultrasonic signals and their differential time of arrival, and further let users infer their positions by listening to the wall/ceil-mounted beacons that form an infrastructure. Good ranging accuracy is achieved with specially designed hardware that ensures instantaneous timestamping of the arrivals of radio and ultrasonic signals.

While the proposed BeepBeep ranging system bears the similarity with above systems in the sense that it also uses acoustic signal and rely on time-of-arrival to calculate the distance, the differences are also obvious, including no need of special hardware design, no dependency on the radio signal,[4] and/or no requirement of clock synchronization among devices and no dependency on the infrastructure.

ENSBox [1] is a platform designed for rapid prototyping distributed acoustic sensing systems with a prominent self-calibration feature. A novel position estimation algorithm was proposed that integrate the microphone array orientation information. High accurate ranging and localization results, up to few centimeters, were achieved. We want to point out that in the design and implementation of the ENSBox, the authors also observed the necessity of local detection and exerted self recording. They used sample counting method to obtain the more accurate timing of the instant at which the sound is emitted. That is, they solved the sending uncertainty to some extent. However, since they still needed to obtain the timestamps with reference to the local clock of each sensor node, tremendous efforts were thus spent to achieve high precision clock synchronization. In contrast, BeepBeep uses a two-way sensing strategy that avoids referring to the local clock and completely removes the requirement on the clock synchronization, which in return enables BeepBeep to be directly applicable to ordinary COTS mobile devices. BeepBeep further removes the receiving uncertainty and delivers more accurate ranging results.

The two-way sensing strategy of BeepBeep also bears some similarity to those schemes where a round-trip time (RTT) measurement is adopted, such as [20], with the difference being that we do not require the other device to bounce back the signal immediately upon receiving. In other words, we allow arbitrary delay between the two trips. Moreover, relying on the signal reflection may have certain requirement on the target size. Note that two-way sensing is not the only beauty of BeepBeep, other techniques like self-recording and sample counting are equally important.

Finally, we want to mention PinPoint [2] which achieves impressive accuracy (up to several feet) using TOA of radio signal directly. Their key idea is a mathematical way for clock difference compensation which leads to very precise timestamp recovery while allowing the nodes' local clocks to run asynchronously. BeepBeep ranging system differs from

[4]The use of WiFi in our system is optional and adopted for convenience. We may use Bluetooth or even directly using audio channel [28].

Environment Setting	Operational Range	Confidence Level	$\text{Avg}_d\,(\text{Avg}_n(\lvert Err\rvert))$ (cm)	$\text{Max}_d\,(\text{Avg}_n(\lvert Err\rvert))$ (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

Table 1. Summary of all experimental results. Subscript d and n indicate that the operations (i.e., avg, max) are conducted over different measured distances for each test case and over n ($n = 50$) times experiments, respectively.

PinPoint in that PinPoint uses radio signal to ranging and requires special hardware design 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.

As a remark, we would like to point out that another category of localization systems, such as microphone arrays [29], utilize differential time-of-arrival (DTOA), a slight variation of TOA. The sound source speaks out a sound, all the receivers (with known positions) record the sound. Through the combinations of DTOAs, the position of the sound source can be determined. In [12], the DTOA idea is reversely used where clock-synchronized sound sources at different locations in return send out sound signal at controlled time and sensors report their detected DTOA to a centralized server for localization through nonlinear optimization. Their idea on the innovative use of DTOA is similar to our use of ETOA. The most differentiating feature that sets apart the BeepBeep ranging system is the complete avoidance of the clock synchronization among devices, while in the two class of DTOA based schemes above, synchronization must be guaranteed among either all the senders or all the receivers.

6.2 System out of COTS Devices

Many localization approaches have been proposed in the literature [30], with varying resulting precision and system complexity such as involvement of specially designed hardware. To the best of our knowledge, few works such as WALRUS [26], Radar [25], Horus [31] and ARIADNE [32] do not involve specially designed hardware.[5] These methods usually give rough proximity information instead of more precise ranging information. They also commonly depend on pre-planned infrastructures and operate in indoor environments.

More specifically, WALRUS can reach room-level precision of mobile devices location with ultrasound signal by utlizing the fact that ultrasound signal does not penetrate walls. It implicitly treats the static PC in each room as infrastructure or landmark. Radar uses received radio signal strength from multiple APs and determines the location using radio propagation model and the radio strength map obtained offline. Radar reaches a precision up to a few meters and leverage the pre-deployed APs as infrastructure. Bearing the same concept, Horus system identifies and combats several causes of wireless signal variation, proposes a location-clustering technique to improve computation efficiency and can achieve 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 only a single actual signal strength measurement and proposes a clustering algorithm for localization. ARIADNE can achieve about two meters accuracy.

Evidently, their dependency on the infrastructure, indoor operating environment setting, and achievable location precision make them not suitable for many of our targeting scenarios that is mobile and ad-hoc in nature and requires high precision.

[5]There are other two systems, namely Place Lab and E911/E112, that provide location service to COTS cell phones, but their precision is too low to be useful in our targeting scenarios.

7 Discussion

In the preceding sections, we have presented and evaluated the BeepBeep ranging system. We have shown that our system can achieve up to one centimeter precision within a few meters using our testing devices. This working range and accuracy may vary depending on the quality of the speakers and microphones on the devices. In general, devices that are equipped with higher fidelity speakers can lead to larger operational range with certain sound volume since their playouts of signals have less distortion. Similarly, a high sensitive microphone is further helpful to precisely detect the signal from background noise. Moreover, raising sound volume (but not cause extra waveform distortion) may also help in increasing the operational range, but this may consume more energy and is also more annoying as in current implementation the ranging signal is audible. With the prevalence of mobile multimedia applications, we believe more and more COTS devices will be equipped with high quality speakers and microphones.

We have adopted a simple linear chirp signal (bandlimited to 2–6kHz) in our ranging system. However, we do not claim that is optimal, and we put optimal design of ranging signal as a future work and some valuable insights have been discovered in [33]. Actually, we are working on some alternative signal designs. One of them is to use coded pseudo-noise (PN) signals. Our preliminary study shows a simple chaotic PN signal may achieve similar accuracy as our simple linear chirp. Further, in the ranging process, different devices may emit different PN signals using different codes. It is easier for each device to associate the signal with a certain device from the recorded data. Finally, with well-chosen codes, PN signals can be orthogonal to one another and therefore can be reliably detected even if multiple signals are overlapped. Such collision resilient property of PN signal suggests that schedule-based protocol may be significantly simplified or even exempted if PN signal are used.

As we have explained earlier, choosing a proper length of sound signal is a tradeoff. On the one hand, in order to achieve high SNR, we prefer the signal length to be long. On the other hand, long sound signals may suffer more due to multipath effect because the signals from secondary paths overlap more on the primary path signal. In current design, we have chosen a fixed signal length (i.e., 50ms) to balance these two requirements. However, in the future, an adaptive scheme can be used. For example, we can use shorter signals when the environment is relatively quiet but subjects to multipath interference (indoor environments); while we use longer sound signal in our-door environment where noise is the dominating factor affecting the ranging precision.

Since what being measured in our BeepBeep ranging system is the actual traversed path length of the two acoustic signals, due to the two-way sensing strategy, it is necessary to ensure the line-of-sight between the two devices so that the measured path length can be correctly converted to the physical distance via a fixed conversion function. In this case, the distance is simply half of the measured path length. As already shown in our derivation, there is also a calibration term (i.e., K in Equation (9)) if the microphone and the speaker are not located together. It is possible to utilize some "orthogonal" sensory modalities for the devices to become aware of the possible non-LOS conditions [33]. Another point worth mentioning is that as long as the LOS can be ensured, the orientation of the devices seems not matter much in the extra experiments we have conducted.

There are also some challenges if the BeepBeep ranging mechanism is to be directly applied onto the extremely weak sensor nodes like Mica2 motes, which has small memory buffer, low computation power and no dedicated I/O processor to perform the sampling of the incoming sound signal. Since the achievable ranging accuracy is directly proportional to the sampling rate, the constraint of small memory buffer can be worked around if the target precision can be relaxed. The computation power is less a problem since we do not require real-time signal detection, i.e., signal detection can be postponed. Because the sampling operation is controlled by the main microcontroller, it is critical that the ranging process should not be interrupted by other processes. This is probably doable with some customized protocol and a shortened ranging process.

Finally, we note that ranging with audible sound might interact with existing applications (e.g. multimedia), but we expect such interaction to be tolerable due to following reasons. Firstly, the ranging process is rather quick in practice (few seconds), and may only happen occasionally.[6] Secondly, as we have shown in the evaluation, the BeepBeep system is rather robust and resistent to both music and human conversations. Thirdly, according to our own experiences and the feedbacks from some labmates, the band-limited chirp signal is still ear pleasing. We believe the BeepBeep system can co-exist well with most of current applications.

[6]Some applications may require continuous measures of distance. However, we find there are still many heuristics can be used to reduce the ranging processes. For example, we may use the Signal Strength of WiFi (RSS) as an indicator of human mobility. We will redo ranging only when there is a significant change in RSS.

8 Conclusion

In this research, we have designed, implemented and evaluated the high-accuracy acoustic ranging system – BeepBeep. It is a pure software-based solution and uses only the most basic set of commodity hardware – a speaker, a microphone, and some form of device-to-device communication – to achieve centimeter accuracy. It operates in a spontaneous, ad-hoc, and device-to-device context without leveraging any pre-planned infrastructure. It is readily applicable to many sensor platforms and to most commercial-off-the-shelf mobile devices like cell phones and PDAs. We believe it will have wide applications in low-cost sensor networks as well as in a compelling set of social related mobile applications that desire the proximity awareness and the fine-grained control over the spatial relationship.

We identified the three uncertainties typically involved in the time-of-arrival measurements and carefully designed the BeepBeep ranging system with three unique features to conquer those uncertainties, namely the two-way sensing strategy to avoid clock synchronization uncertainty, the self-recording strategy to remove the sending uncertainty and the sample counting method to avoid the receiving uncertainty. Experimental results on handy cell phones demonstrate the superior accuracy and excellent consistence of the BeepBeep ranging system. It achieves about 1cm and 2cm average ranging accuracy with less than 2cm standard deviations for typical indoor and noisy outdoor environments, respectively. Because of the minimum hardware assumptions of the BeepBeep ranging system, we believe its simple yet effective ranging mechanism can be directly incorporated into the design of other customized sensor platforms and will lead to significant cost reduction.

To summarize, we have made the following contribution. First, we identified the three major uncertainties common to any time-of-arrival based ranging system and evaluated them on COTS mobile devices. Secondly, we proposed the BeepBeep ranging mechanism that cleverly overcomes all these uncertainties. Thirdly, we designed and implemented the BeepBeep ranging system, purely in software. Finally, we systematically evaluated the system and our design choices under several typical indoor and outdoor environments using COTS mobile devices. We have achieved centimeter accuracy, the best ever reported in the literature.

We have mainly focused on the ranging problem on COTS mobile devices and achieved good results so far. Our ongoing work is to systematically evaluate the various parameters used in the signal detection algorithm. We also plan to build an ad-hoc localization system using multiple mobile phones.

9 Acknowledgement

We would like to thank many colleagues in the Wireless and Networking research group in MSR Asia, for their strong support, encouragement, many fruitful discussions and valuable suggestions throughout the project. We also thank all the reviewers for their insightful comments and valuable suggestions, and Prof. Gaetano Boriello for shepherding the final revision of the paper.

10 References

[1] L. Girod, M. Lukac, V. Trifa, and D. Estrin, "The design and implementation of a self-calibrating distributed acoustic sensing platform," in *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2006, pp. 71–84.

[2] M. Youssef, A. Youssef, C. Rieger, U. Shankar, and A. Agrawala, "Pinpoint: An asynchronous time-based location determination system," in *MobiSys '06: Proceedings of the 4th international conference on Mobile systems, applications and services*. New York, NY, USA: ACM Press, 2006, pp. 165–176.

[3] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," in *MobiCom '99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 1999, pp. 59–68.

[4] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2000, pp. 32–43.

[5] K. Whitehouse and D. Culler, "Calibration as parameter estimation in sensor networks," in *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. New York, NY, USA: ACM Press, 2002, pp. 59–67.

[6] J. Sallai, G. Balogh, M. Maroti, A. Ledeczi, and B. Kusy, "Acoustic ranging in resource-constrained sensor networks," in *ICWN '04: Proceedings of the International Conference on Wireless Networks*. Las Vegas, Nevada, USA: CSREA Press, June 2004.

[7] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*. New York, NY, USA: ACM Press, 2003, pp. 81–95.

[8] R. Stoleru, T. He, J. A. Stankovic, and D. Luebke, "A high-accuracy, low-cost localization system for wireless sensor networks," in *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2005, pp. 13–26.

[9] M. Maroti, P. Volgyesi, S. Dora, B. Kusy, A. Nadas, A. Ledeczi, G. Balogh, and K. Molnar, "Radio interferometric geolocation," in *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2005, pp. 1–12.

[10] R. Stoleru, P. Vicaire, T. He, and J. A. Stankovic, "Stardust: a flexible architecture for passive localization in wireless sensor networks," in *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems*. New York, NY, USA: ACM Press, 2006, pp. 57–70.

[11] M. Kushwaha, K. Molnar, J. Sallai, P. Volgyesi, M. Maroti, and A. Ledeczi, "Sensor node localization using mobile acoustic beacons," in *MASS05: IEEE International Conference on Mobile Adhoc and Sensor Systems Conference*, Nov. 2005.

[12] K. D. Frampton, "Acoustic self-localization in a distributed sensor network," *IEEE Sensors Journal*, vol. 6, no. 1, pp. 166 – 172, 2006.

[13] R. J. Fontana, E. Richley, and J. Barney, "Commercialization of an ultra wideband precision asset location system," in *ICUWB '03: IEEE International Conference on Ultra Wideband Systems and Technologies*, Nov 2003, pp. 369–373.

[14] [Online]. Available: http://www.betanews.com/article/Bluetooth_21_EDR_Ratified_TouchtoConnect_On_the_Way/1175023711

[15] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn, "A relative positioning system for co-located mobile devices," in *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM Press, 2005, pp. 177–190.

[16] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss, "Requirements for photoware," in *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*. New York, NY, USA: ACM Press, 2002, pp. 166–175.

[17] S. Counts and E. Fellheimer, "Supporting social presence through lightweight photo sharing on and off the desktop," in *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM Press, 2004, pp. 599–606.

[18] G. Shen, Y. Li, and Y. Zhang, "Mobius: enable together-viewing video experience across two mobile devices," in *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*. New York, NY, USA: ACM Press, 2007, pp. 30–42.

[19] P. Enge and P. Misra, "Special issue on GPS: The Global Positioning System," in *Proceedings of the IEEE*, January 1999, pp. 3–172.

[20] J. Webr and C. Lanzl, "Designing a positioning systems for finding things and people indoors," *IEEE Spectr.*, vol. 35, no. 9, pp. 71–78, 1998.

[21] Wiki, "Speed of sound." [Online]. Available: http://en.wikipedia.org/wiki/Speed_of_sound

[22] J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley, Y. R. Yang, B. D. O. Anderson, and P. N. Belhumeur, "A theory of network localization," *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1663–1678, Dec. 2006.

[23] K. Whitehouse, C. Karlof, A. Woo, F. Jiang, and D. E. Culler, "The effects of ranging noise on multihop localization: an empirical study," in *IPSN05: The Fourth International Conference on Information Processing in Sensor Networks*, 2005, pp. 73–80.

[24] D. Niculescu and B. R. Badrinath, "Ad hoc positioning system (aps) using aoa," in *INFOCOM '03: the 22st Annual IEEE Conference on. Computer Communications*, San Francisco, CA, 2003.

[25] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM '00: the 19th Annual IEEE Conference on. Computer Communications*. Tel-Aviv,Israel: IEEE Infocom, March 2000.

[26] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, "Walrus: wireless acoustic location with room-level resolution using ultrasound," in *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM Press, 2005, pp. 191–203.

[27] BatSystem. [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/research/wiki/BatSystem

[28] V. Gerasimov and W. Bender, "Things that talk: Using sound for device-to-device and device-to-human communication," *IBM Systems Journal*, vol. 39, no. 3,4, pp. 530–546, 2000.

[29] M. Brandstein and H. Silverman, "A practical methodology for speech source localization with microphone arrays," *Computer, Speech, and Language*, vol. 11, no. 2, pp. 91–126, April 1997.

[30] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE Computer*, vol. 34, no. 8, pp. 57–66, 2001.

[31] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM Press, 2005, pp. 205–218.

[32] Y. Ji, S. Biaz, S. Pandey, and P. Agrawal, "Ariadne: a dynamic indoor signal map construction and localization system," in *MobiSys '06: Proceedings of the 4th international conference on Mobile systems, applications and services*. New York, NY, USA: ACM Press, 2006, pp. 151–164.

[33] L. Girod and D. Estrin, "Robust range estimation using acoustic and multimodal sensing," in *IROS '01: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 2001, pp. 1312–1320.