This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

Stride-in-the-loop Relative Positioning Between Users and Dummy Acoustic Speakers

Wenchao Huang, Xiang-Yang Li, *Fellow, IEEE*, Yan Xiong, Panlong Yang, Yiqing Hu, Xufei Mao, Fuyou Miao, Baohua Zhao, Jumin Zhao

***Abstract*—We propose and implement a novel positioning system, WalkieLokie, which directly calculates the relative position from a smart device to a target. The requirement of the target is simple: it is attached with a "dummy" acoustic speaker, which does not have any other rich capabilities, such as audio recording, communication, or computation. Hence, the proliferation of smart devices, together with the cheap accessory (*e.g.*, dummy speaker) embedded in daily used items (*e.g.*, smart clothes), paves the way for WalkieLokie applications. WalkieLokie leverages the walking motion for locating an acoustic speaker. The key insight is that the distance between the user and the speaker varies in real-time when the user walks, and the pattern of the variance implies the relative position. We design a novel algorithm to estimate the position and signal processing methods to support accurate positioning. The experiment results show that the mean errors of ranging and direction estimation are 0.63m and 2.46 degrees respectively. Extensive experiments conducted in noisy environments validate the robustness of WalkieLokie.**

***Index Terms*—Acoustic signaling, ranging, direction finding.**

I. Introduction

The proliferation of smarts devices has led to fast development of technologies and applications, such as indoor localization, Augmented Reality (AR). Especially, the AR technology is now popular among mobile users, *e.g.*, Pokémon GO [1], Microsoft HoloLens [2], Google Glass [3], Wikitude [4], and Augmented Car Finder [5]. The AR technology can enhance the user experience when the user visits museums or travels, where the user retrieves digital information of surrounding objects via cameras or other sensors (GPS, inertial sensors). Meanwhile, with the development of mobile technologies, new challenges are experienced and waiting to be resolved. One emerging problem is designing positioning schemes that can be applied in various kinds of applications.

Nevertheless, current positioning systems are still not ubiquitous enough due to different kinds of limitations [6], [7]. The GPS system is limited for outdoor use. Some indoor localization systems [8], [9] require special-purpose infrastructures or hardware, though they achieve sub-meter accuracy. Other systems that leverage existing infrastructures (*e.g.*, WiFi [10]–[15], Visible Light [16], [17]) are widely studied, but the accuracy still depends on the density of deployed infrastructures in a specific location. For the relative positioning systems, the accurate direction finding system Swadloon [18], [19] is proposed. Swadloon is light-weighted that the target is only required to be capable of broadcasting audio. However, Swadloon cannot calculate the distance of the target. Though the ranging system BeepBeep [20] can be added for calculating the relative position, BeepBeep requires that the target has rich functionalities, such as audio recording, computation, and wireless communications.

We propose and implement a novel and light-weighted system, WalkieLokie [21], to calculate the *relative position*, *i.e.*, the distance and direction from a smart device to a target. Compared with existing systems, WalkieLokie has less requirements on the target, which only needs to be attached with a *dummy* speaker (*i.e.*, an acoustic source) for broadcasting audio. Thus, WalkieLokie improves the ubiquity that it can be applied to more kinds of applications, including normal indoor localization, and new AR applications that require relative position of targets. For instance, a person walks in a shopping mall and a virtual shopping guide recommends the surrounding goods that are new arrivals or on sale [22]; a person shares her/his virtual business card [23], [24] with people walking around in a party; or gamers play Pokémon GO in indoor environments. Since the dummy speaker is cheap and simple that it does not require other rich features, it is available for the applications in many forms. For example, the speaker can be a cheap accessory in smart clothes (*e.g.*, Project Jacquard by Google [25]), or even a loudspeaker originally for sales promotion in a shopping mall. The broadcast audio is inaudible that the loudspeaker, which used to be a noisy tool for sales promotion, can now be "silent" for the same job by "broadcasting" its relative position. Note that the speaker can also perform "loud promotion" and the "silent positioning" at the same time.

The key insight of WalkieLokie is leveraging the mobility of a walking user for inferring the relative position. Specifically, the distance between the target and the user changes in real time when the user is walking, and the pattern of *displacement* (variance of distance) is associated with the relative position. Hence, by precisely tracking the displacement according to the acoustic signal received from the target, we can simultaneously calculate the distance and direction from the user to the target.

A major challenge of realizing WalkieLokie is to precisely track the displacement, especially when the acoustic signal is weak. We implement several software-based components of signal processing, including Band Pass Filter (BPF), Auto-

Wenchao Huang, Xiang-Yang Li, Yan Xiong, Panlong Yang, Yiqing Hu, Fuyou Miao are with School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. Xufei Mao is with Dongguan University of Technology, Beijing, China. Baohua Zhao is with Global Energy Interconnection Research Institute, Beijing, China. Jumin Zhao is with Taiyuan University of Technology, Taiyuan, China. Corresponding author: Yan Xiong. Email: {huangwc, yxiong, huyiqing, mfy}@ustc.edu.cn, {xiangyang.li, panlongyang, xufei.mao}@gmail.com, zhaobaohua@geiri.sgcc.com.cn, zhaojumin@126.com.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

matic Gain Control (AGC), and Phase-Locked Loop (PLL). Particularly, we design the second-order PLL for tracking the phase, such that the tracked displacement, which is proportional to the tracked phase shift, is accurate when the signal is weak. Furthermore, our PLL is still accurate, when the signal is modulated with more pulses, as explained in the next paragraph.

We also face the challenge of ensuring the robustness of WalkieLokie, which stems from many practical issues. The main issue is that the accuracy of our basic positioning scheme is reduced when the user is far from the device (*i.e.*, $8 \sim 20m$). We address the issue by modulating more pulses in the acoustic signal emitted by the target, and design the pulse detection algorithm and the long-distance ranging mechanism. We also deal with other practical issues. WalkieLokie can still perform relative positioning, when the user turns the walking direction. It is also robust against interferences, such as multipath effects, concurrent speakers, and noisy environments.

We implement WalkieLokie and evaluate the performance in an empty room, an office and a typical shopping mall. For the case when a user is in the vicinity of a speaker, the mean errors of ranging and direction estimation are $0.63m$ and 2.45^o respectively. In the shopping mall, the mean error is $1.28m$ for relative positioning, where the user walks arbitrarily in a $600m^2$ area and speakers in 5 different places are located.

The contributions of the paper are as follows:

- We propose and implement a novel relative positioning system, which has little requirement on the targeted device.
- We design a novel algorithm for relative positioning by leveraging only the received acoustic signal without any additional information from the targets.
- We propose a group of acoustic processing methods to ensure robustness and ubiquity of our system in practical environments.

The rest of the paper is organized as follows. We present the overview of WalkieLokie in Section II. Then, we propose the position estimation in Section III and the acoustic processing in Section IV. We report our extensive experimental results and review some related work in Section V. Finally, we conclude the paper.

II. Overview

A. Problem Definition

System Requirements: The dummy speaker merely broadcasts inaudible audio without any other capabilities. The smart device has a microphone and inertial sensors (*i.e.*, compass, accelerometer, gyroscope), which are common components in almost all smart devices. The smart device is held by a walking user in open air or attached to the user's body, but it cannot be put in the pocket; otherwise, the received acoustic signal is too weak to be processed. When performing relative positioning, the smart device needs to be roughly relatively static to the user's body (*e.g.*, the user holds the device firmly without shaking the device) such that the user's steps can be precisely counted. Note that if we also leverage the context-sensing techniques [26] for detecting whether some assumptions are violated, (*e.g.*, device in the pocket), we can design some strategies (*e.g.*, just ignoring the signal) for improving the robustness. However, it is out of the paper's scope.

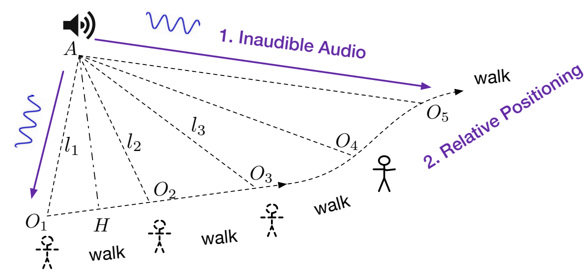


Fig. 1. Example of relative positioning.

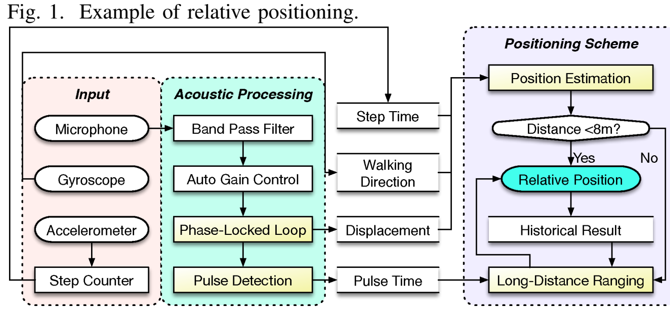


Fig. 2. The architecture of WalkieLokie.

Output: For example, in Figure 1, a user walks and steps on O_1, O_2, ..., and a speaker is placed at A. Assume that $\overline{AH} \perp \overline{O_1O_2}$. WalkieLokie calculates $|\overline{O_1H}|$ and $|\overline{AH}|$, which indicates the relative position between O_1 and A. The relative position can also decompose into distance (*e.g.*, $|\overline{O_1A}|$) and direction (*e.g.*, $= \angle AO_1H$).

B. Intuitive Solution and Observations

The key insight of relative positioning is that the pattern of real-time displacements from the user to the speaker is related to the relative position directly. In Figure 1, we denote l_i as $|\overline{AO_i}|$. Suppose the displacements $d_1(=l_1-l_2)$ and $d_2(=l_2-l_3)$ are measured beforehand and the user's stride length ($|\overline{O_1O_2}|$) is given. Intuitively, $d_1 \approx 0$ implies that O_1 and O_2 are close to H; $d_2 < 0$ implies that the speaker is at the back of the walking user. Hence the coarse-grained direction is inferred. Another observation is that when the distance $|\overline{AH}|$ increases, the value of $|d_2 - d_1|$ decreases, which implies the coarse-grained distance as well. Hence, the relative position between O_1 and A can be estimated, if d_1 and d_2 can be calculated. Note that the distance l_i cannot be directly measured, and we measure the displacement d_i to infer the relative position instead.

Long-distance ranging: Observe that the ranging in the intuitive solution is more inaccurate when the distance increases. Recall that if $|\overline{AH}|$ increases, $|d_2 - d_1|$ decreases. Since there are errors on calculating d_i, and $|d_2-d_1|$ is used for inferring the distance, the ranging is error-prone when $|d_2-d_1|$ is small. Hence, the long-distance ranging is required to be designed. Note that the accuracy of direction finding is not much affected, for the accuracy is determined by the precision of measured d_1.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

C. Architecture

In Figure 2, WalkieLokie is composed of 3 main components: the input of smart device, the acoustic processing, and the positioning scheme.

Input: WalkieLokie gathers continuous data from the microphone and the inertial sensors in the smart device. The microphone records audio for acoustic processing. The accelerometer mainly serves as a pedometer, which records the time of each user's step. The gyroscope is used to calculate the angle of user's rotation, when the user turns direction.

Acoustic signal processing: The component generates intermediate results for the positioning scheme. One result is real-time **relative displacement**, which is calculated by using the Band Pass Filter (BPF), Automatic Gain Control (AGC), and judiciously-designed Phase Locked Loop (PLL), as illustrated in Section IV. Another intermediate result provides additional information for long-distance ranging. More specifically, we encode periodical pulses in the acoustic signals, and the smart device detects the **receiving time** of the periodical pulses.

Positioning scheme: The scheme calculates the relative positions by leveraging the intermediate results. The basic scheme (*i.e.*, position estimation in Section III) estimates position by using the relative displacement, the time when the user steps on the ground, and the user's turning angle. If the computed distance is short ($< 8m$), the position calculated by the basic scheme is accepted as the valid result. Otherwise, the scheme leverages historical results of relative positioning together with the receiving time of periodical pulses to infer the distance, which is the long-distance ranging. Observe that according to the scheme, it is acceptable that there are Non-Line-of-Sight (NLoS) effects occasionally, where the position can also be estimated according to the historical results and current signals.

III. Positioning Scheme

In this section, we design the two main sub-modules of the positioning scheme: the basic positioning estimation according to our insight, and the extended long-distance ranging. We also introduce the method of inertial-sensor processing which generates input of the positioning scheme.

A. Basic Position Estimation

Simple case: We first consider a simple case when a user walks in a straight line, as shown in Figure 3. The case contains the following input:

- h: height difference between the speaker and the smart device. Here, $h = |\overline{AG}|$.
- s: stride length, *i.e.*, $s = \overline{O_iO_{i+1}}$. Here, s is assumed to be close to constant.
- d_i: the displacement of each user's step, *i.e.*, $d_i = l_i - l_{i+1}$ for the step $\overline{O_iO_{i+1}}$.

Then, we calculate the following output, which can be used for directly obtaining the horizontal distance $L_i = |\overline{GO_i}|$ and direction $\psi'_i = \angle GO_iO_n$:

- y: distance from the speaker placed at A to $\overline{O_iO_{i+1}}$. Here, $y = |\overline{AH}|$, where $\overline{AH} \perp \overline{O_1O_n}$.

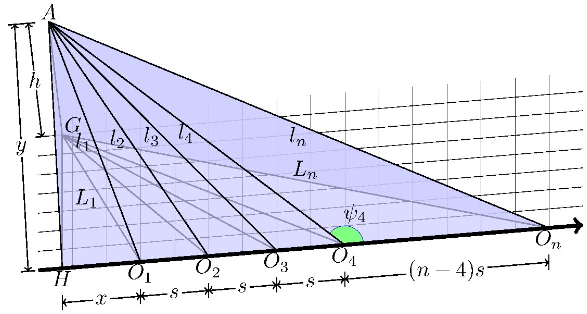


Fig. 3. Positioning when the user walks along a line.

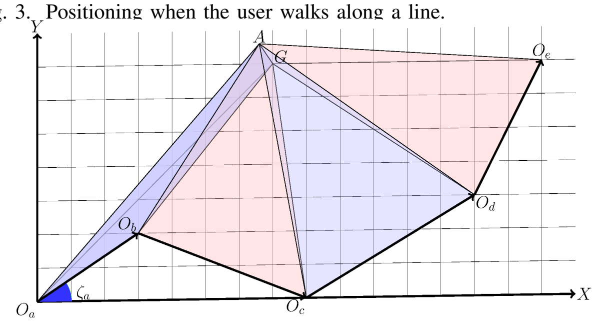


Fig. 4. Positioning when the user walks and turns.

- x: distance from O_1 to $\overline{H}$, *i.e.*, $x = |\overline{HO_1}|$.

The output x and y are calculated by using the maximum likelihood estimation. Specifically, as $|\overline{HO_i}| = x + (i-1)s$, $i = 1, 2, 3, \ldots$, denoting that

$$l'_i = \sqrt{y^2 + (x + (i-1)s)^2} \tag{1}$$

$$e_i = l'_i - l'_{i+1} - d_i \tag{2}$$

For n displacements $d_1, d_2, \ldots, d_n$, x and y can be solved from above n equations by

$$(x, y) = \arg\min_{x,y} \sum_{i=1}^{n} e_i^2 \tag{3}$$

Here we use the Newton's Method [27] to reduce computation overhead.

Practical case: In the practical case, we also consider the situation when a user turns direction while walking. Assume that the user starts from O_a and walks along the linear segment $\overline{O_aO_b}$, $\overline{O_bO_c}$, $\overline{O_cO_d}$, $\overline{O_dO_e}$ in Figure 4. Therefore, besides the input in the simple case, we also leverage the following input:

- n_i: step count along a linear segment. For example, n_a is the step count when the user walks from O_a to O_b.
- ζ_i: user's walking direction. For example, $\zeta_a = \angle O_bO_aX$. Assume that O_a is at $(0, 0)$. O_c is at the position $(c_x, c_y) = (n_a s\cos(\zeta_a) + n_b s\cos(\zeta_b), n_a s\sin(\zeta_a) + n_b s\sin(\zeta_b))$, and so forth.
- d_{c_i}: calculated displacement for each linear segment.

We calculate the relative position $G(g_x, g_y)$, which is the projection of acoustic speaker A at a horizontal plane. Then the real-time relative position can be directly obtained from $G(g_x, g_y)$.

Similar to Eq. (1), Eq. (2), the distance from each stride point to G is

$$l_{c_i} = \sqrt{[c_x + (i-1)s\cos(\zeta_c)]^2 + [c_y + (i-1)s\sin(\zeta_c)]^2 + h^2} \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

Denote the calculated error at the ith step along line $\overline{O_cO_d}$ as

$$e_{c,i} = l_{c_i} - l_{c_{i+1}} - d_{c_i} \tag{5}$$

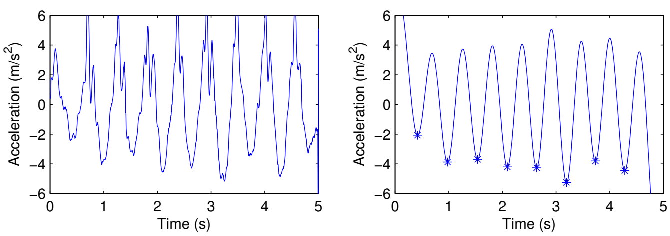
Hence, we obtain the position of G using the following equation:

$$(g_x, g_y) = \arg\min_{g_x, g_y} \sum_{i\in\{a,b,c,d,e\}} \sum_{j=1}^{n_c-1} e_{i,j}^2 \tag{6}$$

B. Processing Inertial Sensors

We generate 2 types of data from inertial sensors as input of relative positioning: the time of user's steps, and the angle of user's turning direction when the user walks.

Step Counter: We design the pedometer according to recent literatures, *e.g.*, [12], [28]. Specifically in Figure 5a, we use the inertial sensors, *i.e.*, the compass, accelerometer, gyroscope, to get real-time acceleration. Then, the samples are passed through a low pass filter as shown in Figure 5b. Finally, we choose bottom points in Figure 5b to indicate the time when the user steps at ground.



(a) Raw Acceleration (b) Step detection

Fig. 5. Step detection using the vertical acceleration.

Calculating the turning direction: When the user turns the walking direction, the rotation angle is calculated mainly by using the gyroscope. Note that WalkieLokie does not require the knowledge of absolute direction [12]. For instance, assume the initial walking direction is ζ_a and the following direction is ζ_b. We do not need the exact value of ζ_a or ζ_b. Instead, we directly calculate the difference of walking direction, *i.e.*, $\zeta_b - \zeta_a$, from the gyroscope. In the implementation, we obtain the device's initial orientation by using the rotation vector collected from the smart device. Then, the rotation vector is translated into quaternion [29], and the difference of walking direction is calculated using the quaternion together with the collected samples from the gyroscope. This implementation can avoid errors caused by magnetic sensor in indoor environments.

C. Ranging in Longer Distances

We design the long-distance ranging algorithm in case the speaker is far away from the smart device. As mentioned earlier, the accuracy of ranging by Eq. (1) is reduced when the distance increases. Though we propose to synthesize all the walking segments when a user walks and turns, the problem is that the method has accumulated errors when we estimate the current position by using the historical measured positions, the estimated walking direction and the pedometer. Especially when the user is far away and the signal is lost for a long time (*i.e.*, the Non-Line of Sight effects), the error increases and the historical measured positions can no longer be used.

The ranging algorithm leverages historical results of the basic positioning scheme and the periodical pulses added in the acoustic signals. When an accurate relative position is estimated (the calculated distance $< 8m$), we derive the distance l from the relative position, which implies the traveling time $t_l = l/v_a$ from the speaker to the smart device. At the same time, the device receives a pulse from the speaker, and the receiving time τ' of the pulse is calculated. Therefore, the sending time of the pulse is $\tau = \tau' - t_l$. Furthermore, the sending time of the latter ith pulse equals $\tau_j = \tau + jT$, where T is denoted as period of pulses. Thus, we calculate the distance l_j when the device receives the jth pulse as follows:

$$l_j = (\tau_j' - \tau_j)v_a = l + (\tau_j' - \tau' - jT)v_a \tag{7}$$

Here, τ_j' is the receiving time of the latter jth pulse.

Positioning after long-distance ranging: We recompute the position (L, ψ') according to the new ranging result as follows.

In Figure 3, assume that l_1 is obtained by synchronization. The distance at the horizontal plane is $L_1 = \sqrt{l_1^2 - h^2}$.

To recompute the direction ψ_1', since $x = -l_1\cos\psi_1$ and $y = l_1\sin\psi_1$, we can infer that $l_i' = \sqrt{l_1^2\sin^2\psi_1 + (-l_1\cos\psi_1 + (i-1)s)^2}$. From Eq. (2),

$$\psi_1 = \arg\min_{\psi_1} \sum_{i=1}^{n} (l_i' - l_{i+1}' - d_i)^2 \tag{8}$$

Then $\psi_1' = \arccos\frac{l_1\cos\psi_1}{\sqrt{l_1^2 - h^2}}$.

IV. Acoustic Processing

In this section, we design the method of modulating acoustic signals emitted from the speaker, and then the module of generating relative displacement and receiving time of periodical pulses, which are used for relative positioning.

A. Brief Design of the Acoustic Wave

The modulated wave $s(t)$ contains two parts $s_1(t)$ and $s_2(t)$, which are used for displacement tracking and long-distance ranging respectively. More specifically, we formally define the wave in the following equations,

$$s(t) = \begin{cases} s_1(t), & kT_2 \le t < kT_2 + T_1 \\ s_2(t), & kT_2 + T_1 \le t < (k+1)T_2 \end{cases} \tag{9}$$

where $T_2 = 0.25s$ is the cycle of the wave and k is the natural number. $T_1 = 0.16s$ is the duration of $s_1(t)$ in each cycle. In Section IV-B, we design $s_1(t)$ and the method of tracking relative displacement according to $s_1(t)$. In Section IV-C, we will discuss how to design and detect the signal $s_2(t)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

B. Tracking Relative Displacement

We adopt the second-order Phase Locked Loop (PLL) to track the phase, where the phase is proportional to the displacement [18]. PLL is a classical method in signal processing and can be regarded as a device that tracks the phase and frequency of a sinusoid. In our design, it is implemented purely by software due to the limited capabilities of smartphone platforms.

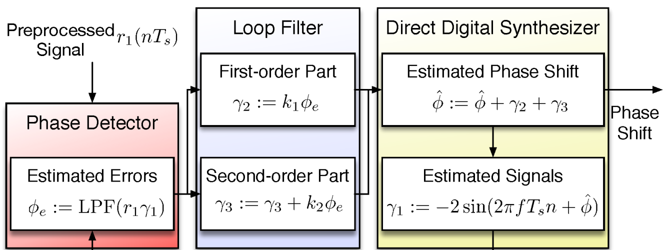


Fig. 6. Design of the Second-Order Phase Locked Loop.

We show our design of PLL in Figure 6. The PLL contains three main components: phase detector, loop filter and direct digital synthesizer (DDS). The phase detector detects the difference $\phi_e = \phi - \hat{\phi}$, where $\hat{\phi}$ is the estimation of ϕ. According to ϕ_e, the loop filter analyzes and predicts the offset $\gamma_2+\gamma_3$ of $\hat{\phi}$ for the next cycle of the loop, where the variances of γ_2, γ_3 are affected by parameter k_1 and k_2 respectively. The DDS updates the next $\hat{\phi}$ by adding the offset and prepares γ_1 for the next phase detection. The loop filter is the key part of PLL [30], and the type and parameter of loop filter should be carefully chosen for different purposes. Here we adopt a second-order filter, *i.e.*, the proportional-plus-integrator [31] filter, as the Loop filter. It uses two updated variables γ_2, γ_3 and two constant parameters k_1, k_2. Particularly, if $k_2 = 0$, it degrades to a first-order PLL.

We explain the reason of adopting the second-order PLL firstly by the following intuitive experiment: a user holds a smart device for a while, moves the phone forward to the speaker and backward for three times, and finally stops at the starting point. Figure 7 shows the results of PLL with different parameters, when the acoustic signal is weak ($l = 32m$). In Figure 7a and 7b, the first-order PLL is chosen (*i.e.*, $k_2 = 0$). For the case in Figure 7a, k_1 is large enough that the calculated displacement can catch up with the real displacement. However, it is affected by the noises which cause jitters on the calculated displacement. Then, the total displacement, which should be close to 0, accumulates to $17cm$ (the total moving length is about $100cm$). On the contrary, in Figure 7b where k_1 is small, the calculated phase cannot catch up with the real phase and jitters frequently when moving. In Figure 7c, we choose the second-order PLL (*i.e.*, $k_2 \neq 0$). Meanwhile, k_1 equals the one in Figure 7b that the PLL is more robust to noises and does not cause observable jitters. Finally, the accumulate error of measured displacement in Figure 7c is less than $2cm$, which is about at least 9 times more accurate than the one in Figure 7a.

Theoretically, the second-order PLL has better performance in noisy environments since its predicted displacement is closer to the real displacement. Specifically, as shown in Figure 6, PLL predicts the displacement by updating γ_2 and γ_3 in each iteration, where $\gamma_2+\gamma_3$ is the calculated phase shift, which corresponds to the relative displacement. When the first-order PLL is chosen (*i.e.*, $k_2 = 0$), $\gamma_3 = 0$ and $\gamma_2 = k_1\phi_e$ is the calculated phase shift. Therefore, k_1 should be large enough (*e.g.*, $k_1 = 5*10^{-2}$) so that $k_1\phi_e$ can catch up with the real shift. However, since ϕ_e is also affected by environmental noises, larger k_1 results in bigger errors in the estimated shift. When the second-order PLL is chosen, (*i.e.*, $k_2 \neq 0$), γ_2 and γ_3 are close to constant, when the relative speed from the smart device to the speaker is constant. Hence, k_1 can be much smaller (*e.g.*, $k_1 = 5 * 10^{-3}$), where the effects caused by noises are reduced. In our scenario, the relative walking speed is close to constant given the sampling frequency (44100Hz), so the second-order PLL is chosen in WalkieLokie.

Moreover, the second-order PLL is still robust against the signals $s_2(t)$ which are used for long-distance tracking and can be regarded as interferences for displacement tracking. In Section IV-C, we design the signal $s_2(t)$ and prove the robustness according to our experiments.

C. Pulse Modulation and Detection

1) Design goals: Several goals require to be achieved when designing the pulses $s_2(t)$ and the pulse detection algorithm:

- **Multiple concurrent speakers:** Each speaker should take different frequency band to avoid interferences, where the narrower bandwidth is preferred. In WalkieLokie, $s_1(t)$ and $s_2(t)$ share the same band to enable narrower bandwidth and more concurrent speakers. The challenge is that $s_2(t)$ should occupy more bandwidth if it can be successfully detected.
- **Robustness in displacement tracking:** $s_2(t)$ can also be used for displacement tracking by PLL. Otherwise, PLL will lose phase locks when processing $s_2(t)$.

2) Pulse modulation: Based on these requirements, we design $s_2(t)$:

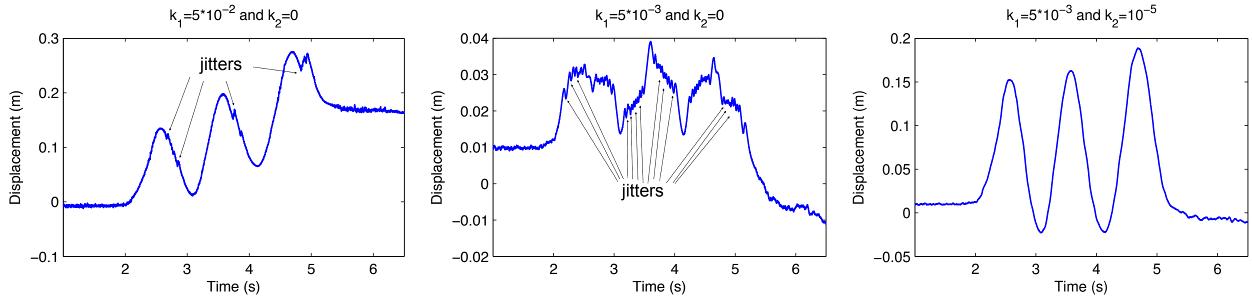
$$s_2(t) = \begin{cases} \cos(2\pi f t + \pi \sin \frac{\pi(t-\tau_i)}{T_p}) & \tau_i \leq t \leq \tau_i + T_p \\ \cos(2\pi f t) & \text{otherwise} \end{cases} \tag{10}$$

where we construct pulses starting at $\tau_1, \ldots, \tau_i$, and the duration of each pulse is T_p. We encode three adjacent pulses per $T_2 = 0.25s$. Three adjacent pulses can be seen as a compensated periodical pulse with the period $T = T_2 = 0.25s$. The time difference of the adjacent pulses is $T_3 = 0.03s$.

Analysis of bandwidth: Since the bandwidth of the pulse is about $\frac{\pi}{T_p}$ [32], we set $T_p = 0.007s$ so that the bandwidth is about 460Hz. As the minimum frequency is 17000Hz when the acoustic is inaudible, and the maximum frequency which is supported by the phone is 24000Hz, the maximum concurrent signals that WalkieLokie supports in one place is $(24000 - 17000)/460 \approx 15$. Essentially, a trade-off exists between the bandwidth of the signals and the number of the concurrent signals.

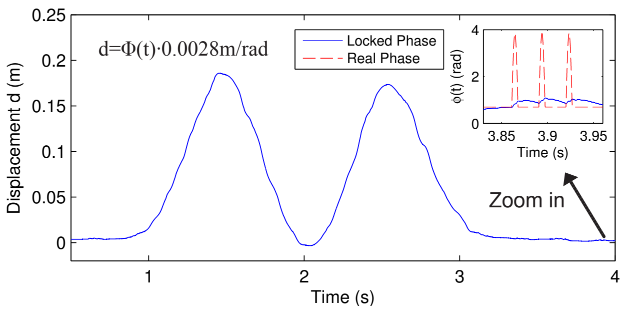
Analysis on effects of displacement tracking: In Figure 8a, the estimated displacement is smooth and there are no

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

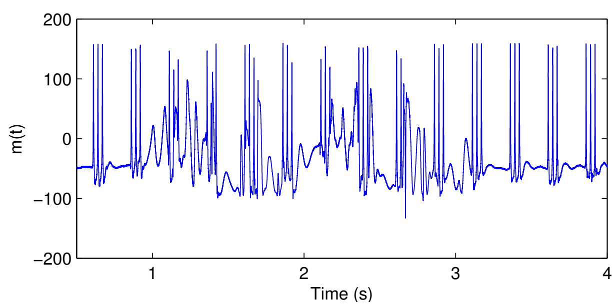


(a) First order, large k_1. (b) First order, small k_1 (c) Second order

Fig. 7. Calculated displacement by PLL with different orders and parameters, when the signal is weak.



(a) Measured Displacement



(b) Detected Pulses

Fig. 8. Calculated displacement and pulses from the same signal.

jitters whenever the phone is static or moving. We zoom in the calculated phase to show the performance of PLL when there are pulses in $s_2(t)$: the calculated phase is not locked to the real phase; instead, the tracked phase is very smooth and phase-shift effects caused by $s_2(t)$ can be omitted. Specifically, while the maximum variation of the real phase is π, the corresponding variation computed by PLL is less than 0.4rad, which corresponds to the displacement of about $1mm$. The cause of the phenomenon is that the parameters (k_1, k_2) of PLL are very small, and the fast changing phase cannot be tracked. Moreover, as the phase at the beginning of a pulse equals the one at the end and the variation is small, the tracked phase shift finally becomes stable and proportional to the displacement.

3) Pulse detection: We first propose basic method of detecting the receiving time $\tau_i' = \tau_i + t_l$ of the ith pulse by leveraging the component $s_3(t) = \pi \sin \frac{\pi(t-\tau_i)}{T_p}$. Assuming the locked phase by PLL is ϕ_r before the pulse starts, the expected pulse is $\tilde{r}(t) = \cos(2\pi f t + \phi_r + \pi \sin \frac{\pi(t-\tau_i')}{T_p})$. Hence, for the kth received sample $r(kT_s)$, we compute the likelihood $m(kT_s) = \sum_{i=k}^{k+T_p/T_s} r(iT_s)\tilde{r}(iT_s)$, *i.e.*, when $m(kT_s)$ reaches the maximum, the corresponding kT_s is the starting time of the received pulse. In Figure 8b, $m(t)$ reaches the peak value (*i.e.*, 150), when there is a pulse at t and the bottom value (*i.e.*, -50) when there are almost no pulses. As a whole, it shows an interesting result that on demodulating $s(t)$, the peak of $m(t)$ is very clear for synchronization in Figure 8b, while the corresponding calculated phase is very smooth for displacement tracking in Figure 8a.

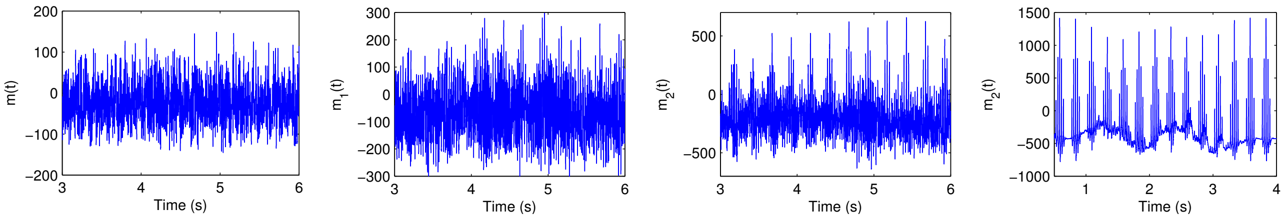
The basic pulse detection mechanism suffers many kinds of interferences, *e.g.*, noises or multipath effects. The moving of the smart device also influences the performance of pulse detection. Facing this problem, we improve the pulse detection mechanism by the following strategies:

Effects by noises and motion: The solution is based on the observation that expected peaks still appear at expected time, though they sink in the noises as shown in Figure 9a. Meanwhile, random peaks have fewer chances to appear periodically. Hence, we assign $m_1(t) = m(t-T_3) + m(t) + m(t+T_3)$ in Figure 9b, where the peaks are more clear to be identified in $m_1(t)$. Then, we assign $m_2(t) = m_1(t-T_2) + m_1(t) + m_1(t+T_2)$ in Figure 9c, where the peaks can be easily detected. Similarly, in the case when the phone is moving as shown in Figure 9d, the peaks are also very clear.

Multipath effects: The detected pulse is prone to be affected by multipath effects, when the phone is static according to the experiments in Section V. To solve the problem, we use $m_3(kT_2) = \sum_{i\in\{x|x=k \mod T_2\}} m(iT_2)$, which sums all the $m_3(t)$ of pulses and makes the detected time of pulses more clear. In Figure 10a, when there is no multipath effect, there are 3 pulses in a period T_2. However, in Figure 10b, which is gathered from the shopping mall, there are 9 pulses at least, which means there are 2 additional paths reflected from walls or other objects. In this case, all the 3 paths are the possible pulses directly received from the dummy speaker.

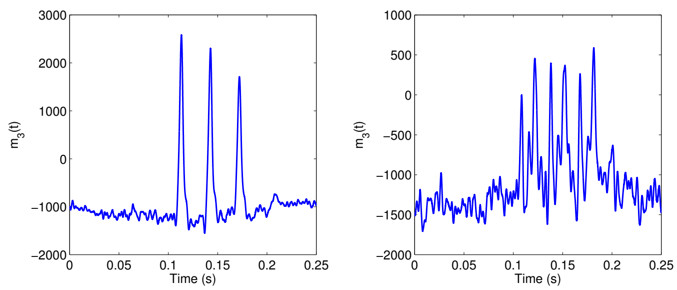
After recognizing the possible propagation paths, we make further steps to obtain the direct path by leveraging the calculated the displacement. Specifically, denote that the displacement calculated by PLL is d when the user walks from A to B. By using pulse detection, the possible receiving time of the pulses is in the set $T_a = \{t_{a1}, t_{a2}, t_{a3}, \dots\}$ and $T_b = \{t_{b1}, t_{b2}, t_{b2}, \dots\}$ when the user is at A and B respectively. Hence we obtain the receiving time $(t_a, t_b) = \arg\min_{t_a\in T_a, t_b\in T_b} |(t_a - t_b)v_a - d|$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications



(a) m(t) of Weak Signal (b) $m_1(t)$ of Weak Signal (c) $m_2(t)$ of Weak Signal (d) $m_2(t)$ of Good Signal

Fig. 9. Detection of the arrival time of the pulse.



(a) Good Case. (b) Bad Case (Multipath).

Fig. 10. Pulse detection in case of multipath effects.

4) Supporting more concurrent targets: we find that further optimizations can be made to support more concurrent speakers as follows: *a) Virtual business card sharing:* we can support more concurrent signals by narrowing the bandwidth of pulses or do not add pulses that are used for long-distance ranging, which is no longer needed when the users are in vicinity. *b) Virtual shopping guide:* we can use only a few speakers (pre-deployed by the WalkieLokie group) for the normal indoor localization, if more shopping guides are required. Our further evaluations in Section V-C prove that WalkieLokie supports unlimited number of shopping guides by simple and sparse deployment of speakers, *i.e.*, the smart device only receives signals from 2 speakers on average, but gains sub-meter accuracy.

V. Performance Evaluation

In the experiment, we first evaluate the accuracy of WalkieLokie in an empty room and then study the robustness against different factors, *e.g.*, relative positions, number of steps, users, turning directions, environments.

Experimental Setup: We perform system evaluation by using two types of speakers: Samsung Galaxy Note 2 and normal dummy speakers. The speaker merely broadcasts acoustic waves and does not perform communications. Specifically, the speaker plays a .wav audio file with the sampling rate of 44100Hz and the central frequency of 19000Hz. We mainly use Google Nexus 4 to receive the acoustic signals. We do not make any modifications to the phone or jailbreak the operating system, and all the components, such as BPF, AGC, PLL, are implemented by the software. We evaluate the performance in an empty room, an office and the shopping mall. The micro benchmarks are made for position estimation and long-distance ranging. We then evaluate the total performance.

A. Position Estimation

We evaluate position estimation in several types of cases, *i.e.*, different related positions from the phone to the speaker, number of walking steps, users, orientation of devices, device diversity and environments, which may affect accuracy of the estimation.

1) Positions: We make evaluations in an empty room as shown in Figure 11a, where the speaker is placed at 16 different positions which are uniformly distributed in a square, *i.e.*, (X, Y) where $X \in \{2, 4, 6, 8\}$, and $Y \in \{2, 4, 6, 8\}$. Here, the empty room is large ($> 1000m^2$) where the multipath effects on measuring the relative displacement can be ignored. We let the user walk for $9 \sim 10$ steps with the walking length of about $6m$. The relative height h is about $0.3m$. For each location, the user holds the phone in hand and walks for 35 times to gather samples, *i.e.*, we get 560 samples in this micro benchmark. Then, we calculate the initial relative position (X, Y) when the user starts walking and the corresponding distance and direction.

In Figure 12a, the accuracy of distance estimation is very close for different X. We further study the distribution of large errors in Figure 12b. We find an interesting fact that the errors are nearly proportional to Y. Hence, when $Y = 2, 4, 6, 8m$, the corresponding errors are within $0.35m$, $0.55m$, $0.97m$, $1.88m$ at the percentage of 80%. For the direction estimation, it is still very accurate when X or Y increases in Figure 12c, 12d. As a total, the mean errors of ranging and direction finding are $0.63m$ and $2.46°$ respectively.

To illustrate why ranging is less accurate when the distance increases, we make simulations by showing relative displacements of each user's step when users walk from different starting points in Figures 12g and 12h. In Figure 12g, $Y = 5m$. When $X = 5m, 10m, 15m, 20m$, the mean variances of displacements ($|d_i - d_{i-1}|$) are 0.055m, 0.016m, 0.004m, 0.002m, respectively. Intuitively, as recalled in Section II-B, since the distance is calculated according to the variance of measured displacements, the ranging is more error-prone, when the variance becomes smaller. Specifically, if the mean error of measured variance is $0.005m$ per user's step, and $Y = 5m$ is given, the errors of calculated X are $0.66m, 1.09m, 5.07m, 17.6m$, when $X = 5m, 10m, 15m, 20m$ respectively. In a similar way, in Figure 12h where $X = 5m$, the errors of calculated Y are $0.74m, 1.78m, 3.5m, 5.9m$, when $Y = 5m, 10m, 15m, 20m$ respectively. Furthermore, when the distance increases, the error of measured variance also increases, because SNR of signals decreases. As a total, the ranging is less accurate which motivates us to design the long-distance ranging scheme.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

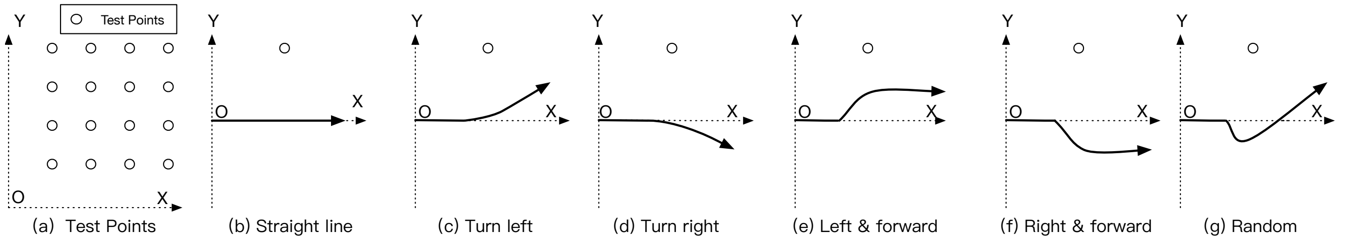


Fig. 11. Experiment configuration in an empty room: test points (a) and different walking paths (b-f).

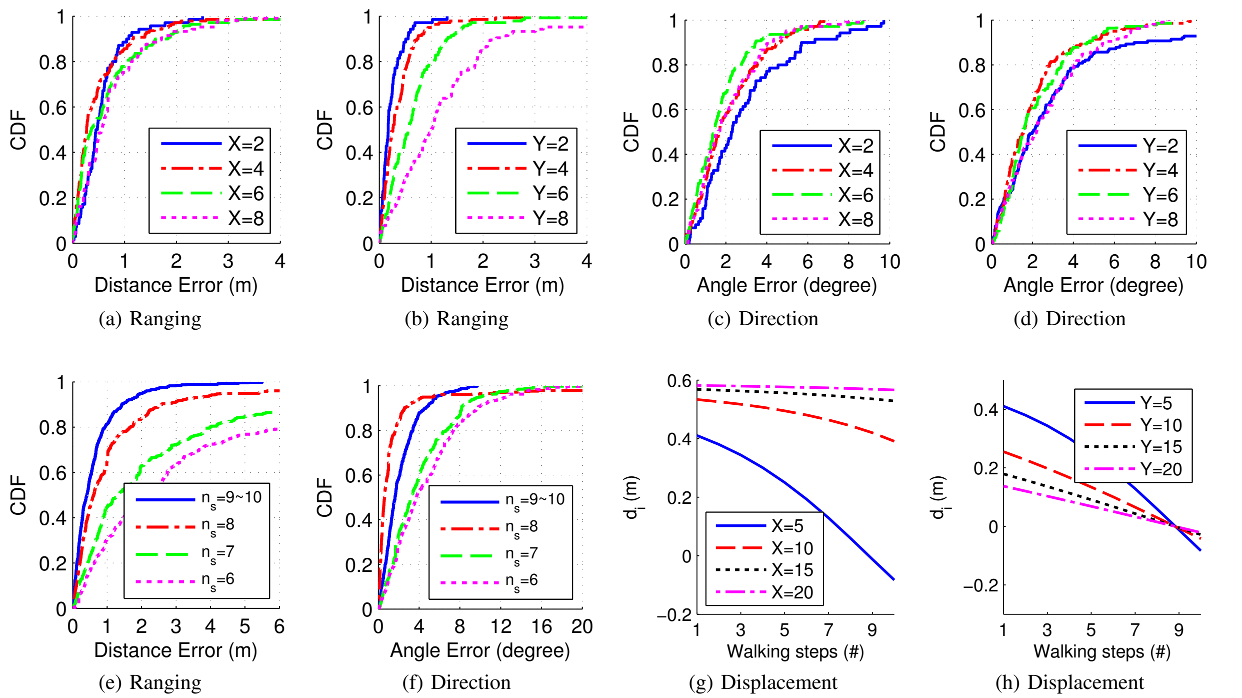


Fig. 12. The accuracy of ranging and direction finding. 1) when the user starts walking at different positions (a)(b)(c)(d), 2) when the user walks for smaller number of steps (e)(f). Relative displacements when users walks from different starting points.

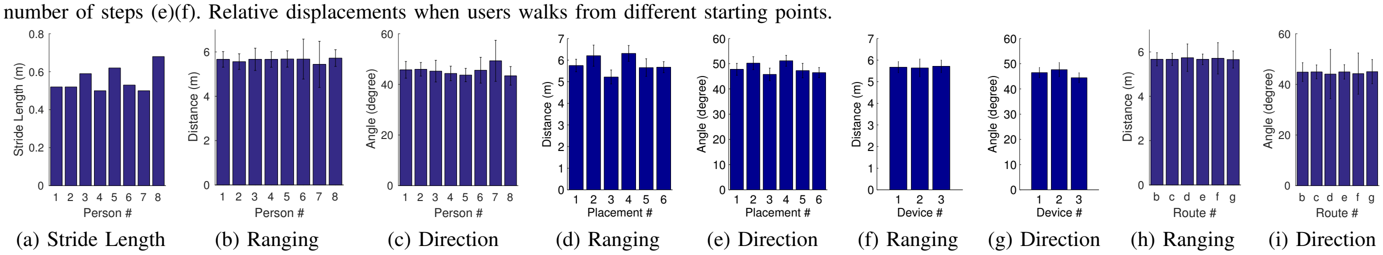


Fig. 13. The mean and standard deviation of ranging and direction estimation for different users (a)(b)(c), placements (d)(e), smart devices (f)(g), and walking paths (h)(i).

2) Number of Steps: The accuracy of the position estimation depends on the number of walking steps. We compare the results when the user walks for smaller number of steps n_s in Figure 12e, 12f. The results show that the ranging errors increase quickly when n_s reduces. The reasons are: 1) the user's stride length varies occasionally. 2) User's phone also shifts left and right regularly, *i.e.*, it does not move strictly in a line, when the user holds the phone and walks. As these facts will have less effects on the accuracy when n_s is larger, it can be foreseen that the accuracy will continue to be improved when $n_s > 10$, though it is already very accurate when $n_s = 10$.

The estimated direction is also affected by the smaller n_s in Figure 12f. But it is still acceptable that the angle errors are under $8°$ at the percentage of 80%, when $n_s = 6$. As a whole, when n_s is small, the direction estimation is still accurate.

According to the experimental results, in Swadloon [18], where the direction is estimated via the phone shaking movement, the small displacement ($< 10cm$) of phone shaking cannot be leveraged for ranging. Nevertheless, the displacement can be used for direction estimation. In the case of virtual shopping guide, the user can walk for more steps to get more accurate relative position. An interesting note is that when the user is walking closer to the speaker (for more steps), the obtained position is more accurate. The changing accuracy happens to meet the user's practical requirement.

3) Users: Different users have different stride lengths and user motions when users walk, which may affect the positioning result. Hence, we recruit 8 volunteers in this experiment: each user walks in a line of about $6m$ for 35 times where

$(X, Y) = (4, 4)$. We have the following observations in Figure 13: the standard deviations (std) of the ranging and direction are small for most users. In Figure 13a, the person 1,2,4,6,7 have small stride lengths while the rest ones have bigger length, but the result is similar among the users (except for the person 6,7). The results imply that the stride length is very stable and the positioning accuracy is not much affected by variation of stride length, though the stride length between different users may be much different.

4) Orientation of Speaker and Microphone: we consider the cases when the speaker or the microphone faces to different directions: (1) (default) the microphone faces to the sky, and the speaker faces to the walking line. (2) microphone, facing to the front. (3) microphone, perpendicular to the walking direction and facing to the speaker. (4) microphone, facing to the ground. (5) microphone, perpendicular to the walking direction and speaker is at the back of the microphone. (6) speaker, facing to the ground. The result in Figure 13d, 13e shows that the std is small in all cases and the result is very stable.

We also find that the mean value of distances increase when the signals are weaker in case (2), (4) and decreases when the signals are stronger in case (3). The reason is that when the signals are weak, PLL will lose some signals and the tracked displacement decreases, which makes the calculated distance become larger. Hence, based on our measurements in displacement tracking, we make calibrations on the calculated PLL. Specifically, the displacement $d = 1.22\frac{v_a}{2\pi f}\Delta\phi$, if $d > 0$; and $d = 1.69\frac{v_a}{2\pi f}\Delta\phi$, if $d < 0$, where $\Delta\phi$ is the tracked phase shift. Note that we make calibration with a constant factor (*i.e.*, 1.22), for the environment has limited effects on the result of PLL when the signals are strong enough. However, when $d < 0$, which means the speaker is at the back of the walking user, d is usually not used for position estimation if the tracked phase is abnormal (*e.g.*, WalkieLokie cannot detect pulses from the tracked phase).

5) Device Diversity: We test several Commercial Off-the-Shelf (COTS) smart devices as acoustic receivers: (1) Nexus 4, (2) Samsung Galaxy Note 2. (3) Nexus 7. We choose $(X, Y) = (4, 4)$, and the error of position estimation is shown in Figure 13f, 13g. The result shows that these smart devices have similar performance.

We also use normal dummy speakers as acoustic sources when we conduct the experiment in a large shopping mall, for we consider the case that the normal speakers serve as virtual shopping guides.

Calibration of clock drift: We find that the normal speaker, which is different from the previous smart devices, has serious clock drift and needs to be calibrated. For instance, when a speaker is supposed to broadcast signals at 19000Hz, the actual frequency of the signals is 19007Hz. If the frequency drift is 0.1Hz, the error of distance measuring is about 600*340*0.1/19000=1.07m, when the smart device works for 10 minutes. To solve this problem, our design of PLL measures the precise clock offset when the receiver is static for seconds. In this case, γ_2 in Figure 6 rapidly converges to a constant value. As γ_2 equals the phase shift per sampling time T_s, the frequency offset equals $\frac{k_2}{2\pi T_s}$. Hence, once we let the smart device be static for seconds, the precise frequency offset is obtained. Afterwards, we calibrate the clock drift in real-time using the constant frequency offset.

6) Turning Directions: We also evaluate the performance when user turns directions. In Figure 11b$\sim$g, we choose 6 different routes: b) straight line, c) walking forwards and turning left (the turning angle is around $30^\circ \sim 40^\circ$), d) walking forward and turning right, e) walking forward, turning left, and turning back (forward), f) walking forward, turning right, and turning back, g) walking randomly. We set the speaker at $(X, Y) = (4, 4)$.

In Figure 13h,13i, the results show that relative positioning is accurate in cases of different routes. Especially in case that the user walks randomly, std of ranging and direction finding are $0.39m$ and 4.8° respectively. Another observation is that when the user turns left (*e.g.*, case c,e), the accuracy increases, *e.g.*, the std of ranging and direction finding in case c) are 0.27m and 2.7° respectively. On the other hand, when the user turns right (*e.g.*, case d,f), the accuracy decreases, *e.g.*, the std of ranging and direction finding in case d) are 0.62m and 9.7° respectively. The reason is that when the user turns right, the relative position changes, *i.e.*, X decreases, and Y increases, as shown in Figure 3. According Figure 12a,12b, the error of ranging increases. Finally, the accuracy of relative positioning is reduced. Though the errors vary according to different turning angles, WalkieLokie is still practical that when the user's walking direction turns to the one that is close to the speaker, the accuracy increases. On the other hand, when the user turns the direction and walks far away from the speaker, which implies that the user is not interested in the information provided by the target, the positioning result becomes less important and more coarse-grained.

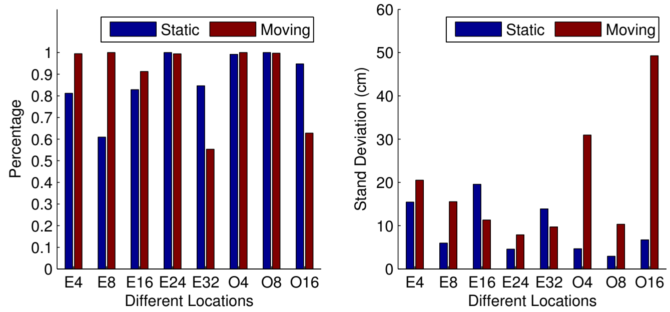
7) Environments: We compare the accuracy of position estimation in the empty room and at different locations in the office. We find that they show the similar results. We further evaluate the effects in a shopping mall in the latter subsection.

B. Long-Distance Positioning

1) Pulse Detection: In Figure 14, we choose 8 locations in an empty room and the office to evaluate the performance of pulse detection. Here, E32 means that the experiment is in the room and the distance from the smart device to the speaker is $32m$, and O16 means that it is in the office and the distance is $16m$. In each position we test two cases: the phone is static or moving back and forth without stop. For each case, the phone records the audio for 100 seconds, which means there are 400 signals for pulse detection in the samples. Then, we evaluate the accuracy of pulse detection. For easier understanding of our results, the error of arrival time t_e is converted to distance measurement error $l_e = v_a t_e$. For instance, if the error is the time interval of 1 acoustic sample, *i.e.*, $t_e = \frac{1}{44100}s$, the corresponding distance error is $l_e \approx 0.8cm$.

Since we find that there are occasional significant errors ($> 3m$), we first set threshold $l_t = 80cm$ and evaluate ratio of successful detection that $l_e < l_t$. In Figure 14a, the successful detection rate is above 80% for most cases when the phone is static. When the phone is moving, the performance is good

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications



(a) $80cm$ criterion. (b) Stand Deviation

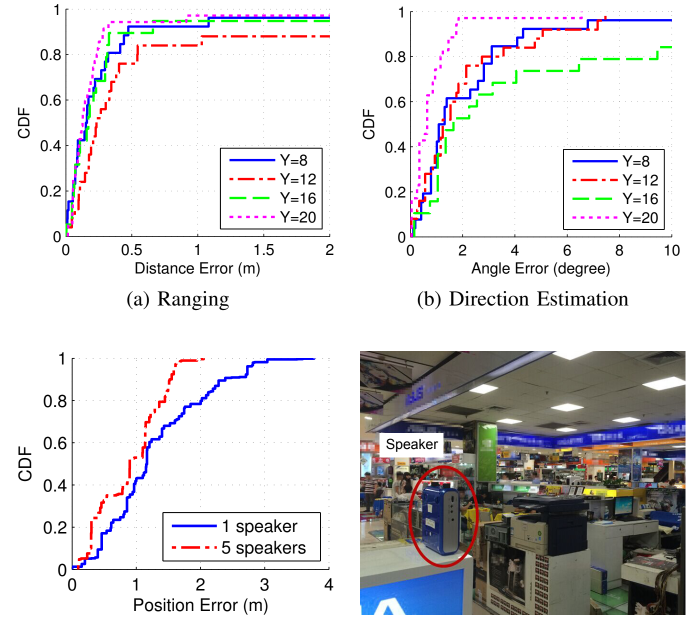
Fig. 14. (a) Percentage of successful experiments at different locations, (b) standard deviation.

as well if the distances are within $24m$ or $8m$ in the empty room or office respectively. In some cases the rate is close to 100%.

There is also an exception that at location E8 when the phone is static, the rate is only 61.0%, while it reaches 100% at the same place when the phone is moving. Hence, we conduct the experiment again at the same place, and the result is close to the previous one. We suppose it is caused by the multipath effects: the phase ϕ_r changes according to the mixed signals and becomes stable when it is static, which affects the result of pulse matching. The reason of high successful rate in case of moving phone is that: though it is also affected by multipath, the phases of reflected signals at different positions are irregular. In other words, the PLL locks the phase of the signals directly from the speaker, *i.e.*, the multipath signals are regarded as noises by PLL. Hence, the performance is better when the phone is moving. We find the location E4, E8, E16 also have the same phenomenon, which validates our hypothesis. Actually, this is a good result for WalkieLokie: when the user is walking, the result of pulse detection is very good and can be directly used for synthesizing; when the user is walking, as the successful detection rate is above 60%, WalkieLokie collects enough samples and then determines the most possible receiving time. In Figure 14b, we show the standard deviation of results in case of successful detections. The std in most cases are around $10cm$ except that the std are $30.9cm$ and $49.2cm$ when the phone is moving at O4 and O16 respectively.

2) Long-Distance Ranging and Direction Estimation: We emulate the process that the user walks for a long period where the synthesizing cannot work due to large accumulated errors in ranging. Then, we evaluate the performance of long-distance relative positioning by the experiment as follows:

1) The user walks in a line where the initial coordinate of the speaker is $(4, 4)$. In this step, we calculate the distance through position estimation and then calculate the sending time of periodical signals $s_2(t)$ by pulse detection.
2) The user then turns, walks and stops at the position where relative coordinate of speaker is (X, Y). In this step, WalkieLokie does not perform any acoustic processing.
3) The user walks again for about $6m$. The position, which is supposed to be (X, Y), is then computed according to the sending time and the acoustic and inertial sensor samples.



(c) Position Errors (d) Shopping mall.

Fig. 15. (a)(b) Positioning after pulse detection. (c) Result of relative and absolute positioning. (d) The shopping mall and the dummy speaker.

We conduct the experiment in the empty room and the office. Specifically, we set $(X, Y) = (4, 12)$ and $(4, 20)$ in the empty room to gather the samples and $(4, 8)$ and $(4, 16)$ in the office.

In Figure 15a, the ranging errors are under $0.32m$ and $0.66m$ at the percentage of 80% and 90% for most cases. There are also occasional errors for each cases which are greater than $2m$. They are caused by the multipath effects in pulse detection. Especially for the case of $Y = 12m$ in the empty room, the big errors are at the percentage of 12%. We can find the corresponding results at E8 and E16 shown in Figure 14a, where the successful detection rate is also much lower than other cases in pulse detection. Actually, since the successful detection rate in pulse detection is above 80% for most cases, the result would converge to the correct value and the abnormal result would be eliminated, if enough sampling time is given.

C. Putting it All Together in a Severe Environment

We evaluate WalkieLokie in a shopping mall, where the environment is quite severe for acoustic based systems: the shopping mall itself is broadcasting loud audios; there are always people walking around who block the sight line of speakers or block the road that we have to turn walking direction. Furthermore, as it may affect the business if we set up speakers on the ceiling and conduct frequent debugging (which may have better results), we only put the speakers at the side of the aisles, as shown in Figure V-C, 15d.

We evaluate the performance of positioning in two cases: a) *relative* positioning by *one speaker*. b) *absolute* positioning by *5 speakers* (like normal indoor localization). We choose a $35m \times 17m$ area (about $600m^2$) in Figure V-C, and put 5 normal dummy speakers in this area. Each speaker broadcasts signals at different central frequencies, which are inaudible

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

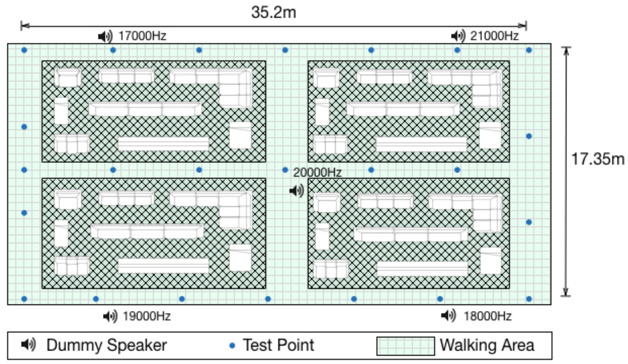


Fig. 16. The map of the shopping mall.

and not discovered by surrounding customers. We emulate the behavior of normal shopping users in evaluation: the experimenter stands at a test point and walks for a few steps (less than 6m) in a line; then he stops or turns the direction and continues walking, and so on. We gather 8 samples per spot. Hence, we can evaluate the performance when leveraging all the walking segments with different walking directions to get the position.

We set the central frequency of the speakers to 17000Hz, 18000Hz, 19000Hz, 20000Hz, 21000Hz, respectively. The smart device differentiates the signals by using the subcomponent BPF in the Figure 2. For example, if we need to analyze the signals of the second speaker (18000Hz), we set the frequency band of BPF which filters the signals at 18000Hz, and other signals are attenuated. Note that in practice the central frequency can be detected according to FFT, and the corresponding parameters of the BPF can be set automatically.

The results show that these speakers have much different performances in relative positioning, though they are the same product model. The signals of speaker at 17000Hz only cover 13% of the area, but the signals of speaker at 19000 and 20000 cover about 54% and 51% of the area. This diversity may be caused by several facts: anchor positions, quality of different anchor speakers, etc. The main reason is that the sound signal sent by speakers has been greatly attenuated when the frequency is higher than 8Hz [20], and the amplitude of attenuation is different among different speakers. We leave the study on configuration of speakers in our future work. Totally, the overall coverage per speaker is 38%, which is about $222m^2$ in our specific area.

We show the accuracy of relative positioning when using one speaker in Figure 15c. Note that we only evaluate the accuracy of the relative position where the starting point is covered by the signals of the speaker. In case of NLoS effects, we simply leverage the inertial sensors (*e.g.*, the step counter and gyroscope) to estimate the real-time position, instead of acoustic processing. Though we can still estimate position according to the historical positioning result when there are no signals, we exclude the results of this case. The results show that for one speaker, the position errors are under $1.2m, 2m$ at the percentage of 50% and 80%. The mean error of relative positioning is $1.28m$.

As mentioned earlier, in order to support more concurrent speakers in virtual shopping guide, we propose the localization scheme using sparse deployed speakers as anchors. Here, we also evaluate the accuracy of calculating the absolute position of the user when all 5 speakers are used, where the absolute positions of the speakers are given as input. We evaluate the accuracy at all test points and the results show that the position errors are under $1.5m$ at the percentage of 90%. Since the average coverage per speaker is 38%, the smart device can receive audio from $38\% * 5 \approx 2$ speakers on average. Therefore, the accuracy is better when using multiple signals for positioning.

D. Overhead

The computation overhead is mainly caused by 3 components: displacement tracking (Including BPF, AGC, PLL), pulse detection and position estimation. We run WalkieLokie using matlab R2013a, and the CPU is 3.1GHz Intel Core i5. For 1 second of received samples, phase tracking, pulse detection, and position estimation take 0.09s, 0.12s, 0.05s respectively. For the computer-vision based annotation system, a convincing way [33] is to leverage Google's Project Tango, which presents the premise that tomorrow's hardware might have computational and (therefore) visual sensory powers far beyond anything on the marketplace today. Comparatively, the COTS desktop computer is sufficient for data processing in WalkieLokie. In fact, there is a trade-off between the overhead and accuracy. For example, we can use infinite BPF instead of finite BPF, which reduces the computation overhead significantly, but incurs larger errors. For the smart devices, it is recommended to send the recorded samples to cloud server, and obtains the result from the cloud, which requires much less computation overhead, meanwhile with low energy consumption. In this case, the major communication overhead is caused by sending acoustic samples (44.1KHz), while the overhead of sending inertial samples can be ignored (0.2KHz). According to our experiments, the receiver Nexus 4 can work for about 180 minutes, and Samsung Galaxy Note 2 can broadcast audio for about 240 minutes. To reduce the communication overhead, another practical solution is to use the smart device to track the displacement via signal processing, and use the cloud to receive the result of tracked displacement and perform the rest of the data processing. Moreover, WalkieLokie can be launched on-demand with energy-efficient context sensing, *e.g.*, [26], which can further reduce the energy cost.

VI. Related Work

WalkieLokie can perform both ranging and direction finding where only a speaker is required, which is achieved by leveraging walking motion of users. It can be used for both relative positioning and indoor localization. As shown in Table I, we compare WalkieLokie with other systems as follows.

A. Ranging

Recent ranging systems [9], [35], [38], [39] require special hardware for the synchronization purpose. A general idea is that the audio sender and receiver records the sending time and the receiving time of the pulses in the audio respectively,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

TABLE I
COMPARISON WITH EXISTING APPROACHES.

System	Requirements of targets or anchors	Relative Positioning	Ranging / Direction	Long Range (>20m)	Density of Anchors	Mean Error
WalkieLokie	A Dummy Speaker	Yes	Both	Yes	5 in $600m^2$	$1.28m$ in localization
Swadloon [18]	Dummy Speakers	No	Direction	Yes	6 in $108m^2$	$0.5m$ in localization
BeepBeep [20]	Smartphone	No	Ranging	No	N/A	$0.02m$ in ranging
Guoguo [9]	Speakers + Sync	No	Ranging	No	9 in $16m^2$	$0.1m$ in localization
RSSI [11], [34]	Wifi or bluetooth	No	Ranging	Yes	N/A	meter-level
Bat [35]	Speaker + Networks	No	Ranging	No	100 in 280 m^3	0.05m in localization
[36]	Smartphone	Yes	Both	No	N/A	$0.19m$ in positioning
[37]	WIFI AP	No	Direction	Yes	N/A	30° in direction

and the distance is calculated according to the difference of the time. In this case, synchronization among the sender and the receiver is required, *e.g.*, exchanging the time information via networks. In Bat System [35], the base-station uses radio channel and communications for synchronization. Cricket [38] uses special device to send the RF signal together with the ultrasound signal at the same time. Then the receiver obtains the distance according to the different traveling time of the two signals. Guoguo [9] uses RF signals to synchronize all the acoustic anchors, the location can be obtained according to the difference of the receiving time by the phone. BeepBeep [20] calculates the distance between the phones. It solves the synchronization problem by letting two phones emit acoustic signals and exchange the sending and receiving time via wireless channel.

Other indoor localization systems perform coarse-grained ranging via Received Signal Strength Indicator (RSSI), *e.g.*, Wifi-based System, [10]–[14], or bluetooth iBeacon [34]. The advantage is that no special infrastructure is required to be deployed. Specially, for some simple location-based services, such as AR advertising, users can get the services when they come near an iBeacon. However, like other ranging systems, at least 3 anchors are required for localization, while WalkieLokie only requires one speaker.

B. Direction Estimation

The popular direction estimation systems use specialized hardware, such as directional antenna [40]–[42] or antenna array [8], [41], [43]. There have been proposals without requirement of specialized hardware as well. [37] emulates the functionality of a directional antenna by rotating the phone around the user's body, to locate outdoor APs. [36] leverages multiple microphones of the smartphone and communication channels for positioning within 4 meters, which is used for short-distance positioning and phone-to-phone games. Some other methods leverage Doppler effects by swinging [44] or shaking [18] the phone. [45] calculates direction by head nodding or shaking using smart glasses. They are based on different frequency shifts when the phone is moving at different directions. Compares to [18], [44], WalkieLokie makes further steps that a user can obtain direction without any additional actions on the phone so that she/he can get the real-time direction while walking. Recall that Swadloon [18] cannot calculate the distance but only the direction from a smartphone to a single acoustic source, due to the short displacement of phone-shaking movement. WalkieLokie performs ranging by leveraging the longer displacement of the walking motion via acoustic signal processing. WalkieLokie also adds more signals in the acoustic wave for additionally supporting long-distance ranging, without the loss of accuracy in displacement tracking via acoustic signals. Furthermore, the bandwidth of the signals is narrowed for supporting multiple concurrent sources.

VII. DISCUSSION

Though WalkieLokie improves the ubiquity for it only requires a dummy speaker to enable relative positioning, it is not the best in every aspect when it is directly applied to ranging or the indoor localization. Compared with the ranging system BeepBeep [20], WalkieLokie does not achieve centimeter-level accuracy, though it can additionally perform direction finding and the target does not require rich functionalities, *i.e.*, audio recording, computation, or wireless communications. WalkieLokie outperforms the existing anchor-based indoor localization systems [9] that it requires less anchors to achieve sub-meter accuracy, since WalkieLokie can calculate both directions and distances. However, many indoor localization systems only require the pre-deployed infrastructures, such as Wifi [10]–[15], Visible Light [16], [17]. Hence, we will explore more ubiquitous positioning systems in our future work.

VIII. CONCLUSION

We propose WalkieLokie to calculate relative position from a user to a target. WalkieLokie can be launched as long as the target is in sight, and the target only needs a dummy speaker that emits acoustic signals at inaudible frequency. Due to its improved ubiquity, it not only can be used for normal indoor localization, but also can be directly applied to new AR applications.

IX. ACKNOWLEDGMENTS

The research is supported by National Natural Science Foundation of China under Grant No.61572453, No.61202404, No.61520106007, No.61170233, No.61232018, No.61572454, No.61632010, Anhui Provincial Natural Science Foundation, No.1508085SQF215, and NSF of Jiangsu for Distinguished Young Scientist, No.BK20150030. The research of Li is partially supported by China National Funds for Distinguished

Young Scientists with No. 61625205, Key Research Program of Frontier Sciences, CAS, No.QYZDY-SSW-JSC002, NSFC with No.61520106007, NSF ECCS-1247944, NSF CMMI 1436786, and NSF CNS 1526638.

References

[1] (2016) Pokémon go. http://www.pokemongo.com.

[2] (2016) Microsoft hololens. https://www.microsoft.com/microsoft-hololens.

[3] (2016) Google glass. http://www.google.com/glass/start/.

[4] (2016) Wikitude. http://www.wikitude.com.

[5] (2016) Augmented car finder. [Online]. Available: http://www.augmentedworks.com/app/find-your-car-with-ar-augmented-car-finder/

[6] Z. Yang, C. Wu, Z. Zhou, X. Zhang, X. Wang, and Y. Liu, "Mobility increases localizability: A survey on wireless indoor localization using inertial sensors," *ACM Comput. Surv.*, vol. 47, no. 3, p. 54, 2015.

[7] H. S. Maghdid, I. A. Lami, K. Z. Ghafoor, and J. Lloret, "Seamless outdoors-indoors localization solutions on smartphones: Implementation and challenges," *CSUR*, vol. 48, no. 4, p. 53, 2016.

[8] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *NSDI*, 2013, pp. 71–84.

[9] K. Liu, X. Liu, and X. Li, "Guoguo: enabling fine-grained indoor localization via smartphone," in *MobiSys*, 2013, pp. 235–248.

[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 *MobiCom*, 2012.

[11] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM*, vol. 2. Ieee, 2000, pp. 775–784.

[12] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in *MobiCom*, 2012, pp. 293–304.

[13] S. He, S.-H. G. Chan, L. Yu, and N. Liu, "Calibration-free fusion of step counter and wireless fingerprints for indoor localization," in *UbiComp*. ACM, 2015, pp. 897–908.

[14] H. Xu, Z. Yang, Z. Zhou, L. Shangguan, K. Yi, and Y. Liu, "Enhancing wifi-based localization with visual clues," in *UbiComp*. ACM, 2015, pp. 963–974.

[15] J. T. Biehl, M. Cooper, G. Filby, and S. Kratz, "Loco: a ready-to-deploy framework for efficient room localization using wi-fi," in *UbiComp*. ACM, 2014, pp. 183–187.

[16] Q. Xu, R. Zheng, and S. Hranilovic, "Idyll: indoor localization using inertial and light sensors on smartphones," in *UbiComp*. ACM, 2015, pp. 307–318.

[17] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, "Epsilon: A visible light based positioning system," in *NSDI*, 2014, pp. 331–343.

[18] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, and Y. Liu, "Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones," in *INFOCOM*, 2014.

[19] W. Huang, Y. Xiong, X. Li, H. Lin, X. Mao, P. Yang, Y. Liu, and X. Wang, "Swadloon: Direction finding and indoor localization using acoustic signal by shaking smartphones," *TMC*, vol. 14, no. 10, pp. 2145–2157, 2015.

[20] 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, pp. 1–14.

[21] W. Huang, X.-Y. Li, Y. Xiong, P. Yang, Y. Hu, X. Mao, F. Miao, B. Zhao, and J. Zhao, "Walkielokie: Sensing relative positions of surrounding presenters by acoustic signals," in *UbiComp*, 2016.

[22] M. Lorenzen and L. Frederiksen, "Why do cultural industries cluster? localization, urbanization, products and projects," *Creative cities, cultural clusters and local economic development*, pp. 155–179, 2008.

[23] N. R. Desai, "Electronic business card system," 1996.

[24] A. Cuervo, "Smart business card," 2014.

[25] (2016) Project jacquard by google. https://www.google.com/atap/project-jacquard/.

[26] S. Nath, "ACE: exploiting correlation for energy-efficient and continuous context sensing," in *MobiSys*, 2012, pp. 29–42.

[27] K. Madsen, H. B. Nielsen, O. Tingleff, and M. Modelling, "Imm methods for non-linear least squares problems," Technical University of Denmark, DTU, 2004.

[28] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *MobiCom*, 2012, pp. 269–280.

[29] S. L. Altman, *Rotations, Quaternions, and Double Groups*. Oxford University Press, 1986.

[30] R. Best, *Phase Locked Loops: Design, Simulation, and Applications*, ser. Professional Engineering. Mcgraw-hill, 2007.

[31] M. Rice, *Digital Communications: A Discrete-Time Approach*. Prentice Hall, 2008.

[32] O. P. Sahu and A. K. Gupta, "Measurement of distance and medium velocity using frequency-modulated sound/ultrasound," *IEEE T. Instrumentation and Measurement*, vol. 57, no. 4, pp. 838–842, 2008.

[33] P. Jain, J. Manweiler, and R. R. Choudhury, "Overlay: Practical mobile augmented reality," in *MobiSys*, 2015, pp. 331–344.

[34] M. Estel and L. Fischer, "Feasibility of bluetooth ibeacons for indoor localization," in *Digital Enterprise Computing (DEC 2015), Böblingen, Germany, June 25-26, 2015*, 2015, pp. 97–108.

[35] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," in *MOBICOM*, 1999, pp. 187–197.

[36] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *SenSys*, 2011, pp. 190–203.

[37] 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, pp. 109–120.

[38] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *MobiCom*, 2000, pp. 32–43.

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

[40] M. Kim and N. Y. Chong, "Direction sensing rfid reader for mobile robot navigation," *TASAE*, pp. 44–54, 2009.

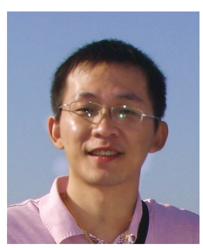
[41] A. Subramanian, P. Deshpande, J. Gaojgao, and S. Das, "Drive-by localization of roadside wifi networks," in *INFOCOM*, 2008, pp. 718–725.

[42] D. Niculescu and B. Nath, "Vor base stations for indoor 802.11 positioning," in *MobiCom*, 2004, pp. 58–69.

[43] K. Joshi, S. Hong, and S. Katti, "Pinpoint: localizing interfering radios," in *NSDI*, 2013, pp. 241–253.

[44] Y. Nishimura, N. Imai, and K. Yoshihara, "A proposal on direction estimation between devices using acoustic waves," in *MobiQuitous*, 2012, pp. 25–36.

[45] 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 *MobiCom*, Maui, HI, USA, 2014, pp. 91–102.

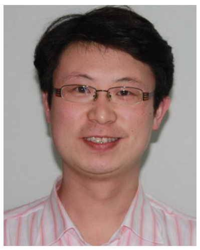


Wenchao Huang received the B.S. and Ph.D degrees in computer science from University of Science and Technology of China in 2005 and 2011, respectively. He is currently an associate professor in School of Computer Science and Technology, University of Science and Technology of China. His current research interests include mobile computing, information security, trusted computing and formal methods.

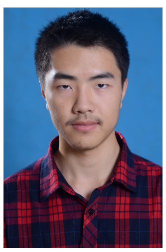
Xiang-Yang Li is a professor at School of Computer Science and Technology, University of Science and Technology of China. He was elected as IEEE Fellow (class 2015), and ACM Distinguished Scientist (class 2015) in 2014. Dr. Li received MS (2000) and PhD (2001) degree at Department of Computer Science from University of Illinois at Urbana-Champaign, a Bachelor degree at Department of Computer Science and a Bachelor degree at Department of Business Management from Tsinghua University, P.R. China, both in 1995. His research interests include mobile computing, cyber physical systems, wireless networks, security and privacy, and algorithms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2017.2679600, IEEE Journal on Selected Areas in Communications

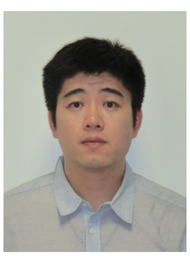
Yan Xiong received the B.S., M.S., and Ph.D degrees from University of Science and Technology of China in 1983, 1986 and 1990 respectively. He is a professor in School of Computer Science and Technology, University of Science and Technology of China. His main research interests include distributed processing,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



Panlong Yang received his B.S. degree, M.S. degree, and Ph.D. degree in communication and information system from Nanjing Institute of Communication Engineering, China, in 1999, 2002, and 2005 respectively. Dr. Yang is now an associate professor in the PLA University of Science and Technology. Dr. Yang has published more than 50 papers in the areas of mobile ad hoc networks, wireless mesh networks and wireless sensor networks.



Yiqing Hu is a Ph.D. candidate in school of Computer Science and Technology, University of Science and Technology of China, His main research interests include indoor localization, smart sensing and data mining.



Xufei Mao received the Ph.D. degree in Computer Science from Illinois Institute of Technology, Chicago in 2010. He received the MS degree (2003) in Computer Science and the Bachelor degree (1999) in Computer Science at Northeastern University and Shenyang University of Technology respectively. He is with Dongguan University of Technology. His research interests span wireless ad-hoc networks, wireless sensor networks, pervasive computing, mobile cloud computing and game theory.

Fuyou Miao received his Ph.D of computer science from University of Science and Technology of China in 2003. He is an associate professor in the School of Computer Science and Technology, University of Science and Technology of China. His research interests include applied cryptography, trusted computing and mobile computing.

Baohua Zhao received His M.S. Degree from Beijing University of Technology in 2016. He is a director in Computing Technology and Applications Research Institute, Global Energy Interconnection Research Institute. His main research interests include trusted computing, information security and computer technology.

Jumin Zhao is a professor at college of information engineering of Taiyuan University of Technology (TUT), Shanxi, China. Dr. Zhao received Ph.D degree (2008) at Electronic Information College from TUT, a Bachelor degree at Electronic Information College and a Bachelor degree at Foreign Language college from TUT both in Jul. 1998. Her research interests include the signal processing, wireless networks, mobile computing, cyber-physical system, and algorithms.