Swadloon: Direction Finding and Indoor Localization Using Acoustic Signal by Shaking Smartphones

Wenchao Huang, Yan Xiong, Xiang-Yang Li, *Fellow, IEEE*, Hao Lin,
Xufei Mao, Panlong Yang, Yunhao Liu, and Xingfu Wang

Abstract—We propose an accurate acoustic direction finding scheme, *Swadloon*, according to the arbitrary pattern of phone shaking in a rough horizontal plane. Swadloon leverages sensors of the smartphone without the requirement of any specialized devices. Our Swadloon design exploits a key observation: the relative displacement and velocity of the phone-shaking movement corresponds to the subtle phase and frequency shift of the Doppler effects experienced in the received acoustic signal by the phone. Swadloon tracks the displacement of smartphone relative to the acoustic direction with the resolution less than 1 millimeter. The direction is then obtained by combining the velocity from the displacement with the one from the inertial sensors. Major challenges in implementing Swadloon are to measure the displacement precisely and to estimate the shaking velocity accurately when the speed of phone-shaking is low and changes arbitrarily. We propose rigorous methods to address these challenges, and apply Swadloon to several case studies: Phone-to-Phone direction finding, indoor localization and tracking. Our extensive experiments show that the mean error of direction finding is around 2.1 degree within the range of 32 m. For indoor localization, the 90-percentile errors are under 0.92 m. For real-time tracking, the errors are within 0.4 m for walks of 51 m.

Index Terms—Direction finding, indoor localization, smartphone

✦

1 INTRODUCTION

D IRECTION finding is attractive in mobile social networks nowadays for supporting various applications, e.g., friending, and sharing. Recent mobile apps have made similar functions, such as Facebook's Friendshake [1] and Google Latitude [2]. However, they are based on GPS and cannot be applied to indoor environment. An accurate method of direction finding is by using antenna array [3], [4] in localization, but it requires specialized hardware and limits the availability to regular users. Several approaches of direction finding by smartphones have been proposed [5], [6], [7]. However, it remains a challenge for accurate direction finding by phone under long distance.

Precise indoor localization is also important for location based services. Those methods achieving high accuracy usually require special hardware not readily available on smartphones [8], or infrastructures expensive to deploy [9].

- W. Huang, Y. Xiong, and X. Wang are with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China. E-mail: {huangwc, yxiong, wangxfu}@ustc.edu.cn.
- X.-Y. Li is with the Department of Computer Science and Technology and TNLIST, Tsinghua University, and the Department of Computer Science, Illinois Institute of Technology, Chicago, IL 60616. E-mail: xli@cs.iit.edu.
- H. Lin is with the School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu, China. E-mail: imlinhao@gmail.com.
- X. Mao and Y. Liu are with the Department of Software Engineering, and TNLIST, Tsinghua University, Beijing, China.
  E-mail: {xufei.mao, yunhaoliu}@gmail.com.
- P. Yang is with the Institute of Communication Engineering, PLAUST, Nanjing, Jiangsu, China. E-mail: panlongyang@gmail.com.

Pure WiFi-based localization can achieve reasonable accuracy (e.g., $3 \sim 4\,\mathrm{m}$), but there always exist large errors (e.g., $6 \sim 8$ m) unacceptable for many scenarios [10]. Though there have been many proposals for improving the accuracy of WiFi based localization (e.g., with 80-percentile errors about 1 m [10]) by exploiting additional signals, low-cost precise indoor localization is still challenging.

We propose Swadloon, a **S**hake-and-**W**alk **A**coustic **D**irection-finding and indoor **LO**calizati**ON** scheme using smartphones. Suppose that there is an acoustic signal emitted from a speaker or a phone. Swadloon exploits the fact that shaking the smartphone or walking with the smartphone will cause Doppler effects on the acoustic signal received by the smartphone. Swadloon precisely measures the real-time *phase and frequency shift* of the Doppler effect, which corresponds to the *relative displacement and velocity* from the phone to the acoustic source respectively. Swadloon then obtains the accurate *direction* of the acoustic source by combining the relative velocity calculated from the Doppler shift with the one from the inertial sensors of the smartphone.

The main challenges of implementing Swadloon are noisy data collected from inertial sensors, and measurement of subtle frequency shift when the motion velocity of the phone is slow or fluctuates continuously. We propose several rigorous methods (discussed in detail in Section 4) in Swadloon to address these challenges, e.g., we use phase locked loop (PLL) to precisely measure the phase and frequency shift.

We evaluate the performance of Swadloon in the case study of phone-to-phone direction finding, where the object phone of direction finding serves as an acoustic source, and

the finder shakes his/her phone gently to produce the Doppler effect. We also explore the feasibility of applying Swadloon to real-time indoor localization, which uses a few anchoring nodes with known locations. The scheme does not rely on any fingerprints and is very easy to use: a user only needs to *shake* the phone for a short duration before *walking* and localization. These anchoring speakers will emit acoustic signals using non-audible frequency (typically around 20 kHz). The smartphones play the role of receivers. As it is difficult for a smartphone to find an accurate *North* as base for absolute direction, our localization method does not exploit the absolute direction. Instead we use a simple "triangulation" method by exploring the accurate opening angle from phone to two anchoring speakers. It measures the direction to the source and its relative displacement for achieving precise localization and real-time tracking respectively. Anchor nodes will not perform any computation or communication. Thus, dummy speakers such as loudspeakers can serve as anchors.

Our extensive experimental results show that Swadloon supports high accuracy for both Phone-to-Phone direction finding and real-time indoor localization. In our testing of Swadloon, the finder only needs to shake the phone *gently* and in *arbitrary* patterns in a rough horizontal plane. For the phone-to-phone direction finding, the mean error of the measured angle is 2.10 degrees within the range of 32 m, and the errors are under 2.06, 4.43, 5.81 degrees at 50, 90, 95 percent respectively, when the acoustic source faces towards to the phone. Since our acoustic direction finding achieves both long distance (about 32 m) and high accuracy (around 2 degrees), it supports a variety of potential applications, such as direction-based advertising that recommends new goods in a shopping mall, or sharing virtual business card with surroundings in a big party, where the former application requires advertisement being broadcasted as long as possible to be detected by users, and the latter one requires accurate direction finding. For indoor localization, we deploy one acoustic source per 6 meters, which broadcasts signals at a predefined frequency. For static localization, Swadloon achieves 90-percentile accuracy of 0.92 m, maximum error of 1.73 m, and the mean error of 0.5 m. For real-time indoor tracking, the error is always kept within 0.4 m even when users walk for more than 50 meters.

The rest of the paper is organized as follows. We review related work in Section 2 and present technical preliminaries in Section 3. We present the design of Swadloon in Section 4, We report our extensive experimental results in Section 5. We conclude the paper in Section 6.

2 RELATED WORK

2.1 Direction Finding

Specialized hardware. Former approaches requires special hardwares for achieving high accuracy, e.g., by using directional antenna [11], [12], [13] or antenna array [3], [4] to implement angle of arrival (AOA) [14] in localization. For example, by rotating the beam of its antenna, a receiver can pinpoint the direction of the AP as the direction that provides the highest received strength [11].

Non-specialized hardware. Zhang et al. [5] effectively emulates the functionality of a directional antenna by rotating the phone around the user's body, to locate outdoor APs. Qiu et al. [7] leverages four microphones for calculating 3D position of each other by using the distance ranging method [15]. As the work is intended for high-speed, locational, phone-to-phone (HLPP) games, it does not show the result when two phones are in long distances. Zhang et al. [16] calculates direction by head nodding or shaking using smart glasses. Other methods [17], [18] close to direction finding are to identify which target the user is pointing at when s/he moves mobile phone towards the target phone.

To the best of our knowledge, the approach closest to ours in direction finding is [6]. It estimates the direction and achieves the mean angular errors within 18 degree while ours is around 2 degree. This approach requires that the searching user generates Doppler Effects to all directions, e.g., the user stretches the arm while holding the searching device, and then swings it through 180 degrees. Correspondingly, as Swadloon tracks the displacement with the resolution under 1 mm, Swadloon only requires that the user shakes the phone gently in an arbitrary path.

2.2 Indoor Localization and Tracking

In indoor localization, to avoid the use of special-purpose infrastructure, e.g., [19], [20], [21], [22], [23], wireless localization, which only leverages an existing infrastructure instead of special-purpose hardware, has attracts many research efforts, e.g., [10], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. However, it is found [10] that pure wireless localization can achieve reasonable accuracy (e.g., $3 \sim 4$ m), but there always exist large errors (e.g., $6 \sim 8$ m) unacceptable for many scenarios. ByteLight [34] claims to be able to provide low-price infrastructure for localization using ceiling-embedded LEDs which send out Morse Code-like signals to be detected by the smartphone's camera. Our case study provides another choice for precise indoor localization, which only needs ceiling-embedded low-price speakers instead.

Leveraging acoustic wave by phone. The methods of leveraging the acoustic wave in smartphone applications have been well addressed. Most of them are leveraging the low speed of the acoustic wave for ranging, such as the mechanism of TOA [15] and TDOA [35], [36]. BeepBeep [15] detects the distance between two smartphones with high accuracy. It has been used by many other schemes, such as HLPP games [7], [37], device pairing [17] and indoor localization [10], [38]. Instead of precise ranging, [39], [40] use acoustic background spectrum for coarse-grained indoor localization.

In this work, we leverage Doppler effects of acoustic waves (i.e., measuring the precise relative displacement and velocity of phone) to design Swadloon. Swadloon is precise enough to be another basic tool of AOA, while it only requires off-the-shelf speakers. Furthermore, Swadloon supports arbitrary number of users and the phones of users do not need to send any signals to get the location, which avoids the signal interference when the number of users increases.

Leveraging the Doppler effects. Doppler effects have been leveraged in wide areas, such as radar, satellite communication, medical imaging and blood flow measurement, etc. There are also localization approaches leveraging the Doppler shift of wireless signals in localization [41] and tracking
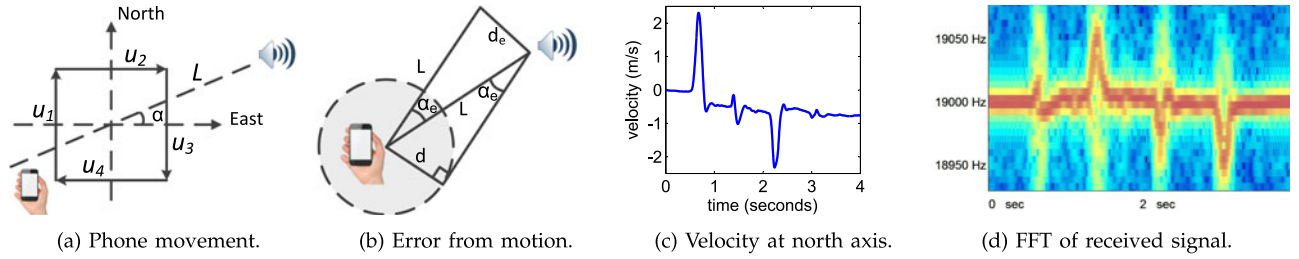
| (a) Phone movement. | (b) Error from motion. | (c) Velocity at north axis. | (d) FFT of received signal. |

Fig. 1. A simple case of calculating the direction α. (a) The phone starts moving north and draw a rectangle. (b) The velocity calculated from the inertial sensors. (c) FFT on the received acoustic signal.

[42] in wireless sensor networks. But it also needs special hardware not available for smartphone users. Meanwhile, by using the phase shift, Swadloon easily implements precise tracking without complicated algorithms compared with [42] which uses frequency shift.

Leveraging the inertial sensors. Inertial sensors have been used for pedestrian dead-reckoning [43] in indoor localization. The challenge is that it suffers from a large accumulation of errors. The complementary approaches to this problem are proposed in [25], [26], [44]. Swadloon uses the accelerometer and gyroscope to obtain the direction of the acoustic source.

3 PRELIMINARY APPROACHES

3.1 Mapping from Doppler Effects to Motion

Our scheme is based on the relationship between Doppler effects and the relative motion from the phone to the acoustic source, when the phone moves and causes Doppler effects on the received acoustic waves. Suppose the acoustic source is emitting the sinusoidal signal at the frequency of f_a, the observed frequency f_r [45] is $f_r = \frac{v_a + v}{v_a + v_s} f_a$. Here v is the velocity of the receiver; positive if the receiver is moving towards the source and negative in the opposite position. v_s is the velocity of the source and v_a is the traveling speed of the acoustic wave.

In this paper, we only consider the circumstance that the acoustic source is motionless or the velocity of the phone is far greater than the source, i.e., $v \gg v_s$. As typically $v_a \gg v_s$, we simplify the computing of the frequency shift f as follows:

$$f = f_r - f_a = \frac{v - v_s}{v_a + v_s} f_a \approx \frac{v}{v_a + v_s} f_a \approx \frac{f_a}{v_a} v. \quad (1)$$

We also assume the acoustic source sends the consecutive sinusoidal acoustic wave at constant frequency f_a. To derive the relative displacement from Doppler effect, we assume that the received signal has the form:

$$r(t) = A(t) \cos\left(2\pi f_a t + \phi(t)\right) + \sigma(t), \quad (2)$$

where $A(t)$ is the amplitude which changes continuously, $\phi(t)$ is the phase which is affected by the Doppler effect and $\sigma(t)$ is the noise. Assuming $\phi(t)$ is a continuous function, the observed frequency f_r at time t is $f_r(t) = \frac{1}{2\pi} \frac{d(2\pi f_a t + \phi(t))}{dt} = f_a + \frac{1}{2\pi} \frac{d\phi(t)}{dt}$. From Eq. (1), the frequency shift f at time t is

$$f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt}. \quad (3)$$

From Eqs. (1), (3), we get the velocity and displacement relative to the acoustic source:

$$\begin{cases} v(t) = \frac{v_a}{2\pi f_a} \frac{d\phi(t)}{dt}, \\ s(t) = \frac{v_a}{2\pi f_a} \phi(t) - \frac{v_a}{2\pi f_a} \phi(0), \end{cases} \quad (4)$$

where $s(t)$ is the relative displacement from the phone to the acoustic source. Specifically, $s(t) = L(0) - L(t)$, where $L(t)$ is the distance between the phone and the source at time t. In Section 4.3, we further show how to calculate $\phi(t)$ in order to obtain $v(t)$ and $s(t)$.

3.2 Basic Direction-Finding Using Doppler Effect for Simple Motion

We make a simple case of phone-to-phone direction finding to illustrate the intuition in designing Swadloon. Then we show the practical limitations of the simple case and implementation challenges. In latter sections, we propose our method on a more complicated case to address these issues.

Assume that the phone and the acoustic source are at the same height and the mobile phone starts moving in north and in a path of rectangle with the constant velocity u_1, u_2, u_3, u_4 in each direction, shown in Fig. 1a. So, frequency shifts are generated, where f_i corresponding to u_i. If the velocities and the frequency shifts are obtained, from Eq. (1), we can calculate the acoustic direction α in the following equations:

$$\begin{cases} u_1 \sin \alpha = \frac{v_a}{f_a} f_1; & u_2 \cos \alpha = \frac{v_a}{f_a} f_2; \\ -u_3 \sin \alpha = \frac{v_a}{f_a} f_3; & -u_4 \cos \alpha = \frac{v_a}{f_a} f_4. \end{cases} \quad (5)$$

Intuitively from Eq. (5), if $u_1 = u_2 = u_3 = u_4$, $f_2 > f_1 > 0 > f_3 > f_4$, which indicates that the $0 < \alpha < 45^o$. Formally, only two equations are needed to calculate α if the velocity in one equation is not parallel to the other. The additional equations can enhance the accuracy by using maximum likelihood estimation.

Note that α is changing while the phone is moving, so it will cause errors on obtaining α. However, it won't affect much on calculating the direction. In Fig. 1b, if the initial distance from the phone to acoustic source is L and the maximum moving range of the phone is d, the maximum angle error is $\alpha_e = \arcsin \frac{d}{L}$. As the phone moves gently, we assume that d is 10 cm at maximum. The maximum errors are 5.7, 1.15, 0.57, 0.19 degree at $L = 1, 5, 10, 30$ m respectively, i.e., the errors get smaller when the distance becomes longer. Note that the user can also shake phone with wider range that the angle error caused by phone movement is ignorable when $L \gg d$.

Moreover, if the phone calculates the position of acoustic source by not only the direction α according to Swadloon but also the distance L according to other techniques such as BeepBeep [15] while the measured L is accurate, the distance d_e from the calculated position to the actual position is $d_e = 2L \sin \frac{\alpha_e}{2} = 2L \sin \frac{\arcsin(d/L)}{2}$. When $d \ll L$, e.g., $10d \le L$, $\arcsin(d/L) \approx (d/L)$ and $\sin(d/2L) \approx d/2L$. So we simplify d_e as $d_e \approx d$. Then the maximum error on computed location caused by shaking is close to the shaking distance d, which is tolerable in direction finding.

However, there are several problems on applying this simple approach. First, the accurate velocity of the phone is hard to be obtained by using the inertial sensors. Though it can be calculated by the accelerometer and other sensors if given the initial velocity of the phone, the errors of the acceleration will be accumulated on its integration, i.e., the calculated velocity. For instance in Fig. 1a, the velocity is zero at the end of moving while the calculated one is $-0.77\,\mathrm{m/s}$ in Fig. 1c. Second, the mobile phone and the acoustic source may not be of the same height. In this case, the calculated f is lowered and the equations in Eq. (5) are not right. Third, it would be *hard and exhausting* to draw the *regular rectangle* for the phone users. Fourth, the velocity of the phone v *cannot be constant* in each direction. So we need a more general solution in cases of different heights and arbitrary motion patterns.

An important practical challenge is that spectrum analysis, such as fast Fourier transform (FFT), is not efficient in calculating frequency shift f, when the user shakes the phone *gently*, e.g., the peak value of the phone speed is $0.2\,\mathrm{m/s} \sim 0.4\,\mathrm{m/s}$ in our experiment in Fig. 7. Specifically, FFT cannot measure the precise value of f if v changes quickly due to the time-frequency resolution problem [46]. That is, for any signal, the time duration ΔT and the spectral bandwidth ΔF are related by $\Delta F \Delta T \ge 1$. For example, if the time resolution is $\Delta T = 8,192/44,100\,\mathrm{Hz} = 0.19\,\mathrm{s}$, the frequency resolution $\Delta F \ge 1/\Delta T = 5.38\,\mathrm{Hz}$. In this case, the frequency resolution is far not enough in our experiment where the peak value of frequency shift is around $11.8\,\mathrm{Hz} \sim 23.6\,\mathrm{Hz}$. Furthermore, it is still challenging when the phone speed is high, e.g., the maximum speed of a user's hand is $2\,\mathrm{m/s}$ [37]. In our simple case, the maximum speed also reaches 2 m/s as shown in Fig. 1c, where the maximum shift is 111.8 Hz theoretically, which seems sufficient for direction finding. However, in Fig. 1d we find that the maximum frequency shift only reaches about 70 Hz. The main reason is that the direction α is about 45 degree and the maximum shift is reduced to $111.8 \cos 45^o = 79\,\mathrm{Hz}$. Meanwhile, most of the time, the frequency shift is far less than 70 Hz. Note that, Spartacus [18] improves the resolution of FFT but also requires peak velocity at $2 \sim 6\,\mathrm{m/s}$ and achieves angular resolution with 10 degree; Swadloon lose the limitation that it also supports slow peak velocity ($0.2 \sim 0.4\,\mathrm{m/s}$) and higher angular resolution (mean error $\simeq 2$ degree). Hence, more accurate frequency measurement is preferred for calculating accurate direction.

Besides the challenge of calculating the frequency shift $f(t)$ for direction finding, the further problem is calculating the phase shift $\phi(t)$, from which $f(t)$ can be obtained by Eq. (3). We also show that the real-time indoor tracking can be implemented by using $\phi(t)$ in Section 5.2.2.
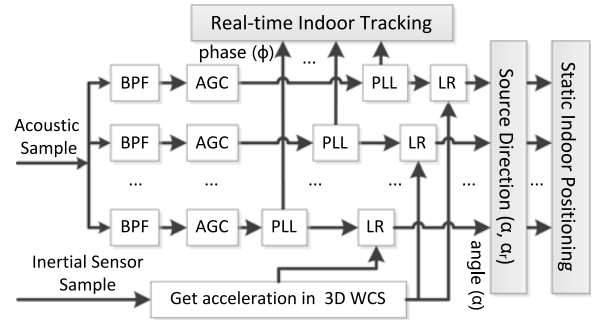

Fig. 2. Implementation of Swadloon.

4 DESIGN OF SWADLOON

We study the more complicated case that the user shakes the phone or walks in an arbitrary path. We show the design of Swadloon in Fig. 2. The phone gathers samples from the microphone and inertial sensors. The data are processed in real time to maximize the utilization of the CPU. The phone dynamically updates the direction of the source according to the previously calculated samples.

In Fig. 2, The noise $\sigma(t)$ and variational amplitude $A(t)$ in Eq. (2) is eliminated by BPF and AGC respectively. The *phase* ϕ and *frequency* f, which corresponds to the relative displacement and velocity respectively, are then obtained by PLL. Swadloon further combines the velocity from the acoustic and inertial sensor samples to get the *source direction* α by linear regression (LR). The phone returns the value of α and ϕ in real time for *direction finding*, *indoor localization* or *tracking*. We describe each component of the design as follows.

4.1 Band Pass Filter (BPF)

To get rid of the interference of other acoustic waves, we assume the phones of different users send acoustic waves in different frequency bands. Hence, in our implementation, the acoustic sample first walks through the band pass filter (BPF) such that only the waves at the exact frequency pass through BPF. Interference by other acoustic sources and low frequency noises that human can hear are both eliminated.

Note that the type of BPF should be seriously chosen. All frequency components of a signal are delayed when passed through BPF. As the frequency is changing in Doppler effect and we need to get the precise phase, the delay at each frequency components must be constant, such that the different frequency component will not suffer distortion, which is known as the linear phase property. As a result, we choose equiripple FIR filter, which satisfies the linear phase property.

Meanwhile, the bandwidth should be wide enough to get the total signal. Normally, the maximum speed of shaking the phone is less than 2 m/s. Thus, if the frequency of acoustic signal is $f_a = 19,000\,\mathrm{Hz}$, the maximum frequency shift $f_{max} = 111.8\,\mathrm{Hz}$. So, the minimum pass band of the filter is 223.6 Hz. For avoiding the interference by other acoustic sources, there should not be multiple signals that pass through the same BPF. Besides, acoustic bandwidth that almost all the smartphones support is limited to maximum of 22,050 Hz (i.e., sample rates of 44,100 Hz) and we find that the lowest frequency

that human can hardly hear is about 17,000 Hz in our experiment. Thus, the maximum number of acoustic sources that can sound simultaneously in a small area (with radius about $30\,\text{m}$) and be successfully detected is limited to $(22{,}050\text{-}17{,}000)/223.6 \approx 23$. However, this is not a challenge for Swadloon as we show that we only need a small number (less than 10) of acoustic sources in a small area for high accuracy. Though there are possible ways to allow more simultaneous acoustic waves such as dividing the signal into different time slots, like TDMA in shared medium network, it is beyond the scope of this paper.

4.2 Automatic Gain Control (AGC)

We adjust the filtered data by automatic gain control (AGC) such that the amplitude of the acoustic signal $A(t)$ in Eq. (2) is replaced by another one that is close to constant. The purpose is to successfully estimate the phase $\phi(t)$ by using PLL in Section 4.3. We adopt the design of AGC from [47]. Suppose T_s is the sampling period of the received signal and k is the step count of sampling, then $t = kT_s$. The main idea is for the input $r_b[k]$ from BPF, we estimate the amplitude $A[k]$ in Eq. (2) by updating $A_1[k]$ with the equation:

$$\log(A_1[k]) = (1 - A_\alpha)\log(A_1[k-1]) + A_\alpha \log(1/A_r[k-1]).$$

Here A_α represents the sensitivity for adjusting $A_1[k]$. $A_r[k]$ represents the coarse-grained estimation of $A[k]$.

Since $\overline{|r_b|} \simeq \frac{2}{\pi}A$ and the calculated $\overline{|r_b|}$ is stable when averaging consecutive 11 samples, in our implementation, $A_r[k] = \frac{\pi}{11*2}\sum_{i=k-10}^{k}|r_b[i]|$ and $A_\alpha = 0.9$. Then, for the received filter data $r_b[k]$, the output

$$r_c[k] = A_1[k]r_b[k].$$

For the amplitude of $r_c[k]$ is close to constant by AGC, if $A_1[k] = A_1[k-1]$, $A_1[k]A_r[k-1] = 1$. Thus, the amplitude of $r_c[k]$ is close to 1. Hence, we get $r_c(t) \approx \cos(2\pi f_a t + \phi(t))$, where $\sigma(t)$ and $A(t)$ in Eq. (2) is approximately eliminated by BPF and AGC respectively.

4.3 Tracking Subtle Displacement by Phase Locked Loop

According to Eq. (4), we use phase locked loops to calculate the phase $\phi(t)$, in order to get the precise relative displacement $s(t)$ and velocity $v(t)$ of the phone. PLL can be thought as a device that tracks the phase and frequency of a sinusoid [47]. In software implementation, we draw the idea from [48]. To get the precise $\phi(t)$, we update an adaptive estimation of $\phi(t)$ in real time, denoted as $\theta(t)$ in order that $\theta(t) \approx \phi(t)$. To make θ converge to ϕ after enough iterations, we define the corresponding function $J_{PLL}(\theta)$ such that J_{PLL} converges to its maximum at the same time. Specifically, $\theta(t)$ is updated in the iterations as:

$$\theta' = \theta + \frac{dJ_{PLL}}{d\theta}. \tag{6}$$

As a result, J_{PLL} should satisfy that

$$\max(J_{PLL}(\theta)) = J_{PLL}(\phi). \tag{7}$$

In Swadloon, we choose J_{PLL} as follows:

$$J_{PLL}(\theta) = \text{LPF}\{r_c(t)\cos(2\pi f_a t + \theta(t))\}$$
$$\approx \frac{1}{2}\text{LPF}\{\cos(\phi(t) - \theta(t))\}.$$

Here, LPF is the low pass filter which excludes the high frequency component in the above approximation. Hence, J_{PLL} satisfies Eq. (7).

Next, we need to change the continuous estimation process of Eq. (6) to the discrete one. Suppose T_s is the sampling period of the received signal and k is the step count of sampling, then $t = kT_s$. Assuming a small step size, the derivation in Eq. (6) with respect to θ at kT_s can be approximated[1]:

$$\frac{dJ_{PLL}}{d\theta} \approx \text{LPF}\left\{\frac{d[r_c[k]\cos(2\pi f_a k T_s + \theta))]}{d\theta}\right\}\Big|_{\theta=\theta[k]}$$
$$= -\text{LPF}\{r_c[k]\sin(2\pi f_a k T_s + \theta[k])\}.$$

As a result, the estimating of $\theta(t)$ is shown as follows:

$$\theta[k+1] = \theta[k] - \mu\text{LPF}\{r_c[k]\sin(2\pi f_a k T_s + \theta[k])\}, \tag{8}$$

where $\theta[k] = \theta(kT_s)$ and μ is a small positive value. Hence, $\phi[k] \approx \theta[k]$ after enough iterations. According to Eq. (4), if the max velocity of the phone is $v_{max} = 2\,\text{m/s}$, $f_s = 44{,}100$ Hz and $f_a = 19{,}000$ Hz, the max offset per sample $|\Delta\phi_{max}| = \frac{2\pi f_a}{v_a f_s}v_{max} = 0.016$. Besides,

$$r_c[k]\sin(2\pi f_a k T_s + \theta[k]) \approx \frac{1}{2}\sin(4\pi f_a k T_s + 2\theta[k]) \leq \frac{1}{2}.$$

Thus, $\mu > 0.03$ in Eq. (8), otherwise, the transition rate of $\theta[k]$ cannot catch up with the real phase. Furthermore, as $\frac{1}{2}\sin(4\pi f_a k T_s + 2\theta[k])$ cannot always be $1/2$, μ needs to be much more than 0.03 to let $\theta[k]$ converge to $\phi[k]$. However, when μ is bigger, the calculated phase is more sensitive to noises, and cannot be precise either. Hence, there is a trade-off on choosing the μ. Specifically, as moving speed of the phone is not always be $2\,\text{m/s}$, we can choose a smaller μ such that Swadloon is more robust to noises. In the implementation, we choose $\mu = 0.03$.

As the relative displacement is proportional to the phase shift by PLL, we estimate the precision of calculated displacement by Eq. (4). If the phase shift is 1 rad and the frequency of the source is 19,000 Hz, the relative displacement is $2.8\,\text{mm}$. We simply measure the phase when the phone is motionless, and find that the phase is oscillating around a constant central value, i.e., the real phase, and the amplitude of the oscillation is 0.005 rad when $\mu = 0.03$. We also let the phone move in an specific path towards the acoustic source with length of $30\,\text{cm}$, and measure the phase shift from the starting point to the end point. The standard deviation is 0.09 rad, which correpsonds the displacement of $0.25\,\text{mm}$. Hence, the measurement resolution of the corresponding displacement is less than $1\,\text{mm}$. In Section 5, We further evaluate Swadloon which depends on accuracy of PLL, to infer the robustness of PLL against multipath effects, noisy environment, etc.

---

1. The proof of the approximation is in G.13 of [48].

4.4   Getting Direction by Linear Regression (LR)

Assuming the direction vector of the acoustic source relative to the phone is $\overrightarrow{\lambda} = (\lambda_x, \lambda_y, \lambda_z)$ and velocity vector of the phone is $\overrightarrow{u} = (v_x, v_y, v_z)$, then $\overrightarrow{u} \cdot \overrightarrow{\lambda} = \frac{v_a}{f_a} f$ according to Eq. (1). For the obtained array $\overrightarrow{u}[k]$ and $f[k]$, they satisfy the following equations

$$\lambda_x v_x[k] + \lambda_y v_y[k] + \lambda_z v_z[k] = \frac{v_a}{f_a} \cdot f[k], \quad \forall k. \qquad (9)$$

Hence, the 3D direction $\overrightarrow{\lambda}$ can be obtained by solving these equations using linear regression, where $f[k]$ can be calculated by Eq. (3), Eq. (8). Ideally, if $u[k]$ is obtained from inertial sensors and there are no errors of $u[k]$, there are three unknowns $\lambda_x, \lambda_y, \lambda_z$ in the equation set. Moreover, using this we can calculate the direction when the phone moves in arbitrary paths, because different motion patterns of the phone merely cause different array $\overrightarrow{u}[k]$ and $f[k]$. We can also translate 3D direction $\overrightarrow{\lambda}$ to 2D direction α as follows:

$$\alpha = \begin{cases} \arcsin \frac{\lambda_y}{\sqrt{\lambda_x^2 + \lambda_y^2}} & \lambda_x \geq 0 \\ \pi + \arcsin \frac{\lambda_y}{\sqrt{\lambda_x^2 + \lambda_y^2}} & \lambda_x < 0. \end{cases} \qquad (10)$$

We now address the non-ideal circumstance with noisy sensor data, i.e., to minimize the error of velocity which is derived from the calculated acceleration in world coordinate system (WCS). In phone-to-phone direction finding and indoor localization, we only need the 2D direction α rather than the 3D direction $(\lambda_x, \lambda_y, \lambda_z)$. Thus, λ_z is not needed. From Eq. (9), if $\lambda_z v_z[k] \approx 0$, i.e., the phone moves in a horizontal plane or the two phones are at the same height approximately, we can calculate the direction by the following equation to eliminate the error of v_z:

$$\lambda_x v_x[k] + \lambda_y v_y[k] = \frac{v_a}{f_a} \cdot f[k]. \qquad (11)$$

Suppose $\hat{a}_x[i] = a_x[i] + \sigma_x[i]$ where $\hat{a}_x[i]$, $a_x[i]$, $\sigma_x[i]$ is the real acceleration, the calculated acceleration, the error of the calculation on the acceleration of the ith sample respectively. We can derive v_x from

$$v_x[k] = v_x[0] + \sum_{i=0}^{k-1} T[i]a_x[i] + \sum_{i=0}^{k-1} T[i]\sigma_x[i],$$

where $T[i]$ is time interval from $a_x[i]$ to $a_x[i+1]$.

The error σ_x is related the natural quality of the inertial sensors and challenging to be measured. In this paper, we simply assume σ_x equals to a constant e_x at a short period. Suppose $t[k] = \sum_{i=0}^{k-1} T[i]$, we get $\sum_{i=0}^{k-1} T[i]\sigma_x[i] = e_x t[k]$. Similarly, we also assume the error of a_y is a constant e_y at a short period.

We also consider the problem that there is clock drift between acoustic source and smartphone. For example, in our experiment the actual frequency of received signal is 19,000.13 Hz when the one of sent signal is 19,000 Hz. Denote the frequency shift caused by clock drift as f_d, the actual frequency of received signal is $f_r' = f_r + f_d$. Since the

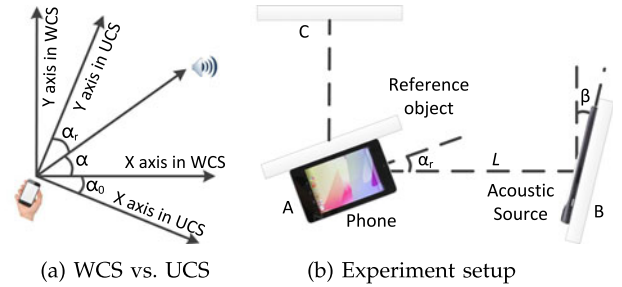
(a) WCS vs. UCS   (b) Experiment setup

Fig. 3. (a) WCS versus UCS when the phone is horizontal. (b) Experiment of direction finding.

frequency shift caused by Doppler effects $f \ll f_r$, it can be inferred that f_d is close to constant when f changes. Hence, the measured frequency shift $f' = f + f_d$ where f_d is close to constant.

As a result, from Eqs. (13), (10), (11), we could calculate the 2D direction by linear regression from the following equation set which has four unknowns ($\lambda_x, \lambda_y, \lambda_0, \lambda_1$)

$$\begin{pmatrix} w_x[0] & w_y[0] & 1 & t[0] \\ w_x[1] & w_y[1] & 1 & t[1] \\ \cdots & \cdots & \cdots & \cdots \\ w_x[n] & w_y[n] & 1 & t[n] \end{pmatrix} \begin{pmatrix} \lambda_x \\ \lambda_y \\ \lambda_0 \\ \lambda_1 \end{pmatrix} = \frac{v_a}{f_a} \cdot \begin{pmatrix} f'[0] \\ f'[1] \\ \cdots \\ f'[n] \end{pmatrix}. \qquad (12)$$

Here, $w_x[k] = \sum_{i=0}^{k-1} T[i]a_x[i]$, $w_y[k] = \sum_{i=0}^{k-1} T[i]a_y[i]$, $\lambda_0 = \lambda_x v_x[0] + \lambda_y v_y[0] + \frac{v_a}{f_a} f_d$ and $\lambda_1 = \lambda_x e_x + \lambda_y e_y$. Note that, we allow that $v_x[0] \neq 0$ and $v_y[0] \neq 0$ in our solution, which means we don't require the phone to be motionless before shaking the phone and calculating the direction. Similarly, we don't need the value of f_d. $v_x[0]$, $v_y[0]$ and f_d are put together as an unknown λ_0 in the equation.

4.5   Choosing the Direction in UCS for Evaluation

Vectors can be transformed between world's coordinate system and user's phone coordinate system (UCS) by the rotation matrix. As the compass is not accurate, we obtain the initial rotation matrix of the phone by sensor fusion of the compass, gyroscope, and accelerometer, but update the dynamic rotation matrix by merely using the gyroscope.

Hence in our world coordinate system, the z axis is considered to be accurate, but the x axis may not point to east due to the error of the compass. Hence, the calculated direction α in WCS may not be the actual direction relative to the east. To evaluate the performance of our direction finding, we will evaluate the direction (denoted as the ground truth and the measured value as α_r and α_r' respectively) of the acoustic source using the UCS of the phone that is placed horizontally such that its z axis is same as the z axis of WCS, as shown in Fig. 3a. When phone is static, the value α_r does not change. Thus, in Section 5.1, we measure α_r' to evaluate the accuracy of direction finding shown in Fig. 3b.

Hence, suppose the phone is horizontal, we get value α by using Swadloon and the opening angle from x axis in UCS to the one in WCS (α_0) by using the rotation matrix from UCS to WCS. α_r is calculated by

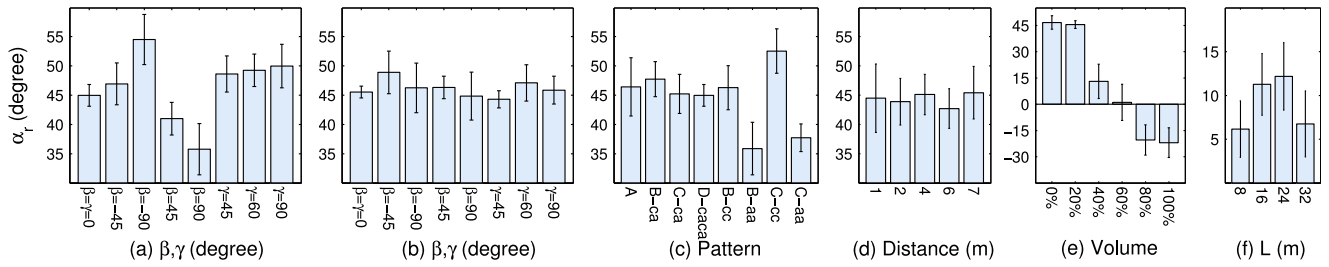$$\alpha_r = \pi/2 - \alpha - \alpha_0. \qquad (13)$$

Fig. 4. Mean and standard deviation of α_r (degree) affected by (a) β and γ when $L = 8\,\mathrm{m}$ (b) β and γ when $L = 32\,\mathrm{m}$ (c) motion pattern (d) non-line of sight (NLOS) (e) man-made multipath (f) multipath from the wall.

5 CASE STUDIES AND EVALUATIONS

5.1 Phone-to-Phone Direction Finding

We use Google Nexus 7 and Motorola XT 910 as smartphone and acoustic source respectively. We mainly use Android API to implement Swadloon and the acoustic source. Specifically, the audio playing and recording is implemented by using class AudioTrack and AudioRecord respectively; the inertial sensor samples are gathered through the function onSensorChanged(). The signal processing components, (i.e., BPF, AGC and PLL), are implemented by pure Java code without the requirement of any third-party libraries or hardware. On calculation direction, we adopt the implementation of linear regression in Michael Thomas Flanagan's java scientific library. The audio sample rate is 44,100 Hz, and sample rate of the inertial sensors is 200 Hz.

5.1.1 Experiment Design

The vertical view of the phone and acoustic source is shown in Fig. 3b. The distance between the phone and the acoustic source is L. The orientation angle of the phone and acoustic source at the horizontal plane is α_r and β respectively. There are reference objects at places A, B, C which are utilized to align the phones. The place C is used to put new acoustic source for further experiment. Additionally, we assume the elevation angle of the acoustic source is γ which is not shown in this 2D figure. The acoustic source is on the floor. The height of phone from the floor is about $40\,\mathrm{cm}$.

The main process of evaluating performance of direction finding is as follows: We vary L, α_r, β, γ by moving the reference objects. We obtain the measured direction α'_r by shaking the phone, aligning the phone to the reference object, and reading the direction value from the phone. We measure α'_r 50 times for each configuration.

5.1.2 Empty Room with Single Acoustic Wave

We first conduct the experiment in a large empty room for examining the accuracy of direction finding when there is only single acoustic wave. The sound pressure in the room is $-41\,\mathrm{dBFS}$ (about 30 dB SPL) measured by Nexus 7. The amplitude of the acoustic source at the distance of $1\,\mathrm{m}$ is $-20\,\mathrm{dBFS}$.

Effect by L and α_r. The case we mostly care about is the performance when the distance L and the orientation of the phone α_r is changing. Hence, we set $\beta = 0$ and $\gamma = 0$, and plot the standard deviations and cumulative distribution function (CDF) of the angular errors when L and α_r are changed in Fig. 5.

The key observation is that the measurement is very accurate when $L \leq 32\,\mathrm{m}$. We examine the reason in Fig. 6, which plots the calculated $\phi(t)$ on random samples with different L values. The calculated $\phi(t)$ is always smooth when $L \leq 24\,\mathrm{m}$, while there are small noises when $L = 32\,\mathrm{m}$ and much bigger noises if $L = 40\,\mathrm{m}$. Hence, the calculated related displacement and velocity become much less scrupulous when $L = 40\,\mathrm{m}$, which affects the calculation of direction. It is similar that most of the following cases mainly affect the calculated phase which finally affect the precision of direction finding.

In Fig. 5a, when $L \leq 32\,\mathrm{m}$, the mean error and standard deviation of the measurement is 2.10 and 2.66 degree. The angular errors are within 2.06, 4.43, 5.81 degree at 50, 90, 95 percent respectively. Though the errors become larger when $L = 40\,\mathrm{m}$, it is still tolerable. We also test angle errors when $L > 40\,\mathrm{m}$, but it becomes much unstable as the signal is too weak. So we do not show the result of this case.

In Fig. 5a, we also find when α_r is chosen from -90 to 90 degree, it has little effect on precision. As the errors are so close for different α_r, we don't show the CDF of different α_r.

Effect by β and γ. We test the errors when the orientation of the acoustic source is not directly pointing to the phone. In this case, we set $\alpha_r = 45^o$. In Figs. 4a and 4b, we show the mean and standard deviation with different choices of β, γ, L.

It shows an interesting result that when β changes, the mean value changes more when $L = 8\,\mathrm{m}$ in Fig. 4a than the one when $L = 32\,\mathrm{m}$ in Fig. 4b. The main reason is that the acoustic source we choose is not omnidirectional, and the signal is much stronger right in front of the source. The signal reflected from the wall affects the result, which is so-called the multipath effect. When the phone is further from the source, the signal reflected from the wall becomes much weaker than the one directly from the acoustic source.
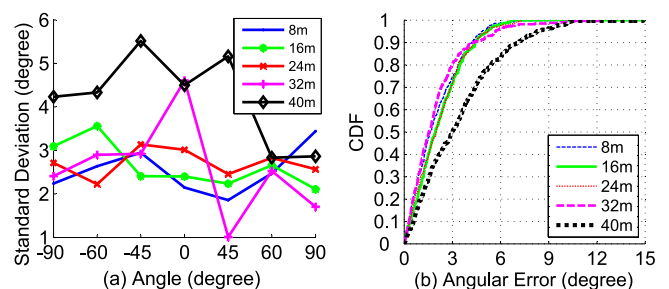

Fig. 5. The result of direction finding in an empty room when $\beta = 0$ and $\gamma = 0$.
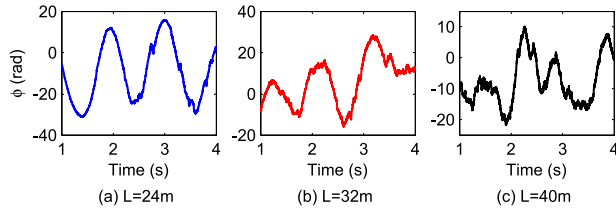
Fig. 6. The calculated phase $\phi(t)$.

Another observation is that if the acoustic source turns up, such as $\gamma = 45^o, 60^o, 90^o$, the mean value will not change a lot no matter $L = 8\,\text{m}$ in Fig. 4a or $L = 32\,\text{m}$ in Fig. 4b. That is, though there is multipath from the ceiling, it has little effect on the mean direction. We find a new phenomenon on multipath effect in latter experiment, which explains these observations here.

Motion pattern. We also analyze the angular errors caused by the inertial sensors. As we claim that Swadloon supports arbitrary pattern of phone movement, we test errors caused by different motion patterns of the phone. In this case, we set $L = 32\,\text{m}$, $\alpha_r = 45^o$, $\beta = \gamma = 0$. As we calculate the direction by Eq. (11) instead of Eq. (9) for better accuracy, it requires $\lambda_z v_z[k] \approx 0$. Note that in most cases of phone-to-phone direction finding, $\lambda_z \approx 0$. Hence, we do not strictly require $v_z[k] = 0$ that the experimenter shakes the phone in *rough* horizontal plane in the experiment.

The experimenter shakes the phone with *arbitrary* patterns in rough horizontal plane, e.g., pattern A in Fig. 8. More specifically, we do not constrain the speed or the amplitude of the phone-shaking movement. Even the subtle movement is tested in the experiment. As the PLL measures the relative displacement with high resolution, the result is acceptable shown in Fig. 4c: the standard deviation of the measurement is 4.96 degree.

We also show results of other regular patterns in Fig. 8. The pattern B, C, D is the circle, the rectangle, and mix of the circle and rectangle respectively. We also specify the

repeat times and the direction of motion pattern. For example, *D-caca* in Fig. 4c means that the phone is shaked for four times in pattern D: **c**lockwise, **a**nticlockwise, **c**lockwise, **a**nticlockwise. The rest of the patterns can be explained similarly. Note that in the experiment, the real shaking pattern is merely close to the specified one, instead of strict match of the two patterns.

The results in Fig. 4c shows that Swadloon is accurate for all motion patterns. We further analyze the detailed results of each pattern in Fig. 7. More specifically, we calculate unknowns $(\lambda_x, \lambda_y, \lambda_0, \lambda_1)$ by Eq. (12). Then, we substitute the solutions into Eq. (12) and compare the values of expression on the left side of Eq. (12) with the ones on the right side. The values represent real-time velocity relatively from a smartphone to an acoustic source. The key observation is that the compared values are very close in all cases. It means that our PLL is accurate to track relative velocity (e.g., values on the right side of Eq. (12)) and Swadloon is very robust in case of different motion patterns. Fig. 7 also shows that Swadloon is very accurate when the phone moves with very small amplitude (5 cm) and velocity ($0 \sim 0.4\,\text{m/s}$).

We also find that when the phone moves clockwise, there is a positive shift in the mean value. When the phone moves anti-clockwise, there is a negative shift. Since for the arbitrary pattern A, there exist both positive and negative shifts in the measurement, the standard deviation becomes a little bigger. We also observed that when the phone was shaken in other regular patterns compared to pattern A, the standard deviation becomes smaller. That is, the error shift is close to invariable in these cases. We also find that when we shake the phone in C-ca, D-caca, the means are close to the same. Based on the results, we choose D-caca as the default motion pattern in the entire experiment. We leave it as a future work to understand why the phenomena happen.

Non-line of sight. We set $L = 8\,\text{m}$, $\alpha_r = 45^o$, $\beta = \gamma = 0$, and test a simple case on the effect by non-line of sight. In Fig. 4d, a person stands between the phone and the acoustic
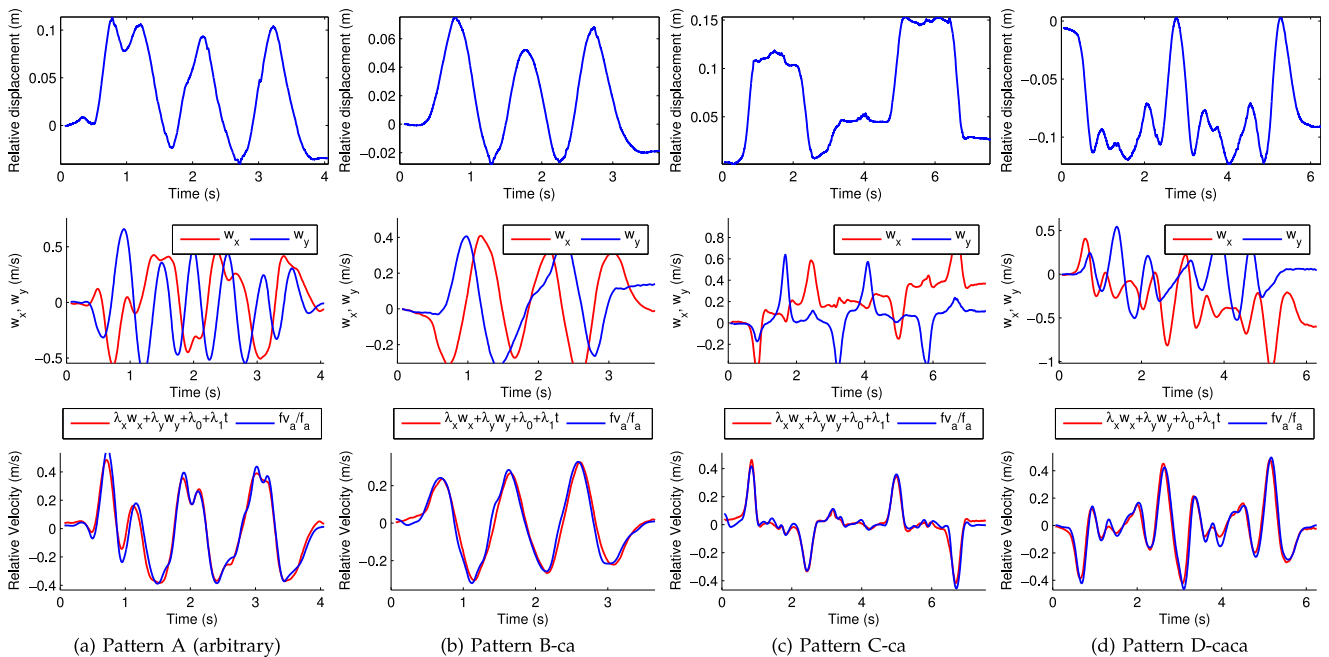

Fig. 7. Examples of direction finding which combines velocity from relative displacement $\frac{v_a}{2\pi f_a}\phi(t)$ with the one from inertial sensors ($w_x(t)$ and $w_y(t)$).
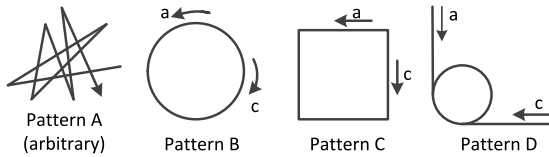
Fig. 8. Tested shaking patterns of the phone.

Fig. 9. (a) Errors on different cases when $L \leq 24\,\mathrm{m}$, $\alpha_r = \beta = \gamma = 0$. (b) The opening angle errors w.r.t. multiple signals.

source, and we measure the errors related to the distance from the person to the phone. It becomes apparent that when the person stands at either ends, the standard deviation is enlarged, while the person stands in the middle, it is close to the one without obstruction. Hence, the person has little effect on direction finding, as long as s/he is not too close to the acoustic source or the receiver. This is also verified in the experiment of noisy environments.

Another case of NLOS is that the user put his back to the source. The signal turns so weak and the result becomes unstable. In this case, the user can turn around to get the precise direction. The other possible complementary method is to let user rotate the phone around the user's body, similar to [5].

Multipath effect. As the multipath effect is hard to measure exactly, we first make a man-made multipath to find its impact. Then, we make a simple real case to verify our finding.

We set $L = 8\,\mathrm{m}$, $\alpha_r = 45^o$, $\beta = \gamma = 0$ and add another phone as acoustic source placed at position C in Fig. 3b. The new source is also 8 meters from the phone. It beeps at the same frequency with the source at B. The volume of the source at B is constant 60 percent. We change the volume of the source at C from 0 to 100 percent, and plot the Fig. 4e. When the volume is less than 20 percent, it has little effect: the standard deviation is low, and the mean value is slightly lowered. There is an interesting phenomenon that when the volume becomes larger, the angle becomes lower which is close to the direction of the new source. However, the standard deviation becomes bigger when both sources have high volume.

We then conduct an experiment with both acoustic source and phone near the wall. The wall is on the right hand side of the user while shaking the phone. We set $\alpha_r = \beta = \gamma = 0$ and $L = 8, 16, 24, 32\,\mathrm{m}$. The mean and stand deviation of α'_r is shown in Fig. 4f. α'_r becomes bigger for all the distances which can be inferred from the above conclusion. It can also be inferred that the strengths of the reflected signals relative to the respective direct signals are different at each L, which causes different mean shifts of α'_r. The additional observation is that the standard deviation is low for each distance. Hence, reflected signal is weak compared to the one directly from the acoustic source.

5.1.3　Empty Room with Multiple Acoustic Waves

To validate the robustness of Swadloon, we conduct two types of experiments: (1) an acoustic source broadcasts multiple signals at different frequencies, (2) multiple sources broadcast signals at different frequencies.

In experiment (1), we measure the angular errors when the acoustic source sends six sinusoidal signals at the frequency from 17,000 to 19,500 Hz. The experiment is performed by setting $\alpha_r = \beta = \gamma = 0$. We find that the results
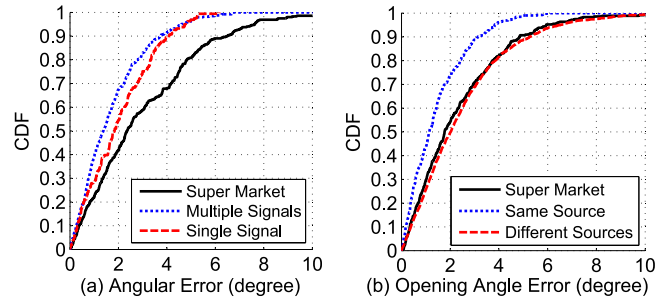
are similar for different L that $L \leq 24\,\mathrm{m}$, while the ones at $L = 32\,\mathrm{m}$ are a little worse. It is because that when the phone sends multiple signals, the signal strength of each component becomes weaker. We plot CDF at $L \leq 24\,\mathrm{m}$ in Fig. 9a. The performance is almost the same with the one sending a single wave. It can be deduced that we can use loudspeakers in the mall as anchor nodes while they are playing music.

We now analyze the performance of direction finding when there are multiple acoustic sources. The performance in this case will have direct impacts on the accuracy of the localization to be studied later in Section 5.2.1. Recall that as the computing of the absolute direction requires the accurate compass which is hard to get, in our localization method we use the opening angle $\angle A_i P A_j$ from the phone with location P to two arbitrary anchor nodes A_i and A_j instead of the absolute orientation of any vector PA_i or PA_j. Thus, here we measure the accuracy of estimated angle $\angle A_i P A_j$ by varying the locations of P, A_i, and A_j.

Fig. 9b shows the opening angle errors in three cases: (1) single source, multiple waves, super market, (2) single source, multiple waves, empty room, (3) multiple source, multiple waves, empty room. We find that the opening angle errors in cases (1), (2) are less than the direction errors in Fig. 9a. Furthermore, we observe that case (3) is much worse than (2). Though it is unfair to compare the two cases that the acoustic sources are different, it shows the possibility of improvement on the precision of indoor localization by using better acoustic sources, as we use the worse case for calculating the latter position.

5.1.4　Noisy Environment

We conduct this experiment in a super market, where it is noisy (-21 dBFS) and there are people walking around and blocking the line from the acoustic source to the phone. We also let the phone send multiple signals. In Fig. 9, the result becomes a little worse than the one in empty room. Almost all errors are less than 10 degrees, which are acceptable.

5.1.5　Overhead

As Swadloon calculates the direction in real time, we only evaluate the CPU usage. When Swadloon processes one acoustic signal, the CPU usage of the phone is 20.5 percent. When processing multiple signals, the pass band of BPF narrows down, which causes higher computation overhead per signal. There are multiple solutions for reducing the overhead, e.g., choosing IIR filter instead of the FIR filter, processing the signal in the network server, etc. Above all,

(a) Trilateration   (b) Acute Angle   (c) Obtuse Angle   (d) Good layout   (e) Bad layout
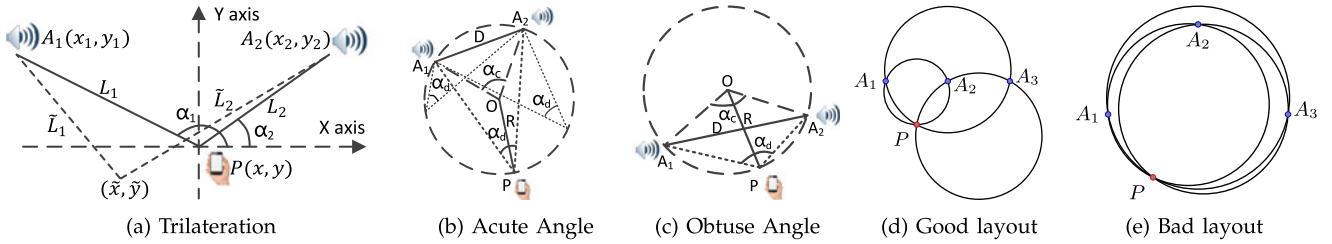
Fig. 10. Indoor localization and tracking: (a) trilateration, (b) pinpoint candidate location to a circle.

as we only need to shake the phone for a short duration to get the directions, the overhead is low that the total computation time is only within several seconds.

5.2 Real-Time Localization

We now describe our basic method of applying Swadloon to fine-grained indoor localization illustrated in Fig. 10a, which is based on the direction α and the phase ϕ in Section 4. Note that there are sophisticated methods leverage merely the Doppler frequency shift for localization, e.g., [42], we just provide a simple method as a case study to evaluate the accuracy of direction finding and phase shift measurement.

We require that there are at least three acoustic sources as anchor nodes installed, which send sinusoid signals at the specific different frequencies. Users need to get the position and frequency of each anchor node from network service. It includes two phases: finding the initial position and real-time tracking.

5.2.1 Static Position Localization

The user needs to shake the phone first in order to get his/her initial position. The phone calculates the direction of each anchor node in WCS and then gets the position. Note that as the compass is not precise, the calculated directions, such as α_1, α_2 in Fig. 10a, are not directly used in calculating the position. However, observe that the opening angle $(\alpha_1 - \alpha_2)$ is fixed no matter which WCS is chosen. We calculate the initial position using this opening angle. Taking the positions (x_1, y_1) and (x_2, y_2) of two anchor nodes A_1 and A_2 and the relative directions PA_1, PA_2 from phone (with unknown position P) to A_1 and A_2, we can compute the distance $D = \|A_1 - A_2\|$ and the opening angle $\alpha_d = \angle A_1 P A_2$, as illustrated in Fig. 10a. It can be inferred that the position P is on a fixed circle illustrated in Figs. 10b and 10c. If α_d is an acute angle as in Fig. 10b, $\alpha_c = 2\alpha_d$. So, the radius of the circle $R = \frac{D}{2\sin\alpha_d}$. Then we get at most two possible solutions of the position of the circumcenter O by using radius R and the given coordinates of two nodes A_1 and A_2. If α_d is an acute angle, then O and P are on the same side of A_1A_2. Similarly, if α_d is an obtuse angle, as in Fig. 10c, O and P are on the opposite side of A_1A_2.

For a system of n anchor nodes, there are $\frac{n(n-1)}{2}$ pairs of anchor nodes. As a result, phone P lies on $\frac{n(n-1)}{2}$ circles. Thus, with at least three anchor nodes, we can get the position of P. It is worth mentioning that for the circle formed by a node pair, the circle is divided into two arcs by the node pair. Node P only lies on one of the arcs, depending on whether α_d is an acute angle or an obtuse angle. Hence, for localization we search for the point P to minimize $\sum_i d_i$ where d_i is the distance from P to the ith arc.

We claim that it will result in better localization accuracy if we place the anchor nodes in a line as in Fig. 10d compared to the one in Fig. 10e. In Fig. 10e, the centers of the circles are too close, which causes big potential errors. The root reason is that the four points A_1, A_2, A_3, P are nearly at the same circle, which means the arbitrary point, e.g., A_1, is close to the circle which is constructed by the rest of three points, e.g., A_2, A_3, P.

Experimental setup. In Fig. 11, we place six phones as anchor nodes in the same empty room in the previous section. The positions are $(0, -3)$, $(6, 0)$, $(12, 0)$, $(18, 0)$, $(24, 0)$, $(30, -3)$ (meters) respectively. The beep frequencies are from 17,000 to 19,500 Hz. We choose spots at $y \in \{-3, -6\}$ and $x \in \{6, 9, 12, 15, 18, 21, 24\}$. We conduct the localization when people stay at these spots, and repeat the experiment 30 times for each spot.

Evaluation. The accuracy of static localization is shown in Fig. 12a. Swadloon achieves localization errors within 0.42, 0.92, 1.08, 1.73 m at the percentage of 50, 90, 95, and 100 percent respectively. The mean error and the standard deviation is 0.50 and 0.59 m respectively. We also find that the localization accuracy at spots with $y = -3$ m is better than the ones on $y = -6$ m. Specifically, on $y = -3$ m, the localization errors are within 0.28, 0.73, 0.91, 1.73 m at the percentage of 50, 90, 95, and 100 percent respectively.

Meanwhile, we find that there are nearly constant error shifts of the calculated position at all locations. Thus, we further adjust the position by linear regression. That is, we build a polynomial function model from the calculated positions to more precise positions by learning the results from half of the samples. We then apply the function to the other half and the result is plotted in Fig. 12b. It shows that the precision is greatly enhanced (i.e., the errors are within 0.67, 0.82, 1.56 m at the percentage of 90, 95, 100 percent respectively).

We then measure the errors of static localization in a large office (-34 dBFS), where the environment is much more complicated. The layout of the anchor nodes is nearly the same with the one in Fig. 11, except the anchor nodes are installed on the ceiling. Fig. 12b shows that the error is within 0.94, 1.23, 2.59 m at the percentage of 80, 90, 100 percent respectively after linear regression.

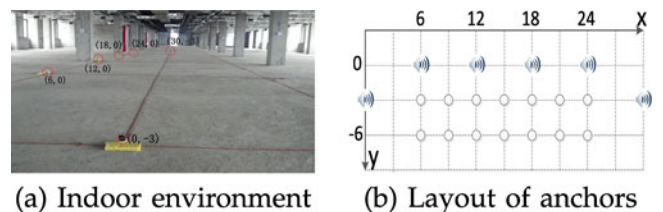(a) Indoor environment   (b) Layout of anchors
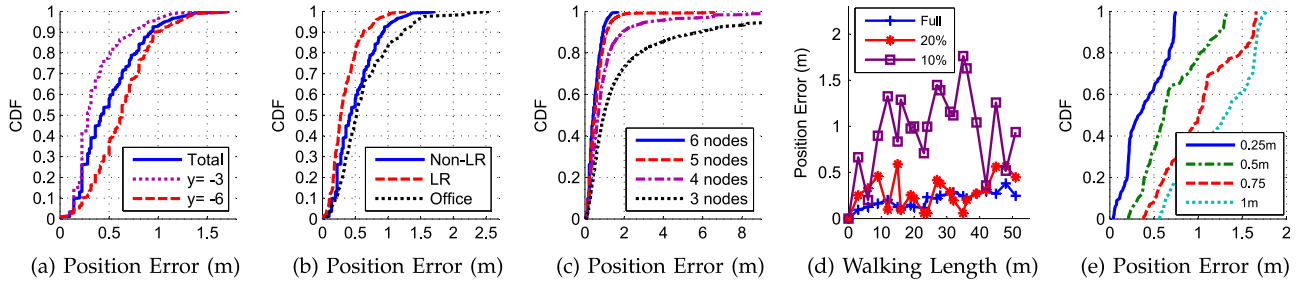
Fig. 11. Indoor localization testing prototype.

Fig. 12. Accuracy of static localization (a) in different locations, (b) in different scenes and by different methods. (c) when parts of the anchor nodes are chosen for calculation. Tracking accuracy (d) along the walking paths, (e) at final point $(24, -3)$ when there are initial position errors at $(6, -6)$.

We also choose specific number of nodes (i.e., $3 \sim 6$) from the six nodes to calculate the position. In Fig. 12c, it shows that the precision is greatly enhanced when the number of nodes increases. Besides, precision in case of three nodes becomes much worse for it is more sensitive by the layout shown in Figs. 12d and 12e.

5.2.2 Real-Time Tracking

In this phase, the user does not need to shake the phone again to obtain the position. The position is tracked in real time by tracking the relative displacement to each anchor using Eqs. (4), (8), and the update rate is 0.05 or 0.25 s in our experiment. In Fig. 10a, if the location of phone at time t has been calculated, denoted as (x, y), we calculate its location $(\tilde{x}, \tilde{y})$ at the latter time $\tilde{t}$ by getting $s(t)$ and $s(\tilde{t})$ using Eqs. (4) and (8). Then we calculate next location according to $(\tilde{x}, \tilde{y})$ iteratively. Specifically, if the user gets the location (x, y), then the distance from (x, y) to (x_i, y_i) is $L_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + h_i^2}$, where h_i is the relative height between the phone and the source (x_i, y_i). Thus, s/he gets the distances from all the available acoustic sources at time t. For $s(t) = L(0) - L(t)$ and $L(0)$ is constant initial distance, we can infer from Eq. (4) that

$$\tilde{L}_i = L_i - \frac{v_a}{2\pi f_a}(\tilde{\phi}_i - \phi_i), \qquad (14)$$

where $\tilde{L}_i = L_i(\tilde{t})$ and $\tilde{\phi}_i = \phi_i(\tilde{t})$. Since ϕ_i and $\tilde{\phi}_i$ can be calculated by Eq. (8) and L_i is already calculated, the latter $\tilde{L}_i$ can be derived from Eq. (14). Then we search for location $(\tilde{x}, \tilde{y})$ near (x, y) to minimize $\sum_i M_i$ where $M_i = |\tilde{L}_i - \sqrt{(\tilde{x} - x_i)^2 + (\tilde{y} - y_i)^2 + h_i^2}|$.

We conduct real time indoor tracking using the same environment as in Fig. 11. In our experiments reported here, a user starts from spot $(6, -6)$ shown in Fig. 13. Then, the user walks in some specific paths with length more than $50\,\mathrm{m}$ with the phone in his/her hand to the destination at spot $(24, -3)$. The errors are kept within $0.4\,\mathrm{m}$ shown in Fig. 13.
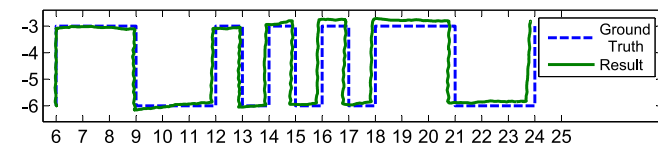

Fig. 13. Precise real-time indoor tracking.

We then consider the case that there are errors on the calculated initial position when the user starts walking. For each test, we uniformly choose a spot which is 0.25, 0.5, 0.75, or $1\,\mathrm{m}$ from $(6, -6)$, and measure the localization accuracies at the destination, i.e., distances from $(24, -6)$ to the calculated final positions in Fig. 12e. We can observe that the errors at the initial position do not affect the real time tracking, where the error is still within $2\,\mathrm{m}$ when the user walks in 51 meters and the initial position error is $1\,\mathrm{m}$.

To reduce the computation overhead, we let the phone process 20 percent of the samples, instead of full samples. Specifically, it processes consecutive samples of $0.05\,\mathrm{s}$ for each $0.25\,\mathrm{s}$. Hence, the phone can deal with the samples and track the position in real time. The result is close to the one which processes full samples in Fig. 13. We plot the localization errors in Fig. 12d. The mean error and standard deviation in this case is 0.29 and $0.34\,\mathrm{m}$ respectively, which is still very precise. CPU usage can also be lowered down by using 10 percent of the sample with the mean error of $1.02\,\mathrm{m}$, if the CPU of some other phone is not fast enough.

Note that the phone still constantly uses extra energy when the user is not shaking the phone. Actually, to determine when we can stop displacement tracking to save energy, we could use the inertial sensors to detect whether the user is static. We can further reduce the energy cost by designing more complicated algorithm, e.g., combining step counting (rough tracking) with our displacement measurement (precise tracking). Since we mainly concern the accuracy and energy cost of displacement tracking by PLL, we do not make further optimization and evaluations in this case.

6 CONCLUSION

In this paper, we propose Swadloon, a novel acoustic-based method to find the direction of the acoustic source. Swadloon effectively leverages the Doppler effects of the acoustic waves received by phones by exploiting the sensors in the smartphone and existing speakers to send sinusoidal signals. Our extensive evaluations show that Swadloon performs extremely well in phone-to-phone direction finding and real-time indoor localization. Note that in localization, we do not directly use the ranging result as accurate ranging often needs either time-synchronization or communication between two nodes, which require special hardware as anchor. Hence, some future work is to develop some mechanism achieving both ranging and direction finding simultaneously, and achieve the mechanism without the requirement of phone-shaking movement.

ACKNOWLEDGMENTS

REFERENCES

[1] (2014). Facebook's friendshake. [Online]. Available: http://www.facebook.com

[2] (2013). Google latitude. [Online]. Available: https://www.google.com.hk/latitude/

[3] K. Joshi, S. Hong, and S. Katti, "Pinpoint: Localizing interfering radios," in *Proc. 10th USENIX Conf. Netw. Syst. Des. Implementation*, 2013, pp. 241–254.

[4] Z. Li, X. Wang, Z. Du, and K. Gong, "Performance evaluation of a four-element antenna array with selection circuits for adaptive mimo systems," *Tsinghua Sci. Technol.*, vol. 15, pp. 294–298, 2010.

[5] Z. Zhang, X. Zhou, W. Zhang, Y. Zhang, G. Wang, B. Y. Zhao, and H. Zheng, "I am the antenna: Accurate outdoor ap location using smartphones," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 109–120.

[6] Y. Nishimura, N. Imai, and K. Yoshihara, "A proposal on direction estimation between devices using acoustic waves," in *Proc. 8th Int. ICST Conf. Mobile Ubiquitous Syst.: Comput., Netw. Serv.*, 2012, pp. 25–36.

[7] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2011, pp. 190–203.

[8] A. Prorok, P. Tome, and A. Martinoli, "Accommodation of nlos for ultra-wideband tdoa localization in single- and multi-robot systems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, 2011, pp. 1–9.

[9] H. Liu, H. Darabi, P. P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern., Part C Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[10] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of WiFi based localization for smartphones," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 305–316.

[11] A. Subramanian, P. Deshpande, J. Gaojgao, and S. Das, "Drive-by localization of roadside WiFi networks," in *Proc. IEEE 27th Conf. Comput. Commun.*, 2008, pp. 718–725.

[12] G. Zhang, Y. Xu, X. Wang, X. Tian, J. Liu, X. Gan, H. Yu, and L. Qian, "Multicast capacity for vanets with directional antenna and delay constraint," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 818–833, May 2012.

[13] J. Iguchi-Cartigny, P. M. Ruiz, D. Simplot-Ryl, I. Stojmenovic, and C. M. Yago, "Localized minimum-energy broadcasting for wireless multihop networks with directional antennas," *IEEE Trans. Comput.*, vol. 58, no. 1, pp. 120–131, Jan. 2009.

[14] D. Niculescu and B. R. Badrinath, "Ad hoc positioning system (aps) using aoa," in *Proc. IEEE 22nd Annu. Joint Conf. IEEE Comput. Commun.*, 2003, pp. 1734–1743.

[15] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: A high accuracy acoustic ranging system using cots mobile devices," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2007, pp. 1–14.

[16] L. Zhang, X. Li, W. Huang, K. Liu, S. Zong, X. Jian, P. Feng, T. Jung, and Y. Liu, "It starts with iGaze: Visual attention driven networking with smart glasses," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 91–102.

[17] C. Peng, G. Shen, Y. Zhang, and S. Lu, "Point&connect: Intention-based device pairing for mobile phone users," in *Proc. 7th Int. Conf. Mobile Syst., Appl. Serv.*, 2009, pp. 137–150.

[18] R. B. P. Z. Zheng Sun, Aveek Purohit, "Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl. Serv.*, 2013, pp. 263–276.

[19] K. Liu, X. Liu, L. Xie, and X. Li, "Towards accurate acoustic localization on a smartphone," in *Proc. IEEE INFOCOM*, 2013, pp. 495–499.

[20] L. Zhang, K. Liu, Y. Jiang, X. Li, Y. Liu, and P. Yang, "Montage: Combine frames with movement continuity for realtime multi-user tracking," in *Proc. IEEE INFOCOM*, 2014, pp. 799–807.

[21] L. Yang, Y. Chen, X. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 237–248.

[22] C. Bo, D. Ren, S. Tang, X. Li, X. Mao, Q. Huang, L. Mo, Z. Jiang, Y. Sun, and Y. Liu, "Locating sensors in the forest: A case study in greenorbs," in *Proc. IEEE INFOCOM*, 2012, pp. 1026–1034.

[23] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, "Epsilon: A visible light based positioning system," in *Proc. 11th USENIX Conf. Netw. Syst. Des. Implementation*, 2014, pp. 331–343.

[24] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 269–280.

[25] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 293–304.

[26] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proc. 10th Int. Conf. Mobile Syst., Appl. Serv.*, 2012, pp. 197–210.

[27] J. Zhao, W. Xi, Y. He, Y. Liu, X.-Y. Li, L. Mo, and Z. Yang, "Localization of wireless sensor networks in the wild: Pursuit of ranging quality," *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 311–323, Feb. 2013.

[28] W. Xi, Y. He, Y. Liu, J. Zhao, L. Mo, Z. Yang, J. Wang, and X.-Y. Li, "Locating sensors in the wild: Pursuit of ranging quality," in *Proc. 8th ACM Conf. Embedded Netw. Sensor Syst.*, 2010, pp. 295–308.

[29] Z. Yang, Y. Liu, and X.-Y. Li, "Beyond trilateration: On the localizability of wireless ad-hoc networks," in *Proc. IEEE INFOCOM*, 2009, pp. 2392–2400.

[30] X. Chen, X. Wu, X. Li, Y. He, and Y. Liu, "Privacy-preserving high-quality map generation with participatory sensing," in *Proc. 7th Eur. Conf. Wireless Sensor Netw.*, 2014, pp. 2310–2318.

[31] J. T. Biehl, M. Cooper, G. Filby, and S. G. Kratz, "Loco: A ready-to-deploy framework for efficient room localization using Wi-Fi," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 183–187.

[32] Y. Zheng, G. Shen, L. Li, C. Zhao, M. Li, and F. Zhao, "Travi-navi: Self-deployable indoor navigation system," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 471–482.

[33] L. Li, G. Shen, C. Zhao, T. Moscibroda, J. Lin, and F. Zhao, "Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 459–470.

[34] (2014). Bytelight technology. [Online]. Available: http://www.bytelight.com/

[35] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. P. Martin, "Detecting driver phone use leveraging car speakers," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw.*, 2011, pp. 97–108.

[36] K. Liu, X. Liu, and X. Li, "Guoguo: Enabling fine-grained indoor localization via smartphone," in *Proc. 11th Annu. Int. Conf. Mobile Syst., Appl. Serv.*, 2013, pp. 235–248.

[37] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, "Swordfight: Enabling a new class of phone-to-phone action games on commodity phones," in *Proc. 10th Annu. Int. Conf. Mobile Syst., Appl. Serv.*, 2012, pp. 1–14.

[38] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, "Centaur: Locating devices in an office environment," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 281–292.

[39] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proc. 9th Int. Conf. Mobile Syst., Appl. Serv.*, 2011, pp. 155–168.

[40] Y. T. H. M. Hiroyuki Satoh, Makoto Suzuki, "Poster abstract: Ambient sound-based proximity detection with smartphones," in *Proc. 11th ACM Conf. Embedded Netw. Sensor Syst.*, 2013, pp. 58:1–58:2.

[41] H. Jin Chang, J. ben Tian, T.-T. Lai, H.-H. Chu, and P. Huang, "Spinning beacons for precise indoor localization," in *Proc. 6th ACM Conf. Embedded Netw. Sensor Syst.*, 2008, pp. 127–140.
[42] B. Kusy, Á. Lédeczi, and X. D. Koutsoukos, "Tracking mobile nodes using RF doppler shifts," in *Proc. 5th Int. Conf. Embedded Netw. Sensor Syst.*, 2007, pp. 29–42.
[43] S. Beauregard and H. Haas, "Pedestrian dead reckoning: A basis for personal positioning," in *Proc. 3rd Workshop Positioning, Navigation Commun.*, 2006, pp. 27–36.
[44] C. Bo, X.-Y. Li, T. Jung, X. Mao, Y. Tao, and L. Yao, "Smartloc: Push the limit of the inertial sensor based metropolitan localization using smartphone," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 195–198, poster.
[45] J. Rosen, and L. Gothard, *Encyclopedia of Physical Science*. Facts on File, New York, NY 10001, 2009.
[46] J. Claerbout, *Earth Soundings Analysis: Processing Versus Inversion*. Blackwell Scientific Publications, Hoboken, New Jersey, 1992.
[47] M. Rice, *Digital Communications: A Discrete-Time Approach*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2008.
[48] C. R. Johnson and W. A. Sethares, *Telecommunication Breakdown; Concepts of Communication Transmitted via Software-Defined Radio*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2003.
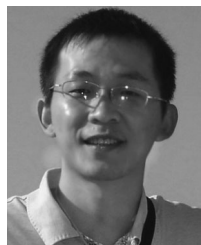
Wenchao Huang received the BS and PhD degrees in computer science from the University of Science and Technology of China in 2005 and 2011, respectively. He is currently a research associate in the School of Computer Science and Technology, University of Science and Technology of China. His current research interests include mobile computing, information security, trusted computing, and formal methods.

Yan Xiong received the BS, MS, and PhD degrees from the University of Science and Technology of China in 1983, 1986, and 1990, respectively. He is a professor in the School of Computer Science and Technology, University of Science and Technology of China. His main research interests include distributed processing, mobile computing, computer network, and information security.

Xiang-Yang Li received the bachelor degrees from the Department of Computer Science and the Department of Business Management from Tsinghua University, P.R. China, both in 1995, the MS degree in 2000, and the PhD degree in 2001 from the Department of Computer Science from the University of Illinois at Urbana-Champaign. He is a professor at the Illinois Institute of Technology and holds an EMC-Endowed visiting chair professorship at Tsinghua University. He is a recipient of China NSF Outstanding Overseas Young Researcher (B). His research interests include wireless networking, mobile computing, security and privacy, cyber physical systems, smart grid, social networking, and algorithms. He is an IEEE fellow and a ACM Distinguished scientist.

Hao Lin received the BS and MS degrees in computer science from Jiangnan University, in 2011 and 2014, respectively. He is currently a software engineer at China Pacific Insurance (Group) Co. Ltd. His current research interests include mobile computing, wireless network, and data mining.
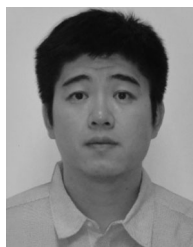
Xufei Mao received the bachelor's degree from the Shenyang University of Technology, the MS degree from Northeastern University both in computer science, in 1999 and 2003, respectively, and the PhD degree in computer science from the Illinois Institute of Technology, Chicago, in 2010. He is with the School of Software and TNLIST, Tsinghua University, Beijing, China. His research interests include wireless ad-hoc networks, wireless sensor networks, pervasive computing, mobile cloud computing, and game theory.
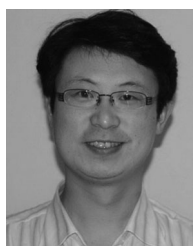
Panlong Yang received the BS, MS, and PhD degrees in communication and information system from the Nanjing Institute of Communication Engineering, China, in 1999, 2002, and 2005, respectively. He is now an associate professor at the PLA University of Science and Technology. He has published more than 50 papers in the areas of mobile ad hoc networks, wireless mesh networks, and wireless sensor networks.

Yunhao Liu received the BS degree in automation from Tsinghua University, Beijing, China, in 1995, and the MS and PhD degrees in computer science and engineering from Michigan State University, in 2003 and 2004, respectively. He is the director of Key Laboratory for Information System Security, Ministry of Education, and Professor in the School of Software, Tsinghua University. His research interests include pervasive computing, peer-to-peer computing, and sensor networks.

Xingfu Wang received the BS degree in electronic and information engineering from the Beijing Normal University of China in 1988, and the MS degree in computer science from the University of Science and Technology of China in 1997. He is an associate professor in the School of Computer Science and Technology, University of Science and Technology of China. His current research interests include information security, data management, and WSN.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.