WalkieLokie: Sensing Relative Positions of Surrounding Presenters by Acoustic Signals

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

ABSTRACT

In this paper, we propose and implement WalkieLokie, a novel acoustic-based relative positioning system. WalkieLokie facilitates a multitude of Augmented Reality (AR) applications: users with smart devices can passively acquire surrounding information in real time, similar to the commercial AR system Wikitude; the surrounding presenters, who want to share information or introduce themselves, can actively launch the function on demand. The key rational of WalkieLokie is that a user can perceive a series of spatial-related acoustic signals emitted from a presenter, which depicts the relation position between the user and the presenter. 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. We design a novel algorithm to estimate the position and signal processing methods to support accurate positioning. The experiment results show that the mean error of ranging and direction estimation is 0.63m and 2.46 degrees respectively. Extensive experiments conducted in noisy environments validate the robustness of WalkieLokie.

ACM Classification Keywords

H.5.1. INFORMATION INTERFACES AND PRESENTATION (e.g., HCI): Artificial, augmented, and virtual realities

Author Keywords

Augmented reality, acoustic signaling, ranging, direction finding.

INTRODUCTION

With[1] the rapid development of smart phones and wearable devices, attractive Augmented Reality (AR) devices and apps have been developed, *e.g.*, Microsoft HoloLens [4], Google Glass [2], Sky Map [3], Wikitude [6], and Augmented Car Finder [1]. The AR technology can enhance a user's experience when traveling or visiting museums, where the user retrieves digital information of surrounding objects via computer vision or other sensors (GPS, inertial sensors). The vision-based systems [18] take an image from the camera and match the image with the ones in the annotated database. Wikitude displays annotations about surrounding objects by using GPS and inertial sensors. A demanding requirement of recent advances is to enable more kinds of AR applications. 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 [25]; or a person shares her/his virtual business card [10, 12] with people walking around in a party. The virtual shopping guide and the person who shares the business card are denoted as presenters who want to be recognized and localized by surrounding people.

Existing techniques are insufficient to enable these applications. The computer vision requires large CPU or communication overhead [18] to implement passive and real-time object (*i.e.*, the presenter) recognition. Moreover, for the business card sharing, presenters may be unwilling to share their photos in a public database for image matching, but may only launch the presenting mode in special cases or places. The popular AR system Wikitude, which uses GPS for localization, cannot be used in indoor environment. Some indoor localization systems [23, 38] require special-purpose infrastructure or hardware to support sub-meter accuracy. Other systems that leverage existing infrastructures (*e.g.*, WiFi [8, 9, 14, 22, 33, 39], Visible Light [21, 40]) are widely studied. However, the accuracy still depends on the density of deployed infrastructures in a specific location. There are other schemes on direction finding (e.g., Swadloon [15, 16]) and ranging (e.g., BeepBeep [30]). Swadloon cannot obtain distance from the object. Though BeepBeep can be added for calculating distance, it requires that the presenter has a device with rich functions, such as broadcasting and receiving acoustic signals, wireless communication and data processing. Overall, a system should satisfy the following conditions [27, 41] to enable the applications: **a) Real-time surfing:** The users *passively* receive the surrounding information in *real time* [11, 44]. **b) On-demand annotation:** The surrounding presenters (*e.g.*, shopping guide, business card sharer) *actively* annotate themselves on demand, while others without the intent do not share any information. **c) Location independence:** The system works in most places where the supplementary infrastructure [37] is not required to be deployed in a specific location in advance. **d) Economical and practical devices:** Users or presenters prefer to use a device they already have or a cheap device [17] to launch the AR function.

[1] 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 Tsinghua University, 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.

UbiComp'16, September 12-16, 2016, Heidelberg, Germany

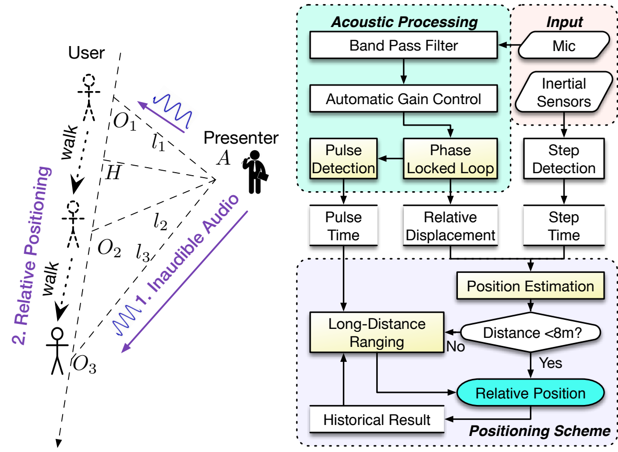
DOI: http://dx.doi.org/10.1145/2971648.2971655

We propose and implement WalkieLokie, which calculates the *relative position* from a user to a presenter. WalkieLokie can serve as a base system for various upper-level applications, including the new AR applications. The only requirement of WalkieLokie is that the presenter is equipped with a *dummy* speaker (*i.e.*, an acoustic source) for broadcasting audio. Hence, the user can passively receive the audio and send intensive computations upstream for analysis, which infers the real-time position. On the other hand, the presenter can launch the speaker on-demand in any places, where the position can be obtained by a user only if the user is in sight (<20m). The dummy speaker is cheap and simple that it does not require any other feature, *e.g.*, audio recording, communication or computation. Hence, the speaker is available for the presenter in many forms, such as a cheap accessory in smart clothes (*e.g.*, Project Jacquard by Google [5]), or even a loudspeaker originally for sales promotion in a shopping mall. Moreover, 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.

Our work is based on the observation that when a user walks, the distance between the presenter and the user changes, and the pattern of *displacement* (variance of distance) is associated with the relative position. In other words, by letting a device receive and analyze the signal (audio signal mixed with non-audible signal) broadcast by an object (speaker), we are able to track the displacement and further compute the relative positions accurately and efficiently, i.e., finishing both ranging and direction estimation simultaneously.

There are several challenges in realizing our work. First, since the displacement is relatively small when a user walks for only a few steps, it is challenging to precisely track the displacement. We propose several components to process the acoustic signal and finally design a Phase Locked Loop to track the phase of the signal which corresponds to the displacement. Therefore, by leveraging the displacement tracking mechanism, WalkieLokie calculates the relative positions with sub-meter accuracy when a user is in the vicinity of the speaker (*i.e.*, $< 8m$). Hence, WalkieLokie is accurate enough for business card sharing. Second, when the user is far from the speaker (*i.e.*, $8 \sim 20m$), even a tiny error in the displacement measurement may cause large errors in ranging (note that the direction finding is still accurate). As long-distance positioning is required by virtual shopping guide, we additionally design the long-distance ranging mechanism by leveraging the historical results of relative positions. To leverage the results, the mechanism adds periodical pulses in the acoustic signal, where the pulses take narrow bandwidth (*i.e.*,460Hz), and have ignorable impacts on the accuracy of displacement tracking.

WalkieLokie also addresses a number of practical issues. We design an enhanced algorithm in case that the user turns the walking direction. WalkieLokie deals with multipath effects on pulse detection by leveraging the tracked phase in PLL. We also propose a quick and light-weighted method to calibrate the clock drift in different speakers by leveraging the property in second-order PLL. We optimize WalkieLokie to support more speakers in a room, and it is robust in noisy environments, *e.g.*, the shopping mall.



(a) Brief Example. (b) Architecture of WalkieLokie.

Figure 1: Example of coarse-grained relative positioning, and architecture of WalkieLokie.

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 error of ranging and direction estimation is $0.63m$ and 2.45^o. 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 positioning system. It is location independent and economically practical for both users and presenters, and it supports more kinds of AR applications.
- We design a novel algorithm for relative positioning by leveraging only the received acoustic signal without any additional information from the presenters.
- 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 first present the overview of WalkieLokie. Then, we propose the position estimation based on displacement tracking and displacement tracking method. Next, we give the details of the long-distance ranging. We report our extensive experimental results and review some related work. Finally, we conclude the paper.

OVERVIEW

Problem Description

In WalkieLokie, the presenter is equipped with a dummy speaker. The dummy speaker merely broadcasts solely inaudible audio. The user holds a smart device in open air or attach the smart device to the body. The smart device has a microphone and inertial sensors (*i.e.*, compass, accelerometer, gyroscope), which are common components in almost all smart devices. Smart device cannot be put in the pocket. Otherwise, the received acoustic signal is weak to be processed.

WalkieLokie calculates the *relative position* between a user and a presenter. For example, in Figure 1a, a user walks and steps on O_1, O_2, and O_3, and a presenter stands on 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.*, $= \angle AO_1H$). Note that 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.

Intuitive Solution

The key insight of relative positioning is that when a user walks along a line, the pattern of displacements from the user to a dummy speaker is related to the relative position directly. In Figure 1a, 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. So, 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.

Main Technical Issues

From the above example, we solve the following technical issues: **1) Formal solution of relative positioning.** Given real-time relative displacement, we calculate the fine-grained relative position, instead of the coarse-grained position in the intuition solution. **2) Tracking relative displacement.** We track the relative displacement before relative positioning. **3) Long-distance ranging.** When $|\overline{AH}|$ becomes much longer, $|d_2 - d_1|$ is much smaller and it is insufficient for ranging. Therefore, we propose a novel acoustic processing scheme to support long-distance ranging. Note that the accuracy of direction finding is not much affected, for the accuracy is determined by the precision of measured d_1.

Architecture

To solve the technical issues, we divide WalkieLokie into 3 main components in Figure 1b: 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. The gyroscope is used to calculate the angle of user's rotation, when the user turns direction.

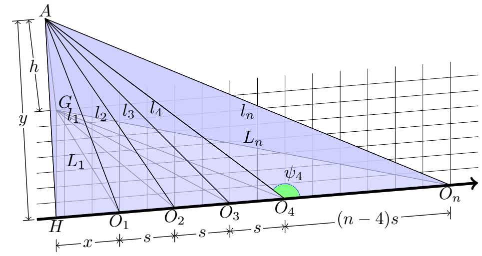


Figure 2: Positioning when the user walks along a line.

Acoustic signal processing: The component generates intermediate results for the positioning scheme. One result is **relative displacement**, which is calculated by using the Band Pass Filter (BPF), Automatic Gain Control (AGC), and judiciously-designed Phase Locked Loop (PLL). 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 scheme estimates position by using the relative displacement and the time when the user steps on the ground. If the computed distance is short ($< 8m$), it is accepted as a valid result. Otherwise, the scheme leverages historical results of relative positioning together with the receiving time of periodical pulses to infer the distance.

ESTIMATION OF RELATIVE POSITIONS

In this section, we propose the method on distance and direction estimation from smart device to speaker, *i.e.*, the relative position, given the measured displacement as input.

Positioning When User Walks Naturally in a Straight Line

To estimate the distance, we first consider a simple scenario when a user starts walking from O_1 and steps at $O_2, O_3, \ldots, O_n$, shown in Figure 2. A is the position of the speaker. Denote height difference between the speaker and the smart device as $h = |\overline{AG}|$. Assume the stride length is $s = \overline{O_iO_{i+1}}$. The other inputs are the displacements of all the steps, *i.e.*, $d_i = l_i - l_{i+1}$ for the step $\overline{O_iO_{i+1}}$. Observe that the distance from the speaker to $\overline{O_iO_{i+1}}$ is constant $y = |\overline{AH}|$, where $\overline{AH} \perp \overline{O_1O_n}$. Hence, we first estimate $x = |\overline{HO_1}|$ and y from those inputs and then estimate the position at each step point O_i according to x and y.

Then, 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

$$l_i' = \sqrt{y^2 + (x + (i-1)s)^2} \quad (1)$$

$$e_i = l_i' - l_{i+1}' - d_i \quad (2)$$

For n displacements $d_1, d_2, \ldots, d_n$, x and y can be solved from above n equations by $(x, y) = \arg\min_{x,y} \sum_{i=1}^{n} e_i^2$. Here we use the Newton's Method [26] to reduce computation overhead. Finally, the horizontal distance $L_i = |\overline{GO_i}|$ and direction $\psi_i' = \angle GO_iO_n$ are calculated by using x, y.

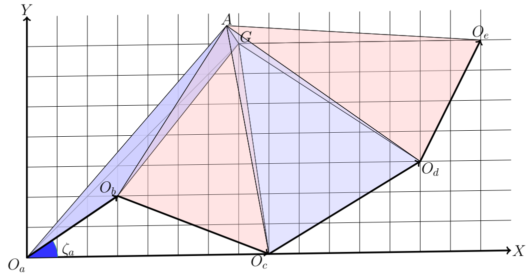


Figure 3: Positioning when the user walks and turns.

Synthesizing When User Turns Direction

We calculate the relative position 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 3. We use the calculated displacements in this case. We also use the pedometer to estimate the linear length n_as, n_bs, n_cs, n_ds, where n_a is the step count when the user walks from O_a to O_b.

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 [33]. 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. Specifically 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 [7], 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.

Calculating position: We calculate the relative position $G(g_x, g_y)$, which is the projection of acoustic speaker A at a horizontal plane. Assume that O_a is at $(0,0)$. O_c is at the position $(c_x, c_y) = (n_as\cos(\zeta_a) + n_bs\cos(\zeta_b), n_as\sin(\zeta_a) + n_bs\sin(\zeta_b))$, and so forth. Given the calculated displacement $d_{c_1}, d_{c_2}, \ldots, d_{c_{n_c}}$, similar to Eq. (1), Eq. (2), the distance from each stride point to G is

$$l_{c_i} = \sqrt{[c_x + (i-1)s\cos(\zeta_c)]^2 + [c_y + (i-1)s\sin(\zeta_c)]^2 + h^2} \tag{3}$$

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

$$e_{c,i} = l_{c_i} - l_{c_{i+1}} - d_{c_i} \tag{4}$$

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

$$(g_x, g_y) = \underset{g_x, g_y}{\arg\min} \sum_{i\in\{a,b,c,d,e\}} \sum_{j=1}^{n_c-1} e_{i,j}^2 \tag{5}$$

TRACKING DISPLACEMENT

In this section, we design the acoustic wave emitted by the dummy speaker and propose the scheme of inferring relative displacement according to the received acoustic wave.

Brief Design of the Acoustic Wave

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

$$s(t) = \begin{cases} s_1(t), & kT_2 \le t < kT_2 + T_1 \\ s_2(t), & kT_2 + T_1 \le t < (k+1)T_2 \end{cases} \tag{6}$$

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.

Intuitively, $s_1(t)$ is a sine wave, and the phase of the corresponding received signal $r_1(t)$ is changed when the distance changes. We prove in the following subsection that the phase of $r_1(t)$ is proportional to the relative displacement. So by tracking the phase of $r_1(t)$, the displacement is tracked. Note that $s_2(t)$ is not only used for pulse detection, but also capable of displacement tracking, like $s_1(t)$. As a result, the performance of displacement measurement is rarely affected by $s_2(t)$.

Phase & Displacement

In order to track the displacement, we define $s_1(t)$ as follows:

$$s_1(t) = \cos(2\pi ft) \tag{7}$$

where f is the frequency. To make the audio inaudible and to make the broadcasting frequency available by commercial speaker, we set $17000Hz < f < 24000Hz$.

The receiving signal $r_1(t)$ has a phase shift ϕ caused by Doppler effects, *i.e.*, $r_1(t) = \cos(2\pi ft + \phi)$. Then, in Figure 1a, the displacement [16] is

$$d = l_1 - l_2 = \frac{v_a}{2\pi f}(\phi_2 - \phi_1) \tag{8}$$

where ϕ_1 and ϕ_2 is the calculated phase at O_1 and O_2 respectively and v_a is the travelling speed of acoustic wave. Hence, the displacement d is proportional to the variance of the phase of the signal $(\phi_2 - \phi_1)$.

Preprocessing Received Signal

Before tracking phase ϕ from $r_1(t)$, we have to preprocess the received signal. For the sent signal $s_1(t)$, the actual received signal $r_{raw}(t)$ does not equal $r_1(t)$. Its amplitude $A(t)$ always changes and it is also mixed with noises $\sigma(t)$, which is affected by the surrounding environment. We denote $r_{raw}(t) = A(t)\cos(2\pi ft + \phi(t)) + \sigma(t)$. Hence, we need to normalize $A(t)$ and attenuate $\sigma(t)$ before tracking the phase $\phi(t)$.

To attenuate the noise $\sigma(t)$, we let $r_{raw}(t)$ pass through a Band Pass Filter (BPF). The processed signal $r_{filter} \approx A(t)\cos(2\pi ft + \phi(t))$. r_{filter} is then processed by Automatic Gain Control (AGC) [34]. After that, $A(t)$ is normalized and the signal can be regarded as $r_1(t)$ [16].

Tracking the Phase

We adopt the second-order Phase Locked Loop (PLL) to track the phase for inferring the displacement. 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 platform.

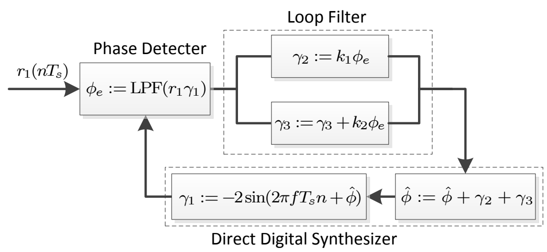


Figure 4: Design of the Second-Order Phase Locked Loop.

We show our design of PLL in Figure 4. 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.

LONG-DISTANCE RANGING

In this section, we propose long-distance ranging mechanism to avoid possible accumulated errors on ranging. Specifically, as mentioned earlier, the accuracy of ranging by Eq. (1) reduces 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. To solve this problem, we calculate the distance by adding periodical pulses $s_2(t)$ in the acoustic signal and use the receiving time of pulses as an additional input for ranging.

Ranging in Longer Distances

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. So, 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:

$$l_j = (\tau'_j - \tau_j)v_a = l + (\tau'_j - \tau' - jT)v_a \tag{9}$$

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

Pulse Modulation and Detection

Design goals:

To design the pulses $s_2(t)$ and the pulse detection algorithm, several problems should be addressed:

- **Bandwidth:** Each speaker should be allocated with less acoustic bandwidth in order to support more presenters. Hence, $s_1(t)$ and $s_2(t)$ should share the same frequency band; otherwise additional bandwidth for $s_2(t)$ is needed. Moreover, bandwidth of $s_2(t)$ needs to be narrow. However, it is challenging that $s_2(t)$ should occupy more bandwidth if it can be successfully detected.
- **Effects on displacement tracking:** $s_2(t)$ can also be used for displacement tracking by PLL. Otherwise, PLL will lose phase locks when processing $s_2(t)$.

Brief design

Modulation: Based on these requirements, we design $s_2(t)$:

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

where we construct pulses starting at $\tau_1, \dots, \tau_i$, and the duration of each pulse is T_p.

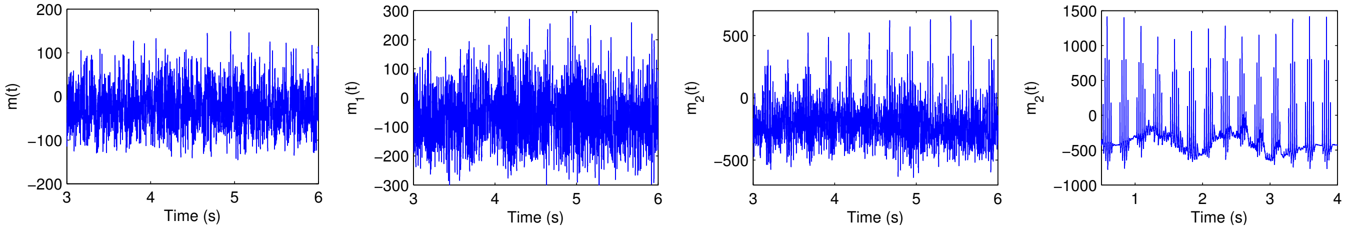
Basic Detection mechanism: We propose to detect the receiving time $\tau'_i = \tau_i + t_l$ of the ith pulse by leveraging the component $s_3(t)$. Assuming the locked phase by PLL is ϕ_r before the pulse starts, the expected pulse is $\tilde{r}(t) = \cos(2\pi f t + \phi_r + \pi \sin \frac{\pi(t-\tau'_i)}{T_p})$. 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.

Bandwidth: 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 $T_3 = 0.03s$. Since the bandwidth of the pulse is about $\frac{\pi}{T_p}$ [35], we set $T_p = 0.007s$ so that the bandwidth is about 460Hz. As the minimum frequency is 17000Hz when the acoustic is non-audible, 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.

Pulse detection

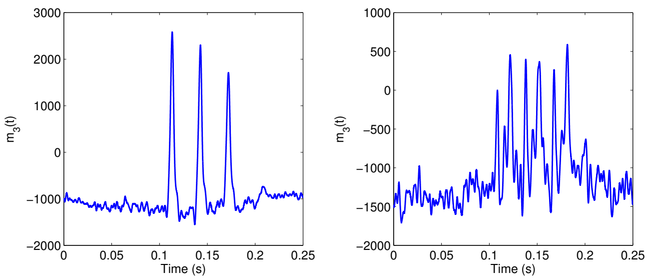
The basic pulse detection mechanism suffers in case of interferences, *e.g.*, noises or multipath effects. The moving of the smart device also influences the performance of pulse detection. Facing this problem, we improve the pulse detection mechanism by the following strategies:

Dealing with noisy environment or phone-moving case: 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 5a. 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 5b, 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 5c, where the peaks can be easily detected. In case when the phone is moving as shown in Figure 5d, the peaks are also very clear.



(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

Figure 5: Detection of the arrival time of the pulse.



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

Figure 6: Pulse detection in case of multipath effects.

Dealing with multipath effects: In the evaluation section, we find that the detected pulse is prone to be affected by multipath effects, when the phone is static. 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 make the detected time of pulses more clear. In Figure 6a when there is no multipath effect, there are 3 pulses in a period T_2. However, in Figure 6b, 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 $(t_a, t_b) = \underset{t_a \in T_a, t_b \in T_b}{\arg\min} |(t_a - t_b)v_a - d|$.

Supporting more concurrent presenters:
We find that further optimizations can be made to support more concurrent presenters as follows: *a) Virtual business card sharing:* In this case, users are usually close to each other, and we can choose to narrow the bandwidth of pulses for long-distance ranging, which reduces accuracy of long-distance positioning but increases the number of presenters that WalkieLokie can support. *b) Virtual shopping guide:* We suggest that if there is requirement of more shopping guides, we can use only a few speakers (pre-deployed by the WalkieLokie group) for normal indoor localization, instead of just relative positioning. Our further evaluations 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.

PERFORMANCE EVALUATION

In this section, 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 operation 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.

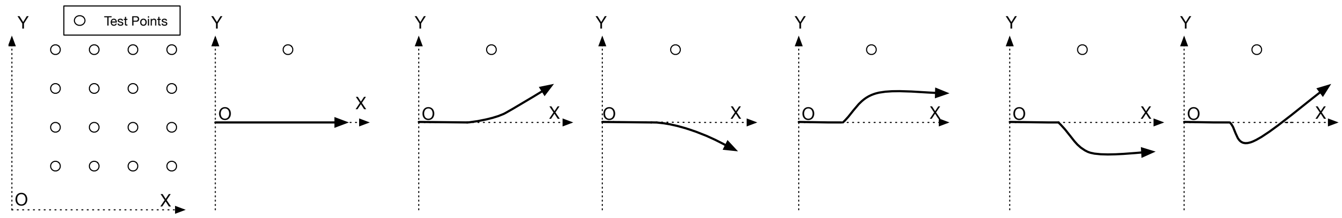
Position Estimation

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.

Positions

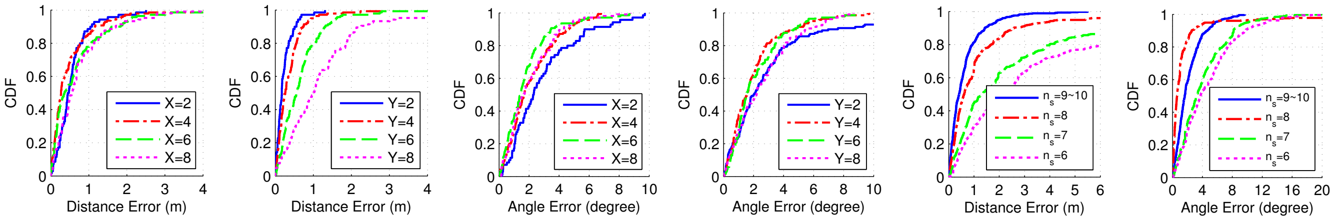
We make evaluations in an empty room as shown in Figure 7a, where the speaker is placed at 16 different positions which are uniformly distributed in a square, *i.e.*, (X,Y) where $X \in \{2,4,6,8\}$, and $Y \in \{2,4,6,8\}$. Here, the empty room is large ($> 1000m^2$) where the multipath effects on measuring 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 8a, the accuracy of distance estimation is very close for different X. We further study the distribution of large errors in Figure 8b. 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 direction estimation, it is still very accurate when the X or Y increases in Figure 8c, 8d. As a



(a) Test Points (b) Straight line (c) Turn left (d) Turn right (e) Left & forward (f) Right & forward (g) Random

Figure 7: Experiment configuration in an empty room: test points (a) and different walking paths (b-f).



(a) Ranging (b) Ranging (c) Direction (d) Direction (e) Ranging (f) Direction

Figure 8: 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).

total, the mean of ranging and angle error is $0.63m$ and 2.46° respectively.

Number of Steps

The accuracy of the position estimation depends on number of walking steps. We compare the results when the user walks for smaller number of steps n_s in Figure 8e, 8f. 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 8f. 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 [16], where the direction is estimated via 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.

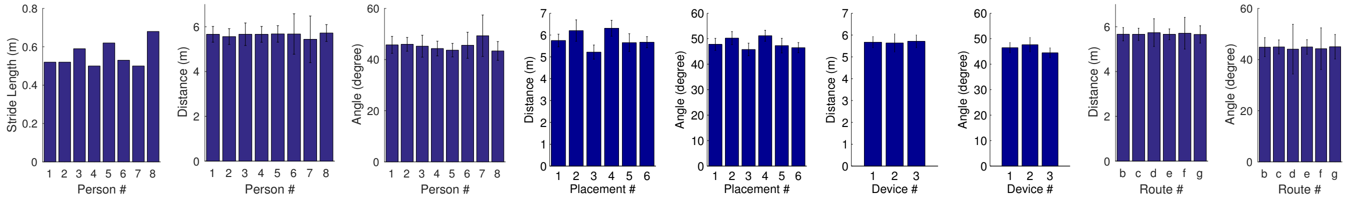
Users

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 9: The standard deviations (std) of the ranging and direction are small for most users. In Figure 9a, 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.

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 9d, 9e 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 constant factor (*i.e.*, 1.22), for the environment has limited effect 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).



(a) Stride Length (b) Ranging (c) Direction (d) Ranging (e) Direction (f) Ranging (g) Direction (h) Ranging (i) Direction

Figure 9: 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).

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 9f, 9g. 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, 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 4 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.

Turning Directions

We also evaluate the performance when user turns directions. In Figure 7b$\sim$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 9h,9i, 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 is $0.39m$ 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) is 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) is 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 2. According Figure 8a,8b, 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 presenter, the positioning result becomes less important and more coarse-grained.

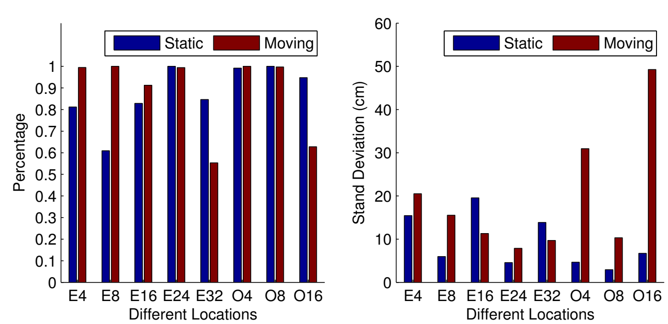
Environments

We compare the accuracy of position estimation in the empty room and at different locations in the office. We find that it shows the similar results. We further evaluate the effects in a shopping mall in the latter subsection.

Long-Distance Positioning

Pulse Detection

In Figure 10, 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 $l_e \approx 0.8cm$.



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

Figure 10: (a) Percentage of successful experiments at different locations (b) standard deviation.

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 10a, the successful detection rate is above 80% for most cases when the phone is static. When the phone is moving, the performance is good as well if the distance is within $24m$ and $8m$ in the empty room and 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. So, 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 10b, we show the standard deviation of results in case of successful detection. The std in most cases are around $10cm$ expect that the std is $30.9cm$ and $49.2cm$ when the phone is moving at O4 and O16 respectively.

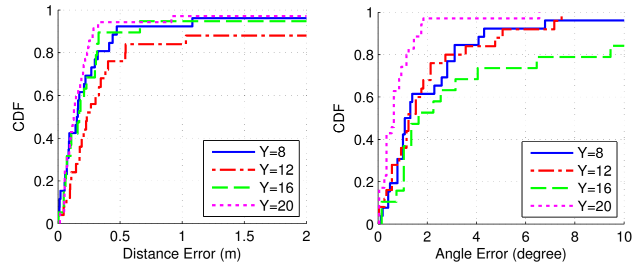
Long-Distance Ranging and Direction Estimation

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 performing 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.

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 11a, 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 $2m$. It is caused by the multipath effects in pulse detection.



(a) Ranging (b) Direction Estimation

Figure 11: Positioning after 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 10a, 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.

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 12, 13a.

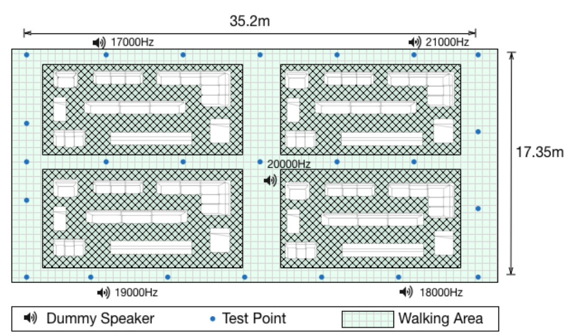
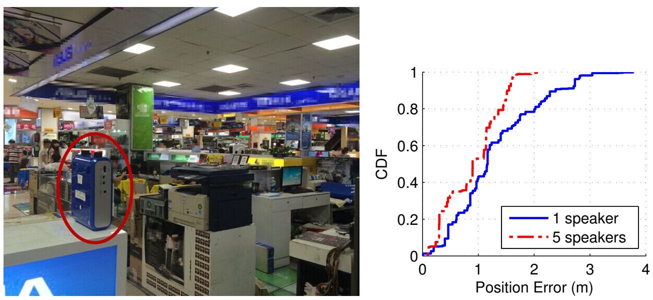


Figure 12: Map of the shopping mall.

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 12, and put 5 normal dummy speakers in this area. Each speaker broadcasts signals at different central frequencies, which are inaudible 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 1b. 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.



(a) Shopping mall. (b) Position Errors

Figure 13: (a) Shopping mall and the dummy speaker. (b) Result of relative positioning (using 1 speaker), and absolute positioning (using 5 speakers).

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. The reason of this diversity may be caused by several facts: anchor positions, quality of different anchor speakers, etc. We leave the study on configuration of speakers in our future work. Totally, the average coverage per speaker is 38%, which is about $222m^2$ in our specific area.

We show the relative position errors when using one speaker in Figure 13b. Note that we only evaluate the accuracy of the relative position where the starting point is covered by the signals of the speaker. Though we can still estimate position according to 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. So, the accuracy is better when using multiple signals for positioning.

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 takes 0.09s, 0.12s, 0.05s respectively. For the computer-vision based annotation system, a convincing way [18] 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). 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.

RELATED WORK

Ranging

There have been many ranging systems [13, 23, 24, 31] used for localization. They achieve considerable accuracy of ranging, but require special hardwares for synchronization purpose. The general idea is that the sender records sending time of signal which is used for ranging, while the receiver detects the arrival time of the signal. Each individuals calculate the sending time or arrival time independently without referring any time information on other devices. Hence, synchronization among devices is needed. In Bat System [13], the base-station uses radio channel and communications for synchronization. Cricket [31] 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 [23] uses RF signals to synchronize all the acoustic anchors, the location can be obtained according to the differences of the receiving time by the phone. BeepBeep [30] 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. Compared with BeepBeep, WalkieLokie uses dummy speaker to implement ranging. It does not need any special hardwares or additional communication channels for exchanging the synchronization information. Besides, WalkieLokie can also perform direction finding. The disadvantage is that the computation overhead is higher due to the continuous signal processing, and WalkieLokie achieves sub-meter accuracy in ranging, while BeepBeep can achieve centimeter-level accuracy. For indoor localization, the difference is that previous systems are only based on ranging results of anchors which requires multiple speakers (≥ 3), while WalkieLokie also implements direction estimation from phone to speaker and only one speaker is sufficient for localization.

Direction Estimation

Most direction estimation systems require specialized hardwares, which use directional antenna [20, 28, 36] or antenna array [19, 36, 38]. For example, by rotating the beam of directional antenna, a receiver can pinpoint the direction of the AP as the direction that provides the highest received strength [36]. For the antenna array [19, 36, 38], the receiving time of the signal by each antenna is different, and magnitude of the difference corresponds to angle of the arrival signal.

There have been proposals without requirement of specialized hardwares as well. [43] emulates the functionality of a directional antenna by rotating the phone around the user's body, to locate outdoor APs. [32] 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 [29] or shaking [16] the phone. [42] calculates direction by head nodding or shaking using smart glasses. They are based on different frequency shift when the phone are moving at different directions. Compares to [16, 29], WalkieLokie makes further steps that a user can obtain direction without any additional actions on the phone so that s/he can get the real-time direction while walking. Recall that Swadloon [16] 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. If Swadloon is applied to indoor localization, at least 3 sources are required as beacons for triangulation, while at least 1 source is required as beacon, if WalkieLokie is applied, since WalkieLokie directly calculates both direction and distance from a smart device to a source.

CONCLUSION

We propose and implement WalkieLokie, a novel system that calculates relative position from a user to a presenter. WalkieLokie can be launched as long as the object is in sight ($< 20m$), without the limitation of applying the system in places where some infrastructure has been deployed. Moreover, the object only needs a dummy speaker that emits acoustic signals at non-audible frequency. So COTS speakers can serve as cheap positioning devices. Hence, WalkieLokie can be applied in various Augmented Reality applications, by which the user can acquire the relative position of the surrounding presenters equipped with a dummy speaker.

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, and Anhui Provincial Natural Science Foundation, No.1508085SQF215.

REFERENCES

1. 2016. Augmented Car Finder. (2016). `http://www.augmentedworks.com/app/find-your-car-with-ar-augmented-car-finder/`
2. 2016a. Google Glass. `http://www.google.com/glass/start/`. (2016).
3. 2016b. Google Sky Map. `https://play.google.com/store/apps/details?id=com.google.android.stardroid`. (2016).
4. 2016. Microsoft HoloLens. `https://www.microsoft.com/microsoft-hololens`. (2016).
5. 2016c. Project Jacquard by Google. `https://www.google.com/atap/project-jacquard/`. (2016).
6. 2016. Wikitude. `http://www.wikitude.com`. (2016).
7. Simon L. Altman. 1986. *Rotations, Quaternions, and Double Groups*. Oxford University Press.
8. Paramvir Bahl and Venkata N Padmanabhan. 2000. RADAR: An in-building RF-based user location and tracking system. In *INFOCOM*, Vol. 2. Ieee, 775–784.
9. Jacob T Biehl, Matthew Cooper, Gerry Filby, and Sven Kratz. 2014. Loco: a ready-to-deploy framework for efficient room localization using wi-fi. In *UbiComp*. ACM, 183–187.
10. Antonio Cuervo. 2014. Smart business card. (2014).
11. Davide Dardari, Pau Closas, and Petar M Djuric. 2015. Indoor tracking: Theory, methods, and technologies. *TVT* 64, 4 (2015), 1263–1278.
12. Nimesh R Desai. 1996. Electronic business card system. (1996).
13. Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, and Paul Webster. 1999. The Anatomy of a Context-Aware Application. In *MOBICOM*. 187–197.
14. Suining He, S-H Gary Chan, Lei Yu, and Ning Liu. 2015. Calibration-free fusion of step counter and wireless fingerprints for indoor localization. In *UbiComp*. ACM, 897–908.
15. Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, XuFei Mao, Panlong Yang, Yunhao Liu, and Xingfu Wang. 2015. Swadloon: Direction Finding and Indoor Localization Using Acoustic Signal by Shaking Smartphones. *TMC* 14, 10 (2015), 2145–2157.
16. Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, XuFei Mao, Panlong Yang, and Yunhao Liu. 2014. Shake

and Walk: Acoustic Direction Finding and Fine-grained Indoor Localization Using Smartphones. In *INFOCOM*.

17. Takahiro Iwase and Ryosuke Shibasaki. 2013. Infra-free indoor positioning using only smartphone sensors. In *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*. IEEE, 1–8.

18. Puneet Jain, Justin Manweiler, and Romit Roy Choudhury. 2015. OverLay: Practical Mobile Augmented Reality. In *MobiSys*. 331–344.

19. Kiran Joshi, Steven Hong, and Sachin Katti. 2013. PinPoint: localizing interfering radios. In *NSDI*. 241–253.

20. Myungsik Kim and Nak Young Chong. 2009. Direction Sensing RFID Reader for Mobile Robot Navigation. *TASAE* (2009), 44–54.

21. Liqun Li, Pan Hu, Chunyi Peng, Guobin Shen, and Feng Zhao. 2014. Epsilon: A visible light based positioning system. In *NSDI*. 331–343.

22. Hongbo Liu, Yu Gan, Jie Yang, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye. 2012. Push the limit of WiFi based localization for smartphones. In *MobiCom*.

23. Kaikai Liu, Xinxin Liu, and Xiaolin Li. 2013a. Guoguo: enabling fine-grained indoor localization via smartphone. In *MobiSys*. 235–248.

24. Kaikai Liu, Xinxin Liu, Lulu Xie, and Xiaolin Li. 2013b. Towards accurate acoustic localization on a smartphone. In *INFOCOM*. 495–499.

25. Mark Lorenzen and Lars Frederiksen. 2008. Why do cultural industries cluster? Localization, urbanization, products and projects. *Creative cities, cultural clusters and local economic development* (2008), 155–179.

26. K. Madsen, H. B. Nielsen, O. Tingleff, and Mathematical Modelling. 2004. IMM METHODS FOR NON-LINEAR LEAST SQUARES PROBLEMS. Technical University of Denmark, DTU. (2004).

27. Halgurd S Maghdid, Ihsan Alshahib Lami, Kayhan Zrar Ghafoor, and Jaime Lloret. 2016. Seamless Outdoors-Indoors Localization Solutions on Smartphones: Implementation and Challenges. *CSUR* 48, 4 (2016), 53.

28. Dragoş Niculescu and Badri Nath. 2004. VOR base stations for indoor 802.11 positioning. In *MobiCom*. 58–69.

29. Yasutaka Nishimura, Naoki Imai, and Kiyohito Yoshihara. 2012. A Proposal on Direction Estimation between Devices Using Acoustic Waves. In *MobiQuitous*. 25–36.

30. Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. 2007. BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices. In *SenSys*. 1–14.

31. Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. 2000. The Cricket location-support system. In *MobiCom*. 32–43.

32. Jian Qiu, David Chu, Xiangying Meng, and Thomas Moscibroda. 2011. On the feasibility of real-time phone-to-phone 3D localization. In *SenSys*. 190–203.

33. Anshul Rai, Krishna Kant Chintalapudi, Venkata N. Padmanabhan, and Rijurekha Sen. 2012. Zee: zero-effort crowdsourcing for indoor localization. In *MobiCom*. 293–304.

34. M. Rice. 2008. *Digital Communications: A Discrete-Time Approach*. Prentice Hall.

35. O. P. Sahu and Anil K. Gupta. 2008. Measurement of Distance and Medium Velocity Using Frequency-Modulated Sound/Ultrasound. *IEEE T. Instrumentation and Measurement* 57, 4 (2008), 838–842.

36. A.P. Subramanian, P. Deshpande, Jie Gaojgao, and S.R. Das. 2008. Drive-By Localization of Roadside WiFi Networks. In *INFOCOM*. 718–725.

37. Agoston Torok, Akos Nagy, László Kováts, and Péter Pach. 2014. Drear-towards infrastructure-free indoor localization via dead-reckoning enhanced with activity recognition. In *Next Generation Mobile Apps, Services and Technologies (NGMAST), 2014 Eighth International Conference on*. IEEE, 106–111.

38. Jie Xiong and Kyle Jamieson. 2013. ArrayTrack: A Fine-grained Indoor Location System. In *NSDI*. 71–84.

39. Han Xu, Zheng Yang, Zimu Zhou, Longfei Shangguan, Ke Yi, and Yunhao Liu. 2015a. Enhancing wifi-based localization with visual clues. In *UbiComp*. ACM, 963–974.

40. Qiang Xu, Rong Zheng, and Steve Hranilovic. 2015b. IDyLL: indoor localization using inertial and light sensors on smartphones. In *UbiComp*. ACM, 307–318.

41. Zheng Yang, Chenshu Wu, Zimu Zhou, Xinglin Zhang, Xu Wang, and Yunhao Liu. 2015. Mobility Increases Localizability: A Survey on Wireless Indoor Localization using Inertial Sensors. *ACM Comput. Surv.* 47, 3 (2015), 54.

42. Lan Zhang, Xiang-Yang Li, Wenchao Huang, Kebin Liu, Shuwei Zong, Xuesi Jian, Puchun Feng, Taeho Jung, and Yunhao Liu. 2014. It starts with iGaze: visual attention driven networking with smart glasses. In *MobiCom*. Maui, HI, USA, 91–102.

43. Zengbin Zhang, Xia Zhou, Weile Zhang, Yuanyang Zhang, Gang Wang, Ben Y. Zhao, and Haitao Zheng. 2011. I am the antenna: accurate outdoor AP location using smartphones. In *MobiCom*. 109–120.

44. Xiaojun Zhu, Qun Li, and Guihai Chen. 2013. APT: Accurate outdoor pedestrian tracking with smartphones. In *INFOCOM*. 2508–2516.