Shake and Walk: Acoustic Direction Finding and Fine-grained Indoor Localization Using Smartphones

Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, Xufei Mao, Panlong Yang, Yunhao Liu

Abstract—We propose an accurate acoustic direction finding scheme, Swadloon, according to the arbitrary pattern of phone shaking in rough horizontal plane. 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^o within the range of $32m$. For indoor localization, the 90-percentile errors are under $0.92m$. For real-time tracking, the errors are within $0.4m$ for walks of $51m$.

I. INTRODUCTION

Direction 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]–[5] in localization, but it requires specialized hardware and limits the availability to regular users. Several approaches of direction finding by smartphones have been proposed [6]–[8]. However, it remains a challenge for accurate direction finding by phone under long distance.

We propose Swadloon, a *Shake-and-Walk* **A**coustic **D**irection-finding and indoor **LO**calizati**ON** scheme using smartphones. Assume 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 the noisy data collected from inertial sensors, and the measurement of the subtle frequency shift when the motion velocity of phone is slow or fluctuates continuously. We propose several rigorous methods (discussed in detail in Section IV) 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 feasiblity 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 20kHz). The smartphones play the role of receivers. It measures the direction to source and its relative displacement for achieving precise localization and real-time tracking respectively. Anchor nodes will not perform any computation or communication. Thus, it supports *arbitrary* number of users with extremely low cost.

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 rough horizontal plane. For the phone-to-phone direction finding, the mean error of the measured angle is 2.10^o within the range of $32m$, and the errors are under 2.06^o, 4.43^o, 5.81^o at 50%, 90%, 95% respectively, when the acoustic source faces towards to the phone. 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 $0.92m$, maximum error of $1.73m$, and the mean error of $0.5m$. For real-time indoor tracking, the error is always kept within $0.4m$ even when users walk for more than 50 meters.

The rest of the paper is organized as follows. We review the related work in Section II and present technical preliminaries in Section III. We present the design of Swadloon in Section IV, We report our extensive experiment results in Section V. We conclude the paper in Section VI.

II. RELATED WORK

Former approaches requires special hardwares for achieving high accuracy, *e.g.*, by using directional antenna [9] or antenna array [3]–[5] to implement Angle of Arrival (AOA)

in localization. For non-specialized hardware, [6] effectively emulates the functionality of a directional antenna by rotating the phone around the user's body, to locate outdoor APs. [8] leverages 4 microphones for calculating 3D position of each other by using the distance ranging method [10]. 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. Other methods [11], [12] 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 [7]. It estimates the direction and achieves the mean angular errors within 18° while ours is around 2°. 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 $1mm$, Swadloon only requires that the user shakes the phone gently in an arbitrary path.

In indoor localization, to avoid the use of special-purpose infrastructure, *e.g.*, [13]–[16], wireless localization, which only leverages an existing infrastructure instead of special-purpose hardware, has attracts many research efforts, *e.g.*, [17]–[22]. However, it is found [20] that pure wireless localization can achieve reasonable accuracy (*e.g.*, $3 \sim 4m$), but there always exist large errors (*e.g.*, $6 \sim 8m$) unacceptable for many scenarios. ByteLight [23] 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.

III. PRELIMINARY APPROACHES

A. 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 [24] 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:

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

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 Eq. (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 IV-C, we further show how to calculate $\phi(t)$ in order to obtain $v(t)$ and $s(t)$.

B. Basic Direction-Finding Using Doppler Effect for Simple Motion

We make a simple case of phone-to-phone direction finding to illustrate the intuition and challenges in designing Swadloon.

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 Figure 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 improve the accuracy by using maximum likelihood estimation.

Note that α is changing while the phone is moving, so it will cause errors on calculating α. However, it won't affect much on calculating the direction. In Figure 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 $10cm$ at maximum. The maximum errors are 5.7^o, 1.15^o, 0.57^o, 0.19^o at $L = 1, 5, 10, 30m$ respectively, *i.e.*, the errors get smaller when the distance becomes longer.

Moreover, if the phone calculates the position of acoustic source by not only the direction α according to Swadloon
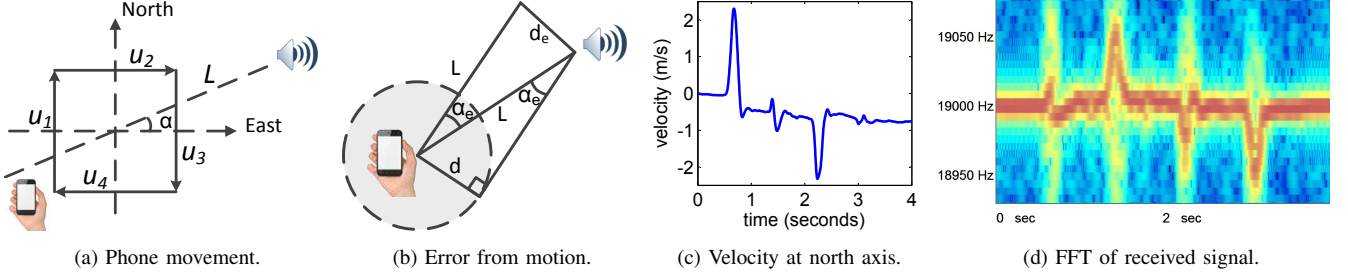
(a) Phone movement. (b) Error from motion. (c) Velocity at north axis. (d) FFT of received signal.

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

but also the distance L according to other techniques such as BeepBeep [10] 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 $d \ll L$, $e.g.$, $10d \leq L$, $\arcsin(d/L) \approx (d/L)$ 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 acceptable 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 Figure 1a, the velocity is zero at the end of moving while the calculated one is $-0.77m/s$ in Figure 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.

Normally, the velocity increases and then decreases, as shown in Figure 1c. The rapid changes of v bring the difficulties on calculating the frequency shift f. Specifically, spectrum analysis, such as Fast Fourier Transform (FFT), is efficient in calculating f, if v is large or close to constant for a while. But FFT cannot measure the precise value of f if v changes quickly due to the time-frequency resolution problem [25]. That is, for any signal, the time duration ΔT and the spectral bandwidth ΔF are related by $\Delta F \Delta T \geq 1$. For example, in Figure 1d, we try to apply FFT on the received signal, where the frequency of the acoustic wave is $f_a = 19000$Hz, the sample rate is 44100Hz, and FFT size is 8192. So, the time resolution is $\Delta T = 8192/44100$Hz$= 0.19s$. Then, the frequency resolution $\Delta F \geq 1/\Delta T = 5.38$Hz. However, we assume that the maximum speed of a user's hand is $2m/s$ [26]. The maximum frequency shift is $f_{max} = 2 * 19000/340 = 111.8$Hz. Even if the maximum speed is satisfied, the relative velocity may not reach $2m/s$. For instance, when the maximum speed of phone is about $2m/s$ shown in Figure 1c, for the phone never moves towards directly to the acoustic source, the

maximum frequency shift is about 60Hz in Figure 1d, which corresponds to the relative velocity $v = 1.1m/s$. Furthermore, in our circumstance, we only require that the user shakes the phone gently, so most of the time the frequency shift is far less than 111.8Hz. The resolution ΔF, which is more than 5.38Hz, is not precise enough to measure the frequency shift.

Hence, if the relative velocity and corresponding frequency shift are close to constant for a period, designers can increase ΔT to get better frequency resolution by FFT. However, in our circumstance, the velocity is always changing, which requires that both ΔT and ΔF is small enough, to get more precise f at smaller time block. Hence, it is in conflict with the time-frequency resolution problem of FFT for estimating f.

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 V-B2.

IV. DESIGN OF SWADLOON

We show the design of Swadloon in Figure 2. The phone gathers samples from the microphone and inertial sensors, when the user shakes the phone or walks in an arbitrary path. 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 previous calculated samples.

In Figure 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 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.

A. 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 specific frequency pass through BPF. The interference by other acoustic sources and the low frequency noises that human can hear are both eliminated.
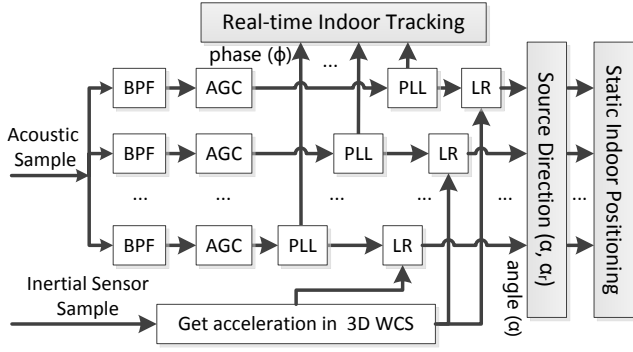
Fig. 2: Implementation of Swadloon.

Note that the type of BPF should be carefully 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.

B. 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 let the magnitude of $(\theta[k+1] - \theta[k])$ in Eq. (8) only be determined by μ, rather than $A(t)$, which is discussed in Section IV-C. We adopt the design of AGC from [27]. 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 eliminated by BPF and AGC respectively.

C. Tracking Subtle Displacement by Phase Locked Loop

According to Eq. (4), we use Phase Locked Loops (PLL) 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 [27]. In software implementation, we draw the idea from [28]. 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:

$$\theta' = \theta + \frac{dJ_{PLL}}{d\theta} \quad (6)$$

As a result, J_{PLL} should satisfy that

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

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

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

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]:

$$\begin{aligned} \frac{dJ_{PLL}}{d\theta} &\approx \text{LPF}\{\frac{d[r_c[k]\cos(2\pi f_a kT_s + \theta))]}{d\theta}\}\Big|_{\theta=\theta[k]} \\ &= -\text{LPF}\{r_c[k]\sin(2\pi f_a kT_s + \theta[k])\} \end{aligned}$$

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

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

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 $v_{max} = 2m/s$, $f_s = 44100$Hz and $f_a = 19000$Hz, the max offset per sample $|\Delta\phi_{max}| = \frac{2\pi f_a}{v_a f_s}v_{max} = 0.016$. Besides,

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

Thus, $\mu > 0.03$ in Eq. (8), otherwise, the transition rate of $\theta[k]$ cannot catch up with the real phase. Furthermore, as $\frac{1}{2}\sin(4\pi f_a kT_s + 2\theta[k])$ 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 tradeoff on choosing the μ. 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. (5). If the phase shift is 1 rad and the frequency of the source is 19000Hz, the relative displacement is $2.8mm$. 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 $30cm$, 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 $0.25mm$. Hence, the measurement resolution of the corresponding displacement is less than $1mm$. In section V, We further evaluate Swadloon which depends on accuracy of PLL, to infer the robustness of PLL against multipath effects, noisy environment, etc.

D. 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

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

[1]The proof of the approximation is in G.13 of [28].

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 3 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 causes different array $\overrightarrow{u}[k]$ and $f[k]$. We can also translate 3D direction $\overrightarrow{\lambda}$ to 2D direction α as follows:

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

We now address non-ideal circumstance with noisy sensor data, *i.e.*, to minimize the error of velocity which is derived from the calculated acceleration in 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:

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

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 the 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.

As a result, from Eq. (12)(10)(11), we could calculate the 2D direction by linear regression from the following equation set which has 4 unknowns $(\lambda_x, \lambda_y, \lambda_0, \lambda_1)$

$$\begin{pmatrix} w_x[0] & w_y[0] & 1 & t[0] \\ w_x[1] & w_y[1] & 1 & t[1] \\ \cdots & \cdots & \cdots & \cdots \\ w_x[n] & w_y[n] & 1 & t[n] \end{pmatrix} \begin{pmatrix} \lambda_x \\ \lambda_y \\ \lambda_0 \\ \lambda_1 \end{pmatrix} = \frac{v_a}{f_a} \cdot \begin{pmatrix} f[0] \\ f[1] \\ \cdots \\ f[n] \end{pmatrix}$$

where $w_x[k] = \sum_{i=0}^{k-1} T[i]a_x[i]$, $w_y[k] = \sum_{i=0}^{k-1} T[i]a_y[i]$, $\lambda_0 = \lambda_x v_x[0] + \lambda_y v_y[0]$ 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. $v_x[0]$ and $v_y[0]$ are put together as an unknown λ_0 in the equation.

E. Choosing the Direction in UCS for Evaluation

Vectors can be transformed between World's Coordinate System (WCS) 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 (WCS), the Z axis is considered to be accurate, but the X axis may not point to east due to the error of the compass. So, 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 α_r) 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 Figure 3a. When phone is static, the value α_r does not change. Thus, in Section V-A, we measure α_r to evaluate the accuracy of direction finding shown in Figure 3b.
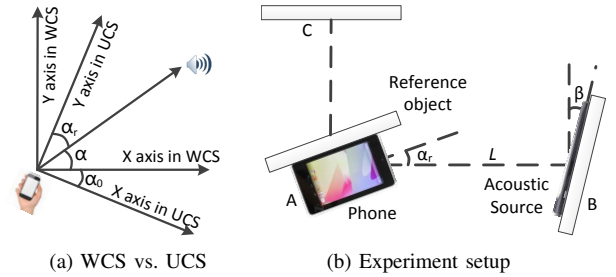

(a) WCS vs. UCS   (b) Experiment setup

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

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

$$\alpha_r = \pi/2 - \alpha - \alpha_0 \quad (12)$$

V. CASE STUDIES AND EVALUATIONS

A. Phone-to-phone Direction Finding

We implement Swadloon on Nexus 7, where all the components, including BPF and PLL, are implemented by using Android APIs. The audio sample rate is 44100Hz, and sample rate of the inertial sensors is 200Hz.

1) Experiment Design: The vertical view of the phone and acoustic source is shown in Figure 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 used to align the phones. The place C is used to put new acoustic source for further experiment. Additionally, we assume 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 $40cm$.

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.

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 of the room is -41 dBFS (about 30 dB SPL) measured by Nexus 7. The amplitude of the acoustic source at the distance of $1m$ is -20 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 Figure 4.
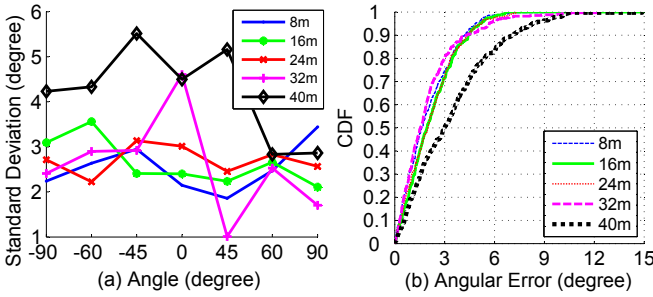

Fig. 4: The result of direction finding in an empty room when $\beta = 0$ and $\gamma = 0$.

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

We also find that α_r has little effect on precision according to Figure 4a. 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 Figure 5a, 5b, 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 in $L = 8m$ than the one in $L = 32m$. 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.

Another observation is that if the acoustic source turns up, such as $\gamma = 45^o$, 60°, 90°, the mean value will not change a lot no matter $L = 8m$ or $L = 32m$. 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. In this case, we set $L = 32m$, $\alpha_r = 45^o$, $\beta = \gamma = 0$. 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.
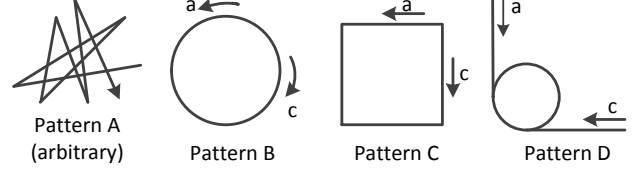

Fig. 6: Tested shaking patterns of the phone.

The experimenter shakes the phone with *arbitrary* patterns in rough horizontal plane, *e.g.*, pattern A in Figure 6. 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 Figure 5c: the standard deviation of the measurement is 4.96^o.

For further analyzing the cause of error and improving the accuracy, we observe the results of other patterns in Figure 6. 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 Figure 5c means that the phone is shaked for 4 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 important observation is that, when the phone moves clockwise, there is a positive shift on the mean value. When the phone moves anti-clockwise, there is a negative shift. As for the arbitrary pattern A, there are 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 constant in these cases. We also find that when we shake the phone in C-ca, D-caca, the means are close to same. Based on the results, we choose D-caca as the default motion pattern in the whole experiment. We leave it as a future work to understand why the phenomena happen.

Non-line of sight. We set $L = 8m$, $\alpha_r = 45^o$, $\beta = \gamma = 0$, and test a simple case on the effect by Non-line of sight (NLOS). In Figure 5d, a person stands between the phone and acoustic source, and we measure the errors related to the distance from the person to the phone. It becomes apparent that when the person stands in 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
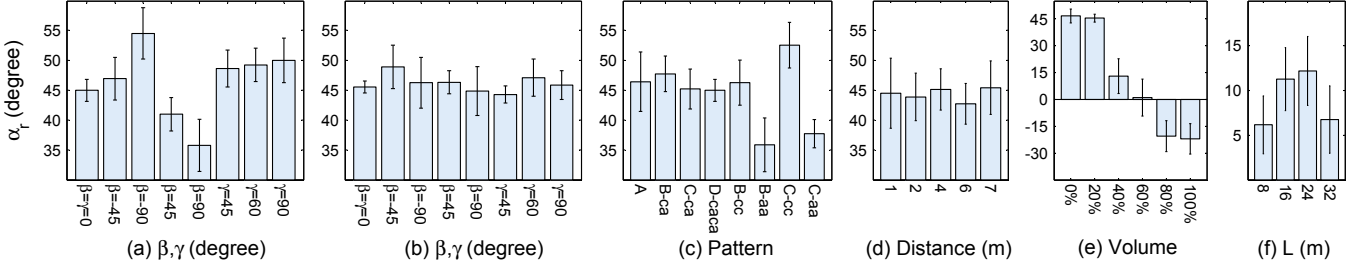
Fig. 5: Mean and standard deviation of α_r (degree) affected by (a) β and γ when $L = 8m$ (b) β and γ when $L = 32m$ (c) motion pattern (d) non-line of sight (e) man-made multipath (f) multipath from the wall.

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 environment.

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 complementory method is to let user rotate the phone around the user's body, similar to [6].

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 $L = 8m$, $\alpha_r = 45^o$, $\beta = \gamma = 0$ and add another phone as acoustic source placed at position C in Figure 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%. We change the volume of the source at C from 0% to 100%, and plot the Figure 5e. When the volume is less than 20%, 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 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 $L = 8, 16, 24, 32m$. The result is shown in Figure 5f. 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 other 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.

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 6 sinusoidal signals at the frequency from 17000Hz to 19500Hz. The experiment is performed by setting $\alpha_r = \beta = \gamma = 0$. We find that the results are similar for different L that $L \leq 24m$, while the ones at
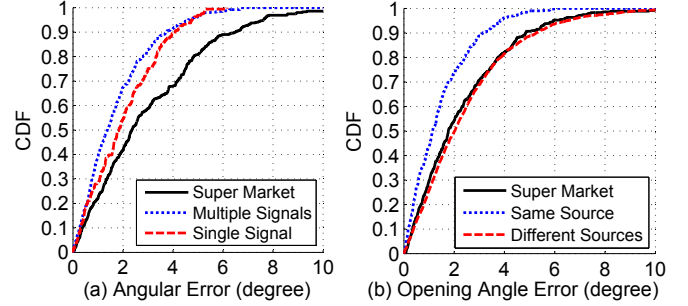

Fig. 7: (a) Errors on different cases when $L \leq 24m$, $\alpha_r = \beta = \gamma = 0$. (b) The opening angle errors when there are multiple signals.

$L = 32m$ are a little worse. It is because that when the phone sends multiple signals, the signal strength of each component becomes weaker. We plot the CDF at $L \leq 24m$ in Figure 7a. The performance is almost the same with the one sending single wave. It can be inferred 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 impact on the accuracy of the localization to be studied later in Subsection V-B1. 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.

Figure 7b 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 Figure 7a. 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.

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 Figure 7, the result becomes a little worse than the one in empty room. Almost all errors are less than 10 degrees, which is acceptable.

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

B. Real-time Localization

We now describe our basic method of applying Swadloon to fine-grained indoor localization illustrated in Figure 8a, which is based on the direction α and the phase ϕ in Section IV. Note that there are sophisticated methods leverage merely the Doppler frequency shift for localization, *e.g.*, [29], we just provide a simple method as a case study to evaluate the accuracy of direction finding and phase shift measurement.

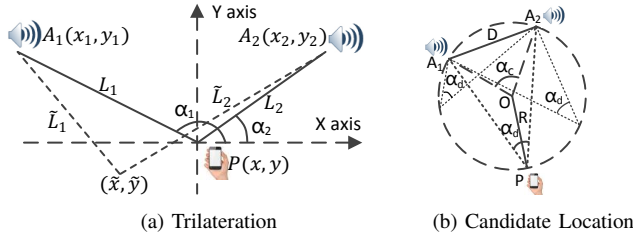(a) Trilateration   (b) Candidate Location

Fig. 8: Indoor localization and tracking: (a) trilateration, (b) pinpoint candidate location to a circle.

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.

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 Figure 8a, 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 Figure 8a. It can be inferred that the position P is on a fixed circle illustrated in Figure 8b. If α_d is a cute angle as in Figure 8b, $\alpha_c = 2\alpha_d$.

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 a cute angle, then O and P are on the same side of $A_1 A_2$. Similarly, if α_d is an obtuse angle, O and P are on the opposite side of $A_1 A_2$.

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 3 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.

Experimental setup: In Figure 9, we place 6 phones as anchor nodes in the same empty room in the previous subsection. The positions are $(0, -3)$, $(6, 0)$, $(12, 0)$, $(18, 0)$, $(24, 0)$, $(30, -3)$ (meters) respectively. The beep frequencies are from 17000 to 19500Hz. 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.
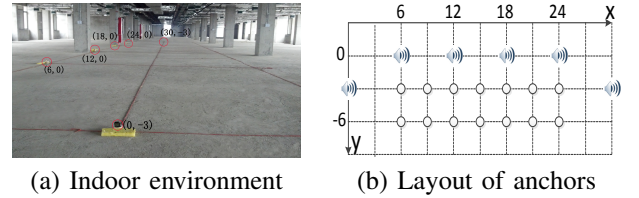

(a) Indoor environment   (b) Layout of anchors

Fig. 9: Indoor localization testing prototype.

Evaluation: The accuracy of static localization is shown in Figure 10a. Swadloon achieves localization errors within $0.42m$, $0.92m$, $1.08m$, $1.73m$ at the percentage of 50%, 90%, 95%, and 100% respectively. The mean error and the standard deviation is $0.50m$ and $0.59m$ respectively. We also find that the localization accuracy at spots with $y = -3m$ is better than the ones on $y = -6m$. Specifically, on $y = -3m$, the localization errors are within $0.28m$, $0.73m$, $0.91m$, $1.73m$ at the percentage of 50%, 90%, 95%, and 100% 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 ploted in Figure 10b. It shows that the precision is greatly enhanced (*i.e.*, the errors are within $0.67m$, $0.82m$, $1.56m$ at the percentage of 90%, 95%, 100% 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 Figure 9, except the anchor nodes are installed on the ceiling. Figure 10b shows that the error is within $0.94m$, $1.23m$, $2.59m$ at the percentage of 80%, 90%,
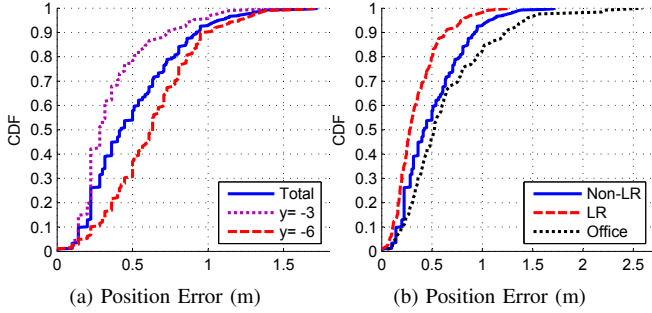
100% respectively after linear regression.

Fig. 10: Accuracy of static localization (a) in different locations, (b) in different scenes and by different methods.

2) Real-time Tracking: After getting the initial location of phone, the phone then updates the real-time location by calculating the relative displacement to each anchor node without shaking the phone again. In this case, the inertial sensors are not used. In Figure 8a, 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 Eq. (4), Eq. (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. According to Eq. (4), Eq. (8) and the definition of s_i, we have

$$\tilde{L}_i = L_i - \frac{v_a}{2\pi f_a}(\tilde{\phi}_i - \phi_i) \tag{13}$$

where $\tilde{L}_i = L_i(\tilde{t})$ and $\tilde{\phi}_i = \phi_i(\tilde{t})$. Then we search for location $(\tilde{x}, \tilde{y})$ near (x, y) to minimize $\sum_i M_i$ where $M_i = \big|\tilde{L}_i - \sqrt{(\tilde{x} - x_i)^2 + (\tilde{y} - y_i)^2 + h_i^2}\big|$.

We conduct real time indoor tracking using the same environment as in Figure 9. In our experiments reported here, users starts from spot $(6, -6)$ shown in Figure 11. Then, the user walks in some specific paths with length more than $50m$ with the phone in his/her hand to the destination at spot $(24, -3)$. The errors are kept within $0.4m$ shown in Figure 11.
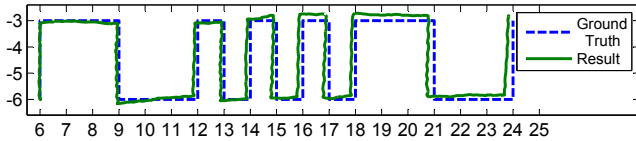

Fig. 11: Precise real-time indoor tracking.

VI. 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, both of which incur overhead. Hence, some future work are to develop a low overhead distance estimation between phone and source for further improving the performances and refine the localization for reducing the number of anchors.

References

[1] Facebook's friendshake. http://www.facebook.com.
[2] Google latitude. https://www.google.com.hk/latitude/.
[3] P. Kulakowski, J. Vales-Alonso, E. Egea-López, W. Ludwin, and J. García-Haro, "Angle-of-arrival localization based on antenna arrays for wireless sensor networks," *Computers & Electrical Engineering*, 2010.
[4] K. Joshi, S. Hong, and S. Katti, "Pinpoint: localizing interfering radios," in *NSDI*, 2013.
[5] J. Xiong and K. Jamieson, "Arraytrack: a fine-grained indoor location system," in *NSDI*, 2013.
[6] 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 *MobiCom*, 2011.
[7] Y. Nishimura, N. Imai, and K. Yoshihara, "A proposal on direction estimation between devices using acoustic waves," in *MobiQuitous*, 2012.
[8] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *SenSys*, 2011.
[9] A. Subramanian, P. Deshpande, J. Gaojgao, and S. Das, "Drive-by localization of roadside wifi networks," in *INFOCOM*, 2008.
[10] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *SenSys*, 2007.
[11] C. Peng, G. Shen, Y. Zhang, and S. Lu, "Point&connect: intention-based device pairing for mobile phone users," in *MobiSys*, 2009.
[12] R. B. P. Z. Zheng Sun, Aveek Purohit, "Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing," in *MobiSys*, 2013.
[13] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, 1992.
[14] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *Personal Communications, IEEE*, 1997.
[15] S. Se, D. G. Lowe, and J. J. Little, "Vision-based global localization and mapping for mobile robots," *IEEE Transactions on Robotics*, 2005.
[16] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *MobiCom*, 2000.
[17] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *MobiSys*, 2005.
[18] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM*, 2000.
[19] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *MobiCom*, 2012.
[20] 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 *MobiCom*, 2012.
[21] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in *MobiCom*, 2012.
[22] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: unsupervised indoor localization," in *MobiSys*, 2012.
[23] Bytelight technology. http://www.bytelight.com/.
[24] J. Rosen and L. Gothard, *Encyclopedia of Physical Science*. Facts on File, 2009.
[25] J. Claerbout, *Earth soundings analysis: Processing versus inversion*. Blackwell Scientific Publications, 1992.
[26] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, "Swordfight: enabling a new class of phone-to-phone action games on commodity phones," in *MobiSys*, 2012.
[27] M. Rice, *Digital Communications: A Discrete-Time Approach*. Prentice Hall, 2008.
[28] C. R. Johnson and W. A. Sethares, *Telecommunication Breakdown; Concepts of communication Transmitted via Software-Defined Radio*. Prentice Hall, August 2003.
[29] B. Kusy, Á. Lédeczi, and X. D. Koutsoukos, "Tracking mobile nodes using rf doppler shifts," in *SenSys*, 2007.