AcouRadar: Towards Single Source based Acoustic Localization

Linsong Cheng, Zhao Wang, Yunting Zhang, Weiyi Wang, Weimin Xu and Jiliang Wang
School of Software and BNRist, Tsinghua University, P.R. China
{chengls14, wangc16, zhangyt15, wang-wy14, xwm18}@mails.tsinghua.edu.cn,
jiliangwang@tsinghua.edu.cn

***Abstract*—Acoustic based tracking has been shown promising in many applications like Virtual Reality, smart home, video gaming, etc. Its real life deployments, however, face fundamental limitations. Existing approaches generally need three sound sources, while most COTS devices (e.g., TVs) and speakers have only two sound sources. We present AcouRadar, an acoustic-based localization system with single sound source. In the heart of AcouRadar we adopt a general new model which quantifies signal properties of different frequencies, distances and angles to the source. We verify the model and show that signal from a single source can provide features for localization. To address practical challenges, (1) we design an online model adaption method to address model deviation from real signal, (2) we design pulse modulated signals to alleviate the impact of environment such as multipath effect, and (3) to address signal dynamics over time, we derive relatively stable amplitude ratio between different frequencies, and thus provide a spectrum based localization method. We implement AcouRadar on Android and evaluate its performance for different COTS speakers in different environments. The results show that the model for localization can be generalized to different speakers. AcouRadar can achieve single source localization with average error less than 5 cm and average angle error of $1.76°$.**

I. Introduction

Acoustic signal based tracking has been shown as a promising technique for many applications like smart home, Virtual Reality (VR), Augmented Reality (AR), video gaming, gesture control, etc. Recent years have witnessed increasing number of acoustic based tracking systems [1][2][3][4] on mobile devices using everyday speakers as signal source. By measuring the distance or distance change to two/three sound sources (e.g., speakers), 2D/3D localization and tracking can be achieved. Acoustic tracking systems use everyday speakers as signal sources, and their computation and signal processing overhead can be relatively low. Thus existing systems are usually suitable to be implemented on commercial devices like COTS smartphone, smart appliance, and embedded devices [1][2], leading to great potential for mobile applications.

Despite of high accuracy and efficiency, existing approaches face practical and fundamental limitations in real deployments. First, existing acoustic tracking approaches usually require periodical localization to bootstrap and alleviate accumulated tracking error. Second, they require three sound sources to achieve 3D tracking and localization, while most COTS devices (e.g., TVs) and speakers have only two sound sources. Such a requirement significantly limits the applicability of acoustic tracking.

To address this limitation, a natural question is, can we achieve 2D (3D) localization with a single (two) acoustic source(s)? Achieving this will improve the performance of acoustic tracking in calculating initial position and updating position, and also enable acoustic tracking to most TVs and commercial speakers with at most two sound sources. This also enables us to explore the limit of acoustic based localization.

We adopt a new acoustic space model for single acoustic source, which quantifies signal properties with respect to different frequencies, distances and angles to the acoustic source. We verify the model and show that signal from a single sound source exhibits diversity and provides features for localization in the space [5]. From the model, we formally derive three basic signal characteristics: (1) distance attenuation: the received signal strength attenuates as increasing of the distance to the signal source. (2) directive radiation: the received signal strength is related to the radiation angle to the source. Signal with the same distance but different direction to the source exhibits different strength. (3) frequency diversity: signals of different frequency exhibit different distance attenuation and directive radiation models. Even at the same location and the same angle, signal strength of different frequency can be different.

Based on the model, we propose AcouRadar, a single source acoustic-based localization approach. With AcouRadar, a mobile phone samples the acoustic signal from a single source. Based on the received signal, AcouRadar leverages the model to derive the angle and distance to the source and thus the location. To be practical, AcouRadar needs to address three non-trivial challenges. (1) Model deviation: Practical signal may deviate from the theoretical model. Even worse, the deviation for different speakers may be different. (2) Environment Impact: Environment factors such as multi-path effect, which tends to be obvious for acoustic signal, impacts the received signal and localization accuracy. (3) Signal dynamics and hardware diversity: Different speakers may transmit signal at different power levels, and signal may fluctuate over time.

To address those challenges, we online adapt the theoretical model to compensate the deviation from practical signal. By analyzing the multi-path effect of different signal patterns and their corresponding impact, we design a pulse based signal pattern and processing method, so as to alleviate the impact of multi-path effect. Moreover, we find that different

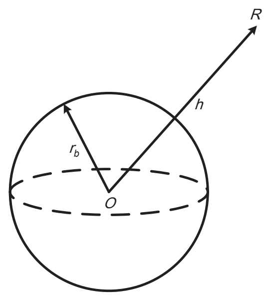


Fig. 1: Spherical sound source model.

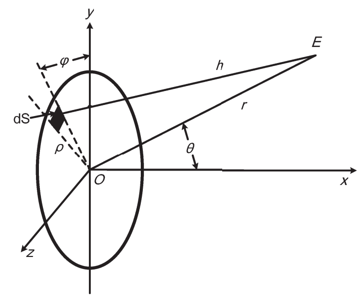


Fig. 2: Virtual beamforming model.

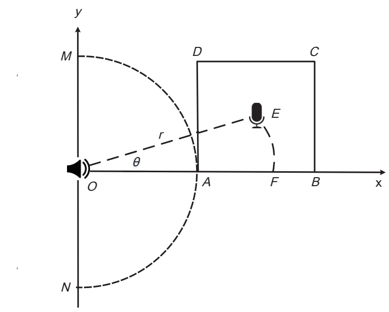


Fig. 3: Localization.

speakers may transmit signal at different power levels and signal amplitude may fluctuate due to various impact factors. However, the relation of signals for different frequencies (e.g., amplitude ratio between different frequencies) remains relatively stable. We analyze and quantify the relatively stable feature for different frequencies, and exploit the feature for localization.

We implement AcouRadar on COTS speakers and smartphones, and examine its performance in different environments. To the best of our knowledge, AcouRadar is the first system of its kind that is able to achieve localization with single source. The evaluation results show that AcouRadar can achieve localization with error less than 5 cm. This shows that this work can be applied to devices such as nowadays smart TVs and other smart devices with speakers. For example, gesture control can be realized with its two-channel speakers on those devices. We believe AcouRadar would significantly extend application scenarios of acoustic based tracking with nowadays devices, like in gesture control in smart home, VR, user tracking with smart speaker, etc.

The contributions of this work are as follows.

- We introduce the single source acoustic space model, which quantifies signal properties with different signal frequencies, distances and angles to the source in the space.
- Based on the model, we propose the design of AcouRadar, a single source localization method suitable for nowadays speakers. AcouRadar addresses practical challenges such as model deviation, multi-path effect, signal dynamics, hardware diversity, etc.
- We implement AcouRadar and examine its performance extensively in different settings. The evaluation results show that AcouRadar can achieve single source localization with average error less than 5 cm.

The remainder of this paper is structured as follows. Section II introduces the single source localization model. Section III presents the main challenges for single source 2D localization based on the model. Section IV describes AcouRadar design in detail. Section V presents our implementation and comprehensive experimental evaluation. Section VI introduces related work. Finally, Section VII concludes this paper.

II. Single Source localization

A. Localization Model

We first show the single source localization model of typical speakers.

Ball Sound Model. Sound is a physical phenomenon and it spreads in waves in the air. In the process of spreading, a sound wave will cause air pressure change during propagation. As a physical phenomenon, it follows basic physical principles.

Ball vibration. Fig.1 shows a spherical sound source of radius r_b. Assume that the vibration velocity of the sound source is $u = u_0 e^{j(\omega t - kr_0)}$, where c_0 is the speed of sound traveling in the air, $k = \frac{\omega}{c_0} = \frac{2\pi f}{c_0}$ is the propagation factor, and kr_0 is the initial phase of the vibration.

Air vibration. The ball vibration results in air vibration and thus sound wave. The sound wave should follow kinetic function, continuity function and state function [5][6]. The sound air pressure p with distance h to the sound source can be calculated as

$$p = \frac{|A|}{h} e^{j(\omega t - kh + \gamma)} \tag{1}$$

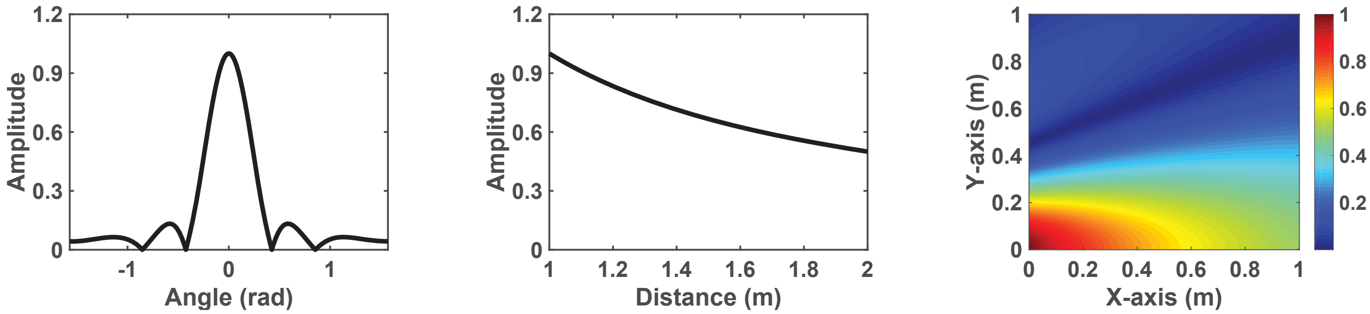
where $|A| = \frac{\rho_0 c_0 k r_b^2 u_0}{\sqrt{1+(kr_b)^2}}$, and $\gamma = arctan(\frac{1}{kr_b})$. For simplicity, we omit the details of calculation and interested readers can refer to [5][6]. From this equation, we can see that the air pressure varies with the vibration of the source. Meanwhile, the air pressure is inversely proportional to the distance to the source.

Point Sound Model. When the radius of the ball approaches 0, we have $kr_0 \to 0$ and $\gamma \approx \frac{\pi}{2}$. Based on Eq. (1), we have

$$p \approx j \frac{k\rho_0 c_0}{2\pi h} Q_0 e^{j(\omega t - kh)} \tag{2}$$

where $Q_0 = 2\pi r_0^2 u_0$. This gives the sound properties of a point source, which builds the foundation for sound properties of a speaker.

Speaker Sound Model. Further, we consider a general model for speakers, i.e., a circular planar piston of radius a with infinite plane baffle wall as shown in Figure 2. Assume the piston vibrates at speed $u = u_0 e^{j\omega t}$. Let the center of the piston be the origin of coordinates, and the surface of piston be the yz plane. The sound field is rotationally symmetric along the x-axis of the piston center. Without loss of generality,



(a) (b) (c)

Fig. 4: The virtual acoustic beamforming model. (a) Amplitude to angles for 10 KHz; (b) Amplitude to distance for 10 KHz; (c) 2D amplitude distribution for 10 KHz.

assume the observation point E in the sound field lies in the xy plane. The distance from E to the origin is r, and the angle to the x-axis is θ.

For a small area dS located at the polar radius ρ, polar angle ϕ, we have $dS = \rho d\rho d\phi$. According to Eq. (2), the sound pressure at E induced by vibration of dS can be calculated as

$$dp = j\frac{k\rho_0 c_0}{2\pi h} u_0 dS e^{j(\omega t - kh)} \tag{3}$$

Thus the sound pressure induced by the entire piston can be calculated as beamforming of small areas in the entire surface, i.e. $P = \iint dp$. Considering the receiver is located at the far field with a distance to the speaker much larger than the speaker radius, i.e. $r \gg a$, we have $h \approx r - \rho cos(\hat{\rho, r})$. As $cos(\hat{\rho, r}) = sin\theta cos\phi$, based on Eq. (3), we have

$$P = j\frac{\omega \rho_0 u_0}{2\pi r} e^{j(\omega t - kr)} \int_0^a \rho d\rho \int_0^{2\pi} e^{jk\rho sin\theta cos\phi} \tag{4}$$

Based on Bessel function $J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{jxcos\phi}$ $d\phi$ and $\int xJ_0(x)dx = xJ_1(x)$, Eq. (4) can be written as

$$P = j\omega\frac{\rho_0 u_0 a^2}{2r}[\frac{2J_1(kasin\theta)}{kasin\theta}]e^{j(\omega t - kr)} \tag{5}$$

According to Eq. (1), we have $v = Hp$ where $H = \frac{1}{\rho_0 c_0}(1 + \frac{1}{jkr})$. Thus the vibration velocity at point E is $v_E = \iint_S H dp = HP$. Thus vibration s of sound signal can be calculated as

$$\begin{aligned} s = &\int v_E dt = H\int P dt = \frac{H}{j\omega}P \\ = &\frac{u_0 a^2}{2c_0 r}(1 + \frac{1}{jkr})[\frac{2J_1(kasin\theta)}{kasin\theta}]e^{j(\omega t - kr)} \end{aligned} \tag{6}$$

Therefore, the amplitude T, i.e. the measured signal strength, can be calculated as

$$T(r,\theta,f) = \frac{u_0 a^2}{2c_0 r}\sqrt{1 + \frac{1}{k^2 r^2}}\left|\frac{2J_1(kasin\theta)}{kasin\theta}\right| \tag{7}$$

The equation shows that the amplitude of the sound signal is related to the distance between the sound source and the observation point (r), the angle between the observation point and the sound source (θ), and the signal frequency (f). We consider this as a *Virtual Acoustic Beamforming Model* (VABM).

B. *Signal Characteristics*

According to VABM, we can derive the following characteristics for acoustic signal from a single source speaker.

Directive radiation characteristic: Given a certain value for the signal frequency (f_0) and distance (r_0), the received signal strength T is related with the received direction θ, i.e.

$$T(r_0, \theta, f_0) = \beta_r \left|\frac{2J_1(\zeta_r sin\theta)}{\zeta_r sin\theta}\right| \tag{8}$$

where β_r is a coefficient, and $\zeta_r = ka = \frac{2\pi f_0}{c}\frac{d}{2}$. Figure 4 (a) shows the received signal strength $T(r_0, \theta, f_0)$ to different angle θ with $r_0 = 1$ for $f_0 = 10$ KHz. We can see that with the same distance, the received signal strength is different with respect to different angle θ.

Distance attenuation characteristic: For the same angle θ_0 and frequency f_0, the signal attenuates as increasing of r. This phenomenon is also widely explored for wireless signal. The signal strength can be calculated as

$$T(r, \theta_0, f_0) = \beta_d \frac{1}{r}\sqrt{1 + \frac{1}{k^2 r^2}} \tag{9}$$

where β_d is a coefficient related to θ_0 and f_0. Figure 4 (b) shows the amplitude of the received signal with different distance to the source for frequency $f_0 = 10$ KHz and $\theta_0 = 0$.

Frequency diversity characteristic: Even for a fixed received angle and distance, received signal strength T is related to the signal frequency:

$$T(r_0, \theta_0, f) = \beta_q \left|\frac{2J_1(\zeta_q f)}{\zeta_q f}\right| \tag{10}$$

where β_q is a coefficient, and $\zeta_q = \frac{\pi d sin\theta}{c}$.

Figure 4 (c) shows the received signal strength of 10 KHz in a square area. For the same frequency, the received signal strength is different with different angels and distance to the source.

C. *Basic Localization Method*

Based on the model, we first show basic localization method. Later we show how to address practical challenges such as signal instability. Considering the 2D localization scenario as in Figure 3, the speaker is located in O. Denote the

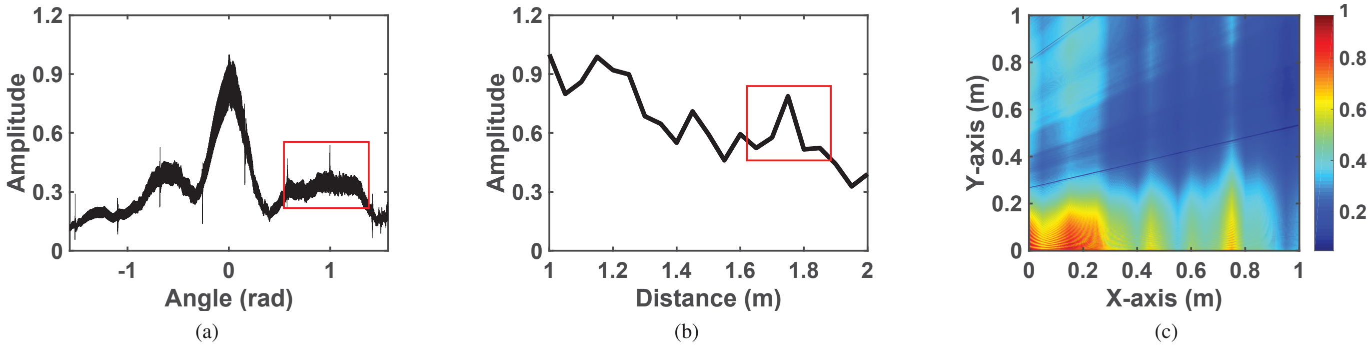


Fig. 5: The real received signal. (a) Amplitude to angles for 10 KHz; (b) Amplitude to distance for 10 KHz; (c) 2D amplitude distribution for 10 KHz.

length of OA as $|OA|$. Suppose a receiver measures the signal strength along the arc from E to F with distance $r_0 = |OE|$ to the source. The signal strength $\tilde{T}$ at position $\tilde{r}$ and $\tilde{\theta}$, given the signal frequency f_1, can be calculated as

$$T(\tilde{r}, \tilde{\theta}, f_1) = \tilde{T}_1 \tag{11}$$

The goal is to calculate $\tilde{r}$ and $\tilde{\theta}$ according to Eq. (7). Basically, if we can calculate all the parameters such as u_0, a and k in Eq. (7), we can calculate $\tilde{r}$ and $\tilde{\theta}$ directly. This is, however, difficult in practice especially for unstable signal.

To obtain $\tilde{r}$ and $\tilde{\theta}$, the receiver first measures the signal strength T_A at A. According to Eq. (9), we can calculate $\beta_d = \frac{|OA|T_A}{\sqrt{1+\frac{1}{k^2|OA|^2}}}$. Meanwhile, according to Eq. (8), the amplitude T_A at A can be calculated as $T_A = \beta_r \left| \frac{2J_1(kasin0)}{kasin0} \right|$, where the Bessel function $\frac{J_1(x)}{x} = \frac{1}{2}$ when $x = 0$. Thus, we can obtain $\beta_r = T_A$.

With β_d, we can calculate the signal strength $T(r, 0, f_0)$ for any distance r according to Eq. (9). Then with $T(r, 0, f_0)$ and β_r, we can calculate the signal strength for any position (r, θ) according to Eq. (8). This indicates we can derive the signal strength in any position (r, θ). This also means that we can find the position ($\tilde{r}$, $\tilde{\theta}$) with signal strength equal to $\tilde{T}_1$.

There may be, however, multiple positions with the same signal strength. To address this issue, we can use multiple frequencies and thus we have

$$\begin{cases} T(\tilde{r}, \tilde{\theta}, f_1) = \tilde{T}_1 \\ T(\tilde{r}, \tilde{\theta}, f_2) = \tilde{T}_2 \end{cases} \tag{12}$$

we solve the equation array to obtain the position.

III. Practical Challenges

The basic localization method in practice, however, faces several challenges.

Model deviation. The real received signal may deviate from the model. Figure 5 (a) shows the measured signal amplitude for different angles. Figure 5 (b) shows the measured signal amplitude for different distance. We can see that the real received signal amplitude in Figure 5 is different from that of the model in Figure 4. There are multiple spikes in the signal, and even worse, the center of the signal is also shifted. This indicates that practice received signal may not strictly follow the theoretical model due to different factors. Different impacting factors (e.g., hardware diversity) may cause the practical signal deviate from the theoretical model. For example, a speaker's vibration surface may not be an ideal plane. Further, we examine the real received signal strength in a square area as shown in Figure 5 (c). To address model deviation, we propose a model adaption method by combining real received signal with the theoretical model. Though practical model may vary across different devices and scenarios, we find that the trend of signal still follows the theoretical model. We extract information from received signal to efficiently adapt to the model.

Multi-path effect. Mutipath effect incurs a significant challenge for our localization method. First, there exist multiple static paths from the speaker to the mobile device. Second, environmental change (e.g., human moving) may generate dynamic paths, which also changes the received sound signals. The influence of multi-path effect for signal of different frequency and signal in different position is different. As shown in Figure 5 (a) and (b), we can see the signal amplitude is elevated or reduced for some angles and distances due to multi-path effect.

In summary, due to model bias and multi-path effect, as shown in Figure 5 (c), the received signal amplitude distribution is much different from that of the theoretical model in the space.

Signal fluctuation and location ambiguity. Ideally, in the basic localization method, we can achieve localization with the intersection of two curves for two different frequencies. In practice, there may exist measurement error which cause the curves deviate from that in the model.

IV. AcouRadar Design

Figure 6 shows the system overview of AcouRadar. AcouRadar mainly consists of three components, i.e., signal modulation and processing, model adaption, and spectrum spreading based intersection for localization.

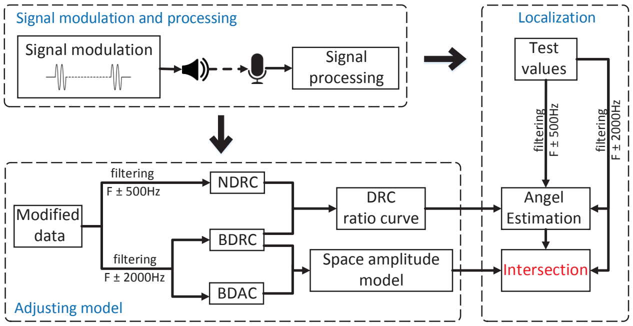


Fig. 6: System overview.

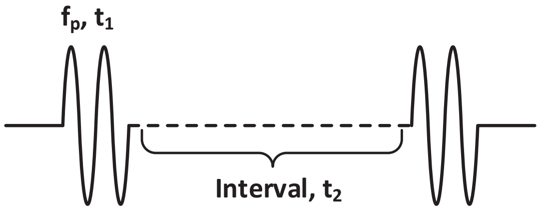
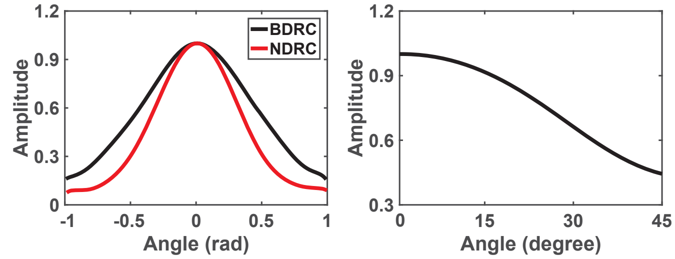


Fig. 7: Modulated sound signal.



(a) (b)

Fig. 8: (a) DRC of different frequency ranges. (b) RDC ratio curve.

A. Signal Modulation and Processing

Acoustic signal tends to be impacted by environment factors. In a typical indoor environment, there are multiple paths for acoustic signal propagation, each of which contributes to a different delayed and attenuated signal. Hence, a received sound signal is the combination of the direct path signal and multiple reflected signals.

Therefore, we should modulate the acoustic signal to minimize the impact of multi-path effect. As shown in Figure 7, AcouRadar employs a repeated pulse signal design with pulse of length t_1 and frequency f_p and inter-pulse interval of t_2. We mainly consider two requirements for the signal modulation to mitigate the impact of multi-path effect. (1) Avoid self-interference. Given a pulse in the signal, the pulse from the direct path should not interfere with that of reflected path. This means t_1 should be smaller than the difference of propagation time between the direct path and reflected paths. Otherwise, the direct path signal may overlap with the reflected signals. (2) Avoid consecutive interference. The pulse of a reflected path should not interfere with the followed pulse. This means t_2 should be large enough so that the impact of current pulse to followed pulses will be minimized.

Pulse length and inter-pulse interval. For the first requirement, considering the minimum difference between the direct path distance and a reflected path distance as 0.2m, the pulse length t_1 should be smaller than $\frac{0.2m}{343m/s} = 0.58ms$. For the second requirement, we conduct experiments and find that a reflected object more than 6 meters away creates very weak reflected signals that can be ignored [7]. Thus the minimum interval t_2 between two pulses can be set to $\frac{6m\times 2}{343m/s} = 35ms$. Besides, considering that we need an emitting frequency larger than 10 Hz, we have $t_2 < \frac{1}{10} - t_1 \approx 100ms$.

Frequency selection. We select the pulse's frequency f_p based on the following three factors. First, the selected frequency should be different from the background noise, such as human voice which is usually under 1 KHz and music instruments which are usually under 2.5 KHz. Second, the selected frequency should not exceed the physical sampling capability of COTS mobile phones, which is usually 22 KHz. Third, to facilitate signal processing and removing noise, each pulse should contain at least 2 cycles of signals. Thus the selected frequency should be larger than $f_p > \frac{2}{t_1} = \frac{2}{0.58\times 10^{-3}} \approx 3.4$ KHz.

Signal processing. First, we partition the received signals into segments, each of length $t_1 + t_2$. Second, we filter unnecessary signals and noise. Later we will show more details of this step and its impact on the localization accuracy. Third, we extract the power delay profile of the received signal. We calculate the highest peak in the profile, which corresponds to the amplitude of the direct path from the speaker to the microphone. To further mitigate noise, we calculate the median amplitude of peaks for multiple segments.

B. Model Adaption

As discussed in Section II-C, we need to recover the signal distribution in the whole space based on the model. Intuitively, we can sample the signal properties of a certain selected frequency on two points and then fit the model with the measured parameters. In practice, however, the energy of single frequency is very low. Thus it is unstable to use a single frequency to build the practical signal model.

Instead of using a single frequency, we leverage the frequency diversity and use a spectrum of frequencies for acoustic space amplitude model. We find that the received signal contains a spectrum of frequency and we can use the spectrum to build the model. By applying a band-pass filter, the resulted signal consists of a spectrum of n frequencies $f_1, f_2, \ldots, f_n$.

We show how to use the spectrum of frequencies to build the model for localization. As an example, we first consider two signals of frequencies f_1 and f_2, denoted as $X_1 = A_1 cos(2\pi f_1 t + \phi_1)$ and $X_2 = A_2 cos(2\pi f_2 t + \phi_2)$, where A_1 and A_2 are the amplitude for f_1 and f_2, respectively. The

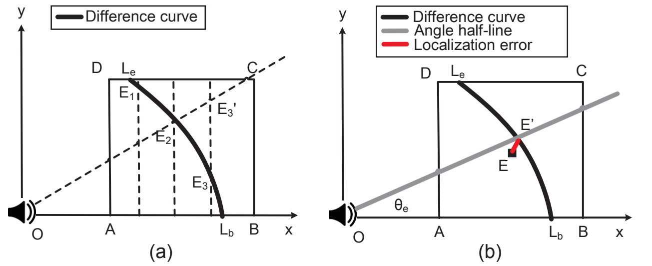


Fig. 9: Localization based on intersection. (a) Monotonicity of the candidate points; (b) Intersection of difference curve and angle half-line.

combined amplitude can be calculated as

$$|T^f| = \sqrt{A_1^2 + A_2^2 + 2A_1A_2cos[(2\pi f_2 - 2\pi f_1)t + (\phi_2 - \phi_1)]} \quad (13)$$

Meanwhile, according to Eq. (7), when the frequencies of A_1 and A_2 are close, we have

$$|T^f| \approx \frac{1}{r}\sqrt{1+\frac{1}{k^2r^2}}crdc(f_1, f_2, \theta) \quad (14)$$

where $k = \frac{2\pi f_1}{c}$ and $cdrc(\cdot)$ can be derived based on Eq. (7) and Eq. (13). We call $cdrc(\cdot)$ the combined directive radiation function of f_1 and f_2. In practice, $\frac{1}{k^2r^2}$ is approximately zero. For example, when $f_1 = 17$ KHz and $r = 1$, $\frac{1}{k^2r^2} = 0.00001$. Therefore, we have $T^f \approx \frac{1}{r}crdc(f_1, f_2, \theta)$. The $cdrc(\cdot)$ function shows that the combined amplitude of two frequencies at a certain position (r, θ) is related to r and θ.

Similarly, we can see that the combined amplitude for a spectrum of frequencies also is related to r and θ. We have

$$T^f = \frac{1}{r}crdc(f_1, f_2, ..., f_n, \theta) \quad (15)$$

This means we can use the spectrum to build the signal model in the space.

As shown in Figure 3, instead of sampling one single point, we sample along a circle of radius $|OA|$ centered at the source O. Based on the sampling, we can build the directive radiation curve along the line AB. We can also build the distance attenuation curve by sampling along the line AB. More specifically, AcouRadar first filters the signal with a broad band-pass filter. The amplitude sum of multiple frequencies exhibits the same characteristics with that for a single frequency. AcouRadar uses a fitting based method to derive the broadband directive radiation curve (BDRC) and broadband distance attenuation curve (BDAC). Using BDRC and BDAC, we can obtain the broadband space amplitude model $T_B^f(r, \theta)$ similar to Section II-C. The reason for choosing a band-pass filter is to ensure that the amplitude model is stable.

Further, AcouRadar filters the signal using a narrow band-pass filter with a higher center frequency. It then derives narrowband directive radiation curve (NDRC) and narrowband distance attenuation curve (NDAC). Correspondingly, we can obtain the narrowband space amplitude model $T_N^f(r, \theta)$ using NDRC and NDAC. Figure 8 (a) shows the BDRC and NDRC.

Using the frequency diversity, we find that though the signal amplitude of a single frequency may fluctuate, the relation between the amplitude for two spectrums (broadband and narrowband) is more stable. We use the curves for BDRC and NDRC for angle estimation. According to Eq. 15, we calculate the curve ratio CR of two curves of different spectrum in $f_1, f_2, \ldots, f_n$ and $f_1', f_2', \ldots, f_n'$ as

$$CR = \frac{T_B^f(r,\theta)}{T_N^f(r,\theta)} = \frac{\frac{1}{r}crdc(f_1, f_2, ..., f_n, \theta)}{\frac{1}{r}crdc(f_1', f_2', ..., f_n', \theta)} = \frac{crdc(f_1, f_2, ..., f_n, \theta)}{crdc(f_1', f_2', ..., f_n', \theta)} \quad (16)$$

The curve ratio is related to θ and irrelevant to r. As the curve ratio is monotonous, we can estimate the received angle according to the ratio of $T_B^f(r, \theta)$ and $T_N^f(r, \theta)$. Figure 8 (b) shows the curve ratio according to BDRC and NDRC of Figure 8 (a). This indicates if we can measure the curve ratio, we can calculate the corresponding angle θ.

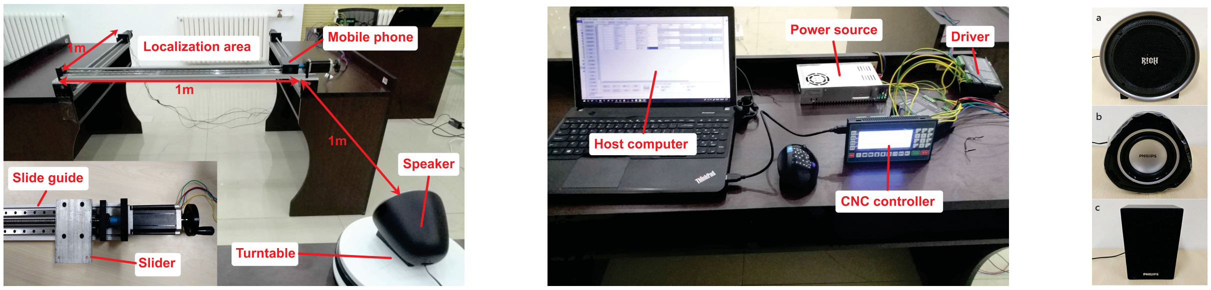
C. Intersection based Localization

Intuitively, we have two monotonous curves. The curve ratio CR is monotonous decreasing with the increasing of angle θ. The distance attenuation curve is monotonously decreasing with increasing of distance r. After measuring the curve ratio and amplitude at a certain position, the localization method can work as follows.

We can see that the amplitude of received signal decreases with the increasing of both angle θ and distance r. Given x, both θ and r increase with the increasing of y. Thus the amplitude monotonously decreases with the increasing of y. For each fixed x_i, we use the space amplitude model to search for candidate point $E_i(x_i, y_i)$ whose model amplitude T_{E_i} is closest to the measured amplitude T'. Here T_{E_i} can be calculated based on $T_B^f(\sqrt{x_i^2 + y_i^2}, \arctan\frac{y_i}{x_i})$ and $T_N^f(\sqrt{x_i^2 + y_i^2}, \arctan\frac{y_i}{x_i})$. For different x_i, we obtain a difference curve by fitting the candidate points as shown in Figure 9 (a). L_b is the start point of the difference curve and L_e is the end point.

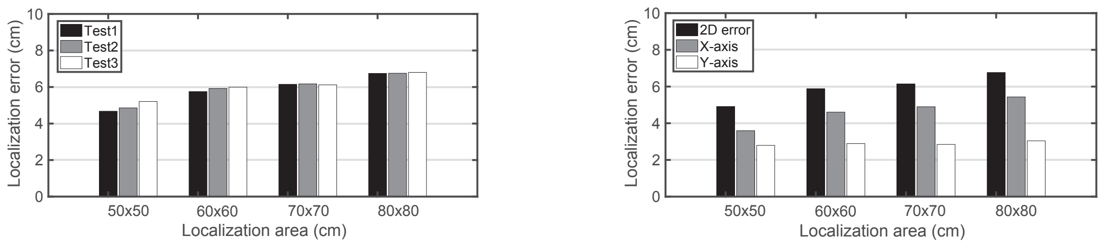
As an example shown in Figure 9 (a), denote the coordinate of X-axis for E_1, E_2, and E_3 as x_1, x_2, and x_3 where $x_1 < x_2 < x_3$. For any two points on the difference curve (e.g., E_2 and E_3), we first calculate E_3' as the intersection of ray OE_2 and line $x = x_3$. As $|OE_3'| > |OE_2|$, we have $T_{E_2} > T_{E_3'}$. T_{E_3} is equal to T_{E_2} and the amplitude monotonously decreases with the increase of y along line $x = x_3$. Thus the y coordinate of E_3 is smaller than that of E_3' and the angle $\angle E_3OA$ is smaller than the angle $\angle E_2OA$. Therefore, for all points on difference curve, the corresponding angle θ monotonously decreases with the increase of x.

For the estimated angle θ_e, we can find a unique point in the difference curve with angle closest to θ_e. Specially, when θ_e is larger than the angle at point L_b, L_b is considered as the intersection; when θ_e is smaller than the angle at point L_e, L_e is considered as the intersection. Thus, we can obtain a unique intersection E' as shown in Figure 9 (b). The distance of point E and E' is the localization error as denoted by the red line.



(a) (b) (c)

Fig. 10: (a) Experiment scenario; (b) CNC Control system; (c) Speakers in experiments



(a) (b)

Fig. 11: (a) Average localization errors for different localization area sizes. (b) Average localization errors for different dimensions.

V. Evaluation

A. Methodology

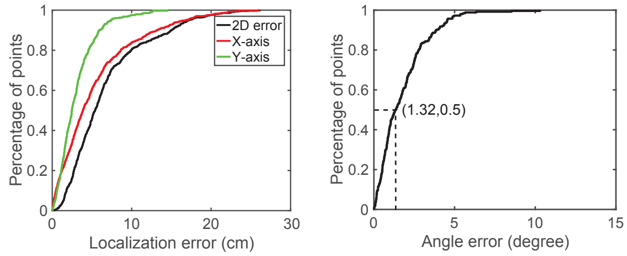
We implement AcouRadar on Android and evaluate its performance for COTS speakers. In our implementation, we use the speaker Dostyle SD316 as the sound source and the Samsung phone (SM-G9300) as the receiver. To obtain the ground truth of the mobile phone location, we use a computer numerical control (CNC) sliding trail system to control the position of mobile phone. Figure 10 (a) shows the sliding trail system. The CNC sliding trail system consists of three sliding trails, a control system and a slider modular on the moving trail, supporting moving in 3D space. It can also record real time position of the slider modular. We attach a mobile phone with AcouRadar on the sliding modular. In our experiment, there are multiple paths from the speaker to the device (e.g., the signal can be reflected by the sliding trails or desk). The sliding system also create non-negligible noise to our localization system. Figure 10 (b) illustrates the CNC control system. Figure 10 (c) shows different COTS speakers in our experiment. We mainly evaluate the performance of AcouRadar from the following aspects.

B. Overall Accuracy

We conduct four sets of experiments to evaluate AcouRadar's overall accuracy for different localization area with side length from 0.5 m to 0.8 m. For each size, we gradually move the mobile phone in a grid with interval of 5 cm on the X axis or Y axis. We record the location from the CNC system and compare it with the output of AcouRadar. On each position, we conduct three runs of tests. For example, the test number of a 0.8m$\times$ 0.8m localization area is $3 \times 16 \times 16 = 768$.

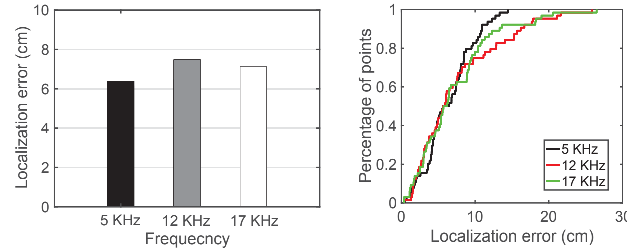
For all tests, Figure 11 (a) plots the average localization error for different localization area size. Overall, we have three observations. First, the average localization error for different localization area size is 4.91 cm, 5.89 cm, 6.15 cm, 6.77 cm. Second, we can also see that the average error increases with the area size. Third, for the same localization area size, the average localization errors of three tests are similar. This indicates the localization results of AcouRadar is relatively stable. Besides, we evaluate the influence of different distances and angles between the speaker and mobile device. Even if the distance increases to 3 m, average localization error is less than 10 cm. Therefore, for many common scenarios (e.g., using speakers of TV), AcouRadar can provide acceptable performance for most of acoustic based tracking applications, such as gesture control.

We further evaluate AcouRadar's average localization errors for different dimensions as Figure 11 (b) shows. From the result, we can see that the average localization errors on X-axis increase with the increasing of localization area size. However, the average localization errors on Y-axis are relatively stable, i.e., 2.89 cm in average. Figure 12 (a) plots the cumulative distribution function (CDF) of AcouRadar's average localization errors AcouRadar achieves performance with 50 percentile error of 5.38 cm, 3.85 cm, and 2.45 cm for 2D localization in X-axis and Y-axis, respectively.



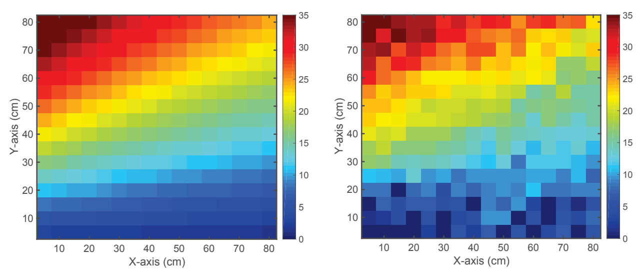
(a) (b)

Fig. 12: (a) CDF of average localization errors; (b) CDF of angle errors.



(a) (b)

Fig. 14: (a) Average localization errors at different frequencies; (b) CDF of errors.



(a) Model angle (b) Estimated angle

Fig. 13: Comparison of theoretical angles and estimated angles.

C. Angle Accuracy

To evaluate the angle accuracy, we first calculate the angle of mobile phone based on its position. Then we estimate the angle using the directive radiation characteristic ratio curve of BDRC and NDRC. We calculate the angle errors of all test points in the localization area. Figure 13 (a) and (b) show the result of real angle and estimated angle distribution. We can see that the estimated angle is very close to the real angle.

To quantitatively examine the angle error, Figure 12 (b) shows the cumulative distribution function of angle error. AcouRadar achieves good performance in angle estimation with an average error around 1.76 degrees, 50 percentile error of 1.32 degrees, and 90 percentile error of 3.89 degrees.

D. Impact of Frequency

We conduct three sets of experiments with sound signals at 5 KHz, 12 KHz, and 17 KHz, respectively. Sound signal at 5 KHz can be heard easily by people, sound signal at 17 KHz almost cannot be heard by most people, and sound signal at 12 KHz is between them. As shown in Figure 14 (a), the average localization errors of sound signals at 5 KHz, 12 KHz, and 17 KHz are 6.37 cm, 7.47 cm, and 7.13 cm, respectively. The localization error for signal of different frequencies is similar, which demonstrates that AcouRadar can work well with sound signal of different frequencies. The error for 17 KHz is slightly higher than that of lower frequency. Figure 14 (b) shows that most positions exhibit similar accuracy. We find that the CNC sliding trail system emits some high frequency noise. Thus higher frequency is more sensitive to error.

VI. Related Work

Acoustic localization. Recent years, acoustic tracking and localization attract many research efforts. AAMouse [8] accurately calculates the moving speed of mobile phone based on Doppler effect and achieves 3D tracking with three speakers. CAT [3] further improves the signal design and proposes to use Frequency Modulated Continuous Waveform (FMCW). LLAP [1] proposes an accurate localization method by measuring phase change with received acoustic signal. Vernier [2] further uses a method to efficiently measure the phase change with a shorter delay. There are also many applications [9] [10] [11] [12] [13] [14][15] built on tracking and localization using acoustic signal. For example, AIM [16] proposes an acoustic based imaging method on mobile phone. BatTracker [4] leverages echoes from nearby objects and uses distance measurements from them to correct error accumulation in inertial sensor based device position prediction. BeepBeep [17] shows an efficient method to measure the distance between two mobile devices. SwordFight [18] estimates the distance between two phones.

RF localization. Many RF based localization approaches are also proposed [19][20][21][22][23][24][25]. Widraw [26] leverages WiFi signals from commodity mobile devices to enable hands-free drawing in the air. INTRI [27] employs the concept of trilateration in fingerprint-based WIFI environment. For example, Tagoram [28] leverages RFID for accurate tracking and achieves a mm-level accuracy. MobiTagbot [29] marks the book with an RFID tag and uses a robot equipped with an antenna to locate the tag on each book.

Light localization. Light localization systems [30] [31] [32] mainly use visible light sources (e.g., LED) for localization. CELLI [33] maps the position of the space with the LED pixels. The positioning is based on analyzing the time of the received signal from the LED. SmartLight [34] associates the spatial position to a circle on the LED panel. iLAMP [35] uses frequency and color spectrum as features to identify each light and derive the location by triangulation.

VII. Conclusion

We present AcouRadar, an acoustic-based localization method with a single sound source. We introduce a single source based virtual acoustic beamforming model to quantify signal amplitude with different frequencies, distances and angles to a single source. The model builds the theoretical foundation for single source localization. We address practical challenges while applying the model for localization. To address model deviation from real signal, we propose an online adaption model with received signal to compensate the deviation. To alleviate the impact of environment such as multi-path effect, we present a pulse based signal design and signal processing method. To address signal dynamics in practice, we leverage relatively stable amplitude ratio between different frequencies. We implement AcouRadar on Android, and evaluate its performance with mobile devices and COTS speakers in different environments. The experiment results show that AcouRadar achieves accurate single source localization, which shows its great potential in real applications.

ACKNOWLEDGEMENTS

This work is in part supported by National Key R&D Program of China 2017YFB1003000, National Natural Science Fund for Excellent Young Scholars (No. 61722210), National Natural Science Foundation of China (No. 61932013, 61532012, 61432015).

REFERENCES

[1] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *International Conference on Mobile Computing and NETWORKING*, 2016, pp. 82–94.

[2] Y. Zhang, J. Wang, W. Wang, Z. Wang, and Y. Liu, "Vernier: Accurate and fast acoustic motion tracking using mobile devices," in *Proceedings of ACM INFOCOM, 2018*.

[3] W. Mao, J. He, and L. Qiu, "Cat: high-precision acoustic motion tracking," in *International Conference on Mobile Computing and NETWORKING*, 2016, pp. 69–81.

[4] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Battracker: High precision infrastructure-free mobile device tracking in indoor environments," in *The ACM Conference*, 2017, pp. 1–14.

[5] Mechel and F. P, *Formulas of Acoustics*. Berlin/Heidelberg: Springer Berlin Heidelberg, 2004.

[6] P. M. Morse and K. U. Ingard, *Theoretical acoustics*. New York, McGraw-Hill, 1968.

[7] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Batmapper: Acoustic sensing based indoor floor plan construction using smartphones," in *International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 42–55.

[8] S. Yun, Y. C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 15–29.

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

[10] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 45–57.

[11] K. Liu, X. Liu, and X. Li, "Guoguo: Enabling fine-grained smartphone localization via acoustic anchors," *IEEE Transactions on Mobile Computing*, vol. 15, no. 5, pp. 1144–1156, 2016.

[12] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Batmapper: Acoustic sensing based indoor floor plan construction using smartphones," in *International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 42–55.

[13] J. Qiu, D. Chu, X. Meng, and T. Moscibroda, "On the feasibility of real-time phone-to-phone 3d localization," in *ACM Conference on Embedded Networked Sensor Systems*, 2011, pp. 190–203.

[14] X. Bian, G. D. Abowd, and J. M. Rehg, "Using sound source localization in a home environment," in *International Conference on Pervasive Computing*, 2005, pp. 19–36.

[15] H. Zhang, W. Du, P. Zhou, M. Li, and P. Mohapatra, "Dopenc: Acoustic-based encounter profiling using smartphones," in *Proceedings of ACM MobiCom*, 2016.

[16] W. Mao, M. Wang, and L. Qiu, "Aim: Acoustic imaging on a mobile," in *Proceedings of ACM MobiSys*, 2018.

[17] C. Peng, G. Shen, and Y. Zhang, *BeepBeep: A high-accuracy acoustic-based system for ranging and localization using COTS devices*. ACM, 2012.

[18] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, "Swordfight: enabling a new class of phone-to-phone action games on commodity phones," in *International Conference on Mobile Systems, Applications, and Services*, 2012, pp. 1–14.

[19] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections," in *Usenix Conference on Networked Systems Design and Implementation*, 2013, pp. 317–329.

[20] S. Kumar, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in *Usenix Conference on Networked Systems Design and Implementation*, 2016, pp. 165–178.

[21] J. Gjengset, J. Xiong, G. Mcphillips, and K. Jamieson, "Phaser: enabling phased array signal processing on commodity wifi access points," in *International Conference on Mobile Computing and NETWORKING*, 2014, pp. 153–164.

[22] J. Han, C. Qian, X. Wang, D. Ma, J. Zhao, W. Xi, Z. Jiang, and Z. Wang, "Twins: Device-free object tracking using passive tags," *IEEE/ACM Transactions on Networking (TON)*, vol. 24, no. 3, pp. 1605–1617.

[23] K. Qian, C. Wu, Z. Yang, Z. Zhou, X. Wang, and Y. Liu, "Enabling phased array signal processing for mobile wifi devices," *IEEE Transaction on Mobile Computing (TMC)*, vol. 17, no. 8, pp. 1820–1833, 2018.

[24] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, 2013.

[25] F. Xiao, Z. Wang, N. Ye, R. Wang, and X. Li, "One more tag enables fine-grained rfid localization and tracking," *IEEE/ACM Transactions on Networking*, vol. 26, no. 1, pp. 161–174, Feb 2018.

[26] L. Sun, S. Sen, D. Koutsonikolas, and K. H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *International Conference on Mobile Computing and NETWORKING*, 2015, pp. 77–89.

[27] S. He and S. H. G. Chan, "Intri: Contour-based trilateration for indoor fingerprint-based localization," *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1676–1690, 2017.

[28] L. Yang, Y. Chen, X. Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: real-time tracking of mobile rfid tags to high precision using cots devices," in *International Conference on Mobile Computing and NETWORKING*, 2014, pp. 237–248.

[29] L. Shangguan and K. Jamieson, "The design and implementation of a mobile rfid tag sorting robot," in *International Conference on Mobile Systems, Applications, and Services*, 2016, pp. 31–42.

[30] B. Xie, K. Chen, G. Tan, Y. Liu, Y. Liu, J. Wu, and T. He, "Lips: A light intensity–based positioning system for indoor environments," *Acm Transactions on Sensor Networks*, vol. 12, no. 4, p. 28, 2016.

[31] R. Gao, Y. Tian, F. Ye, G. Luo, K. Bian, Y. Wang, T. Wang, and X. Li, "Sextant: Towards ubiquitous indoor localization service by photo-taking of the environment," *IEEE Transactions on Mobile Computing*, vol. 15, no. 2, pp. 460–474, 2016.

[32] Y. S. Kuo, P. Pannuto, K. J. Hsiao, and P. Dutta, *Luxapose: Indoor positioning with mobile phones and visible light*. ACM, 2014.

[33] Y. L. Wei, C. J. Huang, H. M. Tsai, and C. J. Lin, "Celli: Indoor positioning using polarized sweeping light beams," in *International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 136–147.

[34] S. Liu and T. He, "Smartlight: Light-weight 3d indoor localization using a single led lamp," in *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '17. New York, NY, USA: ACM, 2017, pp. 11:1–11:14. [Online]. Available: http://doi.acm.org/10.1145/3131672.3131677

[35] S. Zhu and X. Zhang, "Enabling high-precision visible light localization in today's buildings," in *International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 96–108.