PAMT: Phase-based Acoustic Motion Tracking in Multipath Fading Environments

1st Yang Liu
Shanghai Institute of Microsystem and Information Technology (SIMIT), CAS; University of Chinese Academy of Sciences
Shanghai, China
yang.liu@mail.sim.ac.cn

2nd Wuxiong Zhang
Shanghai Institute of Microsystem and Information Technology (SIMIT), CAS Shanghai Research Center for Wireless Communications(WICO)
Shanghai, China
wuxiong.zhang@mail.sim.ac.cn

3rd Yang Yang
Shanghai Institute of Fog Computing Technology (SHIFT) ShanghaiTech University
Shanghai, China
yangyang@shanghaitech.edu.cn

4th Weidong Fang
Shanghai Institute of Microsystem and Information Technology (SIMIT), CAS
Shanghai, China
weidong.fang@mail.sim.ac.cn

5th Fei Qin
University of Chinese Academy of Sciences
Beijing, China
fqin1982@ucas.ac.cn

6th Xuewu Dai
Northeastern University
Shenyang, China
daixuewu@mail.neu.edu.cn

***Abstract*—Motion tracking technologies have been widely used in mobile interaction applications, such as Virtual Reality (VR), healthy monitoring, and virtual touch control. Compared with dedicated hardware devices, mobile phones use reliable speakers and microphones, and can serve as ubiquitous devices for cheap acoustic-based motion tracking solutions. However, for complex indoor environments, it is very difficult for acoustic-based methods to achieve accurate motion tracking due to multipath fading and limited sampling rate at mobile devices. In this paper, a new parameter named Multipath Effect Ratio (MER) is defined to indicate the multipath fading effect on received signals at different frequencies. Based on MER, a novel multipath effect mitigating technique is developed to calculate the phase change of acoustic signals and track the corresponding moving distance by using multiple speakers. A Phase-based Acoustic Motion Tracking (PAMT) method is then proposed and implemented on standard Android smartphones. Experiment results show, without any specialized hardware, PAMT can achieve an impressive millimeter-level accuracy for localization and motion tracking applications in multipath fading environments. Specifically, the measurement errors are less than 2mm and 4mm in one-dimensional and two-dimensional scenarios, respectively.**

***Index Terms*—Motion Tracking; Multipath Fading; Mobile Interaction;**

I. Introduction

Motion tracking technologies have been widely used in mobile interaction applications, such as Virtual Reality (VR). Based on electromagnetic signals, such as Wi-Fi signal [1], [2], visible light [3], [4], and millimeter wave [5], [6], different motion tracking techniques have been proposed to use dedicated hardware devices for achieving sub-meter-level accuracy in different scenarios. However, these methods are not good enough for tracking a user's gesture or posture in real time.

Compared with electromagnetic signals, acoustic signals have much lower frequencies and slower propagation speeds. Therefore, they are very suitable to be used for high-accurate and low-latency motion tracking applications. Further, unlike dedicated hardware devices, mobile phones use reliable speakers and microphones, and can serve as ubiquitous devices for cheap acoustic-based motion tracking solutions.

For complex indoor environments, it is particularly challenging to achieve accurate motion tracking using acoustic signals. Specifically, the low sampling rate on commercial mobile devices (*e.g.*, 48KHz for typical mobile phones) limits the resolution of traditional methods based on Doppler effect or Time-of-Arrival/Time-Difference-of-Arrival (ToA/TDoA) [15] [7]. Therefore, it is very difficult for those traditional methods to track small or slow movements of mobile devices. In addition, due to severe fading effects in indoor environments, normal mobile devices using traditional methods can hardly distinguish Line-of-Sight (LoS) signals from Non-LoS (NLoS) signals with slightly different delays from multiple propagation paths.

There have been some works on improving the accuracy of acoustic motion tracking [7]–[9]. However, they either require specified hardware or have limited performance in multipath fading environments. For example, CAT [7] proposes an enhanced FMCW method to improve the measuring accuracy, but it also cannot track small and slow movement as traditional FMCW methods. In SoundTrak [8], a dedicated smart watch tracks the motion of a micro-speaker using the inaudible acoustic signals, but it requires a specified microphone array placed on the smart watch to track the micro-speaker's motion. Vernier [9] proposes an active motion tracking approach by

This work is partially supported by the National Natural Science Foundation of China (No. 61571004), the Shanghai Natural Science Foundation (No. 16ZR1435200), and the Science and Technology Innovation Program of Shanghai (No. 17DZ1200302, No. 17DZ2292000, No. 16510711600)

calculating the phase change of received signals and then estimating moving distance. However, multipath fading effects are not taken into consideration in Vernier because the target device is moved in a small area without the interference from multiple paths.

In this paper, a Phase-based Acoustic Motion Tracking (PAMT) method is developed to mitigate multipath fading effects and provides fine-grained motion tracking for a typical mobile device, *i.e.*, smart phone. The received signal is a superposition of signals from different paths and it's difficult to distinguish LoS signals from NLoS signals on normal mobile devices. Thus, it's even harder to calculate the actual phase of LoS signals, which are vulnerable to multipath fading in indoor environments. In order to achieve accurate and robust motion tracking, we should accurately measure the transmission time change of LoS signals in multipath fading environments. Some works, such as [31], find the first peak of the cross-correlation between emitting signals and received signals to calculate arrive time change of LoS signals, and then estimate distance change. However, due to the limited sampling rate on mobile phones, it's challenging to achieve mm-level accuracy using the cross-correlation based methods. We leverage the fact that the multiple fading effects don't always affect the phase based measurements of different frequencies at the same time since signals at different frequencies have different wavelengths and phases during measurements. Specifically, we use multiple sources to transmit inaudible acoustic signals at different frequencies. Then the target mobile device receives these transmissions, and derives the distance change to each source based on the phases of received signals. In order to obtain the phase change that best matches moving distance, we select the frequency suffer from the least multipath fading interference to estimate actual moving distance. We now ask the following question: *How to determine the impact of multipath fading on the signals at different frequencies?* In order to select the signal suffering from the least interference, a new parameter named Multipath Effect Ratio (MER) is defined to indicate the multipath fading effects on received acoustic signals at different frequencies in each frame. Based on MER, PAMT can effectively mitigate multipath effects, obtain the accurate phase change in multipath fading environments, and then detect small movements for real-time motion tracking.

The key contributions of this paper are summarized as follows:

- To address the technical challenges of motion tracking in complex indoor environments, a new parameter named Multipath Effect Ratio (MER) is defined to capture the multipath fading effects on received acoustic signals at different frequencies.
- Based on MER, a novel Phase-based Acoustic Motion Tracking (PAMT) method is developed to measure subtle phase changes at different received signals in multipath fading environments, and derive the corresponding small movements for achieving accurate motion tracking.
- A prototype system is implemented on a standard Android smart phone. Experiment results show our PAMT method can achieve an impressive millimeter-level accuracy for localization and motion tracking applications in multipath fading environments. Specifically, the measurement errors are less than 2mm and 4mm in one-dimensional (1-D) and two-dimensional (2-D) scenarios, respectively.

In summary, our work achieves: (1) feasible motion tracking with mm-level accuracy on mobile devices, (2) strong robustness in dense multipath indoor environments.

The rest of this paper is organized as follows. Related work is reviewed in Section II. Section III provides the system design for phase calculation, multipath effect mitigation, and 1-D ranging. The details of the proposed PAMT method are given in Section IV. Implementation details and experimental results are presented and discussed in Sections V and VI, respectively. Some limitations and future work are analyzed in Section VII. Finally, Section VIII concludes this paper.

II. Related work

There are many existing work presenting different techniques of motion tracking, which are clearly different from our work.

Acoustic based motion tracking: Existing acoustic based measurements mainly use traditional technologies based on ToA/TDoA [10]–[12], or the Doppler effect based [13]–[15]. Those technologies have limited resolution due to the low sampling rate at mobile devices, and only provide coarse-grained measurements (*e.g.*, dozens of centimeters). Recently, there has been some works on motion tracking based on the phase change of acoustic signals [7]–[9], [17], [18]. In [7]–[9], acoustic source and receiver are different devices, and the phase offsets between receiver and source due to asynchronous system clocks are approximately compensated as a fixed value at each frequency. However, the rough approximations of the phase offsets limit accuracy and cause accumulating measuring error over time. Further, unlike with FMCW methods which can mitigate multipath using their high-band signals, such as [7], multipath effects aren't taken into consideration in typical active ranging approaches [8], [9], because the target devices in their works are moved in a small area without the interference from strong multiple paths, such as an area of about $10cm \times 10cm$ in [8]. For larger areas, the multiple effects can't be neglected and hinder the performance of motion tracking. In LLAP [17] and Strata [18], acoustic source and receiver are the same device, and the phase offset caused by system clock between them can be ignored. However, they use the reflection of acoustic signals to track user's motion instead of direct path signals, thus, their methods have limited coverage, usually less than a meter on mobile phones. Further, it's challenging to distinguish the reflected signals from multiple targets.

RF based motion tracking: Radio Frequency (RF) based schemes can be divided into two aspects: device-based and device-free schemes. Device-based measurements mainly use ToA/TDoA technologies to build theoretical models that geometrically quantifies the relationships between signal charis-

matics (*e.g.*, Received Signal Strength Indicator (RSSI)) and the user's location [2], [5], [6]. Most RF based device-free tracking systems use the RF signal reflected by human bodies to estimate the localization of indoor users [19], [20]. However, the propagation speed of light is fast, which makes it difficult to achieve accurate motion tracking. Therefore, existing RF signals based tracking schemes cannot achieve fine-grained accuracy using commercial devices.

Signature-based schemes: Signatures can be obtained from a variety of sources, such Wi-Fi signals [21], acoustic signals [22], [23], ambient noise [24], and magnetic fields [25]. The characteristics of the signals, such RSSI, Channel State Information (CSI), and spectrum, can be aggregated into a location signature. The schemes rely on the recording of the signal characteristics as signatures at a set of know location in an off-line phase. During the on-line tracking phase, users match characteristics with those recorded characteristics, and choose the closest match as the estimated users location. However, these signatures vary over time and environmental mobility, and need to be updated when the environment changes. Compared with signatures-based localization methods, we use model-based techniques to avoid the extensive measurements for priori knowledge.

III. System Design

In this section, the limitations for existing approaches are firstly presented. Then, the overview of our approach is provided. Finally, the system design is described, including phase calibration, multipath effect mitigation, and phase based ranging.

A. Limitation of Doppler based approaches

Existing sound based ranging systems mainly use Doppler effect to measure the relive distance and direction [13], [14], [26], [27]. The Doppler effect is a natural phenomenon where the frequency of the sound waves received by a mobile device is changed by a relative velocity between the device and the source. The moving direction and moving speed is obtained from the shift usually measured by Short Time Fourier Transform (STFT). However, the resolution of moving speed v_{res} is limited by the window size N_{STFT} of STFT [14]. We have

$$v_{res} = \frac{F_s}{N_{STFT} F_c} c. \tag{1}$$

Where c is velocity of sound in air. F_s is the sampling rate. The maximum F_s is 48 KHz for typical smartphones. F_c is the original frequency of the signal. When F_c is 18 KHz and N_{STFT} is 4096, the resolution of moving speed is about 22 cm/s. The resolution can be further improved by increase N_{STFT}, but system's dynamic performance would be reduced. Thus, Doppler based methods couldn't detect slow movement which is vital in mobile interaction. Further, Doppler based methods assume that the device moves uniformly in each STFT's moving window, which is noise-prone in practical system.

B. Limitation of FMCW based approaches

Some works use Frequency Modulated Continuous Wave (FMCW) signals with high bandwidth to estimate the distance between receiver and source [7]. In their work, FFT is usually used to analyze the information in frequency domain. However, the resolution of the frequency analysis is related to the bandwidth of FMCW. The resolution d_{res} is [9]

$$d_{res} = \frac{c}{B}. \tag{2}$$

Where c is sound speed, and B is the bandwidth of FMCW. For a bandwidth of 7 KHz, which is the bandwidth of inaudible signals in most smartphones, the resolution of acoustic ranging using FMCW is about 4.9 centimeter when c is 340 m/s. The resolution is insufficient to track small movement in mobile interaction. In our work, we calculate the distance change in time domain, instead of using frequency analysis, whose ranging resolution is not restricted by the bandwidth of signals.

C. PAMT Overview

Because of the above limitations of existing approaches, we propose a phase-based ranging approach for acoustic signals using LoS signals, which has more robust and accuracy performance. In order to make the sound inaudible, a static audio source, like a commercial speaker, continually transmits sinusoidal signals at single frequency in the range of 17-23 KHz. The sound in the range can be generated and received by many commercial devices without introducing audible noises.

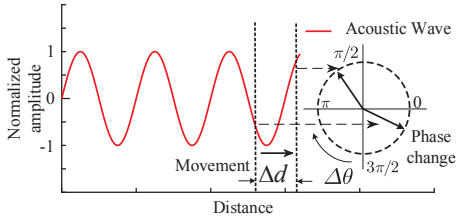


Fig. 1: Active phase based distance measurement.

The signals received by smartphone's microphone are composed of LoS signals and the NLoS signals reflected by static objects, such as wall and table. In indoor environments without dense multipath fading, when the receiver moves close/away, the length of the LoS paths would change as the moving distance and the reflection paths change slowly compared with LoS paths. When the receiver, like a smartphone, moves close/away, the phase of acquired signal would increase/decrease due to the length change of LoS path, as shown in Figure 1. As the phase of the signal increases by 2π, the path length would decrease by one wavelength of the sound wave. We can use phase changes $\Delta\theta$ to determine the movement direction and calculate the real-time relative moving distance Δd of the receiver. However, the phase change is very vulnerable to the interference of multipath effects in indoor environments, which could lead to incorrect measurement. We propose a novel multipath effect mitigation method to reduce the impact of multiple paths, and improve the performance of the phase based measurement.

D. Calculating Phase change using Park transformation

Before phase based ranging, we need to calibrate the phase offsets between sender and receiver due to their asynchronous system clocks. Even when the receiver is static, the received signal frequency is slightly different from the sender because the asynchronous system clocks lead to the linear increases on the phases of received signals. The phase offsets will further bring errors to distance measurement. The increases of phase offsets are slow and difficult to be measured by frequency analysis methods, *e.g.,* Fast Fourier Transform (FFT). Inspired by the phase calibration method used to calibrate the phase offsets between the WiFi wireless cards [32], we use the similar way to calculate the acoustic phase offsets between the receiver and source.

Then, we need calculate the phase changes of received signals. Some works calculate the phase change by multiplying the signal with orthogonal signals [17] [18]. The method brings in high frequency noise that need to be removed by low pass filter, which increases the computation and causes additional delay. Vernier [9] proposes a lightweight phase change calculating method which uses the number of maximum value in a moving window to calculate the number of cycles in the window. However, each cycle doesn't always have an obvious maximum value due to the superposition of signals from different paths, so the method has limited application in practical systems. In this subsection, we propose a phase change calculating method using *Park* transformation, which doesn't introduce high frequency noise and has low latency.

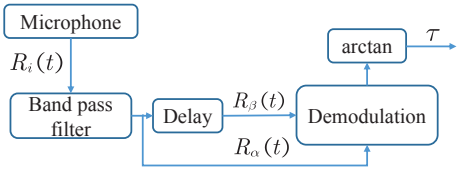


Fig. 2: Calculating phase change using *Park* transformation.

We now give an overview of PAMT when a source transmits the signal at a single acoustic frequency, namely $A\cos(2\pi f_c t)$. A is the amplitude of sound and f_c is the frequency of the sound. The mobile device obtains acoustic signal $R_i(t)$ from its microphone, as the MIC block in Figure 2. In order to measure the phase change at each frequency independently, we use a Band Pass Filter (BPF) with narrow band which could pass the signals around center frequency f_c, and rejects signals at other frequencies, as shown in the figure. We delay the filtered signal R_α by a quarter of fundamental-wave period. The delayed signal R_β is orthogonal to R_α,

$$\begin{aligned} R_\alpha(t) &= A'\cos(2\pi f_c t - \tau), \\ R_\beta(t) &= A'\sin(2\pi f_c t - \tau). \end{aligned} \tag{3}$$

Where A' is the amplitude of the signal after transmission attenuation and filtering. τ is the phase delay due to sound's propagation from source to receiver at time t. R_α and R_β can be further used to estimate the impact of multipath effects in Section III-E. We calculate the phase change using *Park* transformation [30]. After multiplying a transformation matrix $\mathbf{P}$ to R_α and R_β,

$$\begin{bmatrix} R_d(t) \\ R_q(t) \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} R_\alpha(t) \\ R_\beta(t) \end{bmatrix} \tag{4}$$

where $\mathbf{P}$ is the *Park* transformation matrix

$$\mathbf{P} = \begin{bmatrix} \cos(2\pi f_c t) & \sin(2\pi f_c t) \\ -\sin(2\pi f_c t) & \cos(2\pi f_c t) \end{bmatrix}.$$

We can obtain two based band signals without the carrier frequency f_c component, corresponding to R_d and R_q:

$$\begin{aligned} R_d(t) &= A'\cos(\tau), \\ R_q(t) &= -A'\sin(\tau). \end{aligned} \tag{5}$$

Then, we calculate the phase delay τ at time t using inverse tangent transformation, and the phase change in each frame.

E. Combating Multipath using MECF

In practical system, due to the reflection of wall, floor, and other objectors, the received signal is a superimposition of the LoS signals and the reflected signals. However, the influence of multipath effects in active ranging system is neglected in prior works [7], [9]. According to our extensive experiments, the multipath effects should be taken into consideration in active tracking system. The multipath effects would lead to periodic attenuation of the received signals, which brings error in the estimation of moving distance.

Suppose that the received signals are the superimposition of signals from N paths and each path has different delay and attenuation. In the paths, the i-th signal $R_i(t)$ has delay τ_i and amplitude a_i. Then, the receiver signal $R(n)$ is

$$R(t) = \sum_{i=1}^{N} R_i(t). \tag{6}$$

where $R_i(t) = a_i\cos(2\pi f_c t - \tau_i)$. Thus, it is difficult to obtain the actual phase change from the superimposition of signals traveled from different paths. To address this issue, we use a speaker to transmit the signal at different frequencies. The receiver could measure the phase change at each frequency independently using the BPF. The signals at different frequencies are transmitted through the same multiple paths to the receiver. We leverage the fact that the multiple effects don't always affect the phase based measurement of different frequencies at the same time due to their different wavelengths and phases. Thus, in each frame, we can use the phase change without affected by multiple paths to estimate the moving distance.

So how to determine which signal's phase isn't affected by the multipath effects in each frame? Previous work [17] uses linear regression to remove abnormal estimation at certain frequencies and calculates the distance change using the remaining frequencies. However, this method needs high bandwidth to ensure good regression results, and couldn't been used in our system due to limited bandwidth in each speaker. We propose a novel method to define whether a signal's phase is affected by the multipath effects in each frame. We call

this Multipath Effect Combating in Frame (MECF). First, we calculate normalized signal's trace diagram in a frame using R_α and R_β as coordinate axis, as shown in Figure 3c and Figure 3d.

Lemma 1. *When the device is static, the trace of signal at each frequency is a circle.*

Proof: It is simpler to treat this case in complex form

$$\begin{aligned} R(t) &= \sum_{i=1}^{N} R_i(t) = \mathrm{Re}\left(\sum_{i=1}^{N} a_i e^{j(2\pi f_c t - \tau_i)}\right) \\ &= \mathrm{Re}\left(e^{j2\pi f_c t}\left(a_1 e^{-j\tau_1} + a_2 e^{-j\tau_2} ... + a_N e^{-j\tau_N}\right)\right) \\ a_0 e^{-j\tau_0} &= \left(a_1 e^{-j\tau_1} + a_2 e^{-j\tau_2} ... + a_N e^{-j\tau_N}\right) \\ R(t) &= \mathrm{Re}\left(a_0 e^{j(2\pi f_c t - r_0)}\right) = a_0 \cos\left(2\pi f_c t - r_0\right). \end{aligned} \tag{7}$$

The superimposition of signals from multipath paths is a cosine signal. a_0 is the amplitude and τ_0 is the phase of signal. a_0 and τ_0 are related to a_i, and τ_i, and independent with time and carrier frequency. When the device is static, the received signals travel from static paths, and a_i and τ_i are fixed for each path, so that a_0 stays the same. As a result, $R_\alpha^2 + R_\beta^2 = a_0^2 \cos^2(2\pi f_c t - r_0) + a_0^2 \sin^2(2\pi f_c t - r_0) = a_0^2$, the trace has fixed radius a_0 for each frequency and is a circle. ■

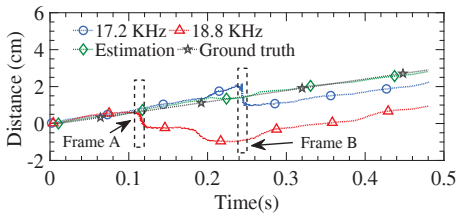
Lemma 2. *When the device moves, the trace of signal at each frequency is a circular ring. The width of the ring indicates the impact of multipath effects.*

Proof: The radius of the trace is a function of a_i, and τ_i. The length of multiple paths would change during the movement, which leads to the change of a_i and τ_i. As a result, the radius a_0 changes during the movement, and the trace is a circular ring for each frequency. If the impact of multiple paths is small in the frame, the width of ring is small, because the amplitude of signals changes smoothly due to the smooth propagation loss of the LoS path. Thus, a_0 changes slowly in the frame. When the impact becomes serious, the width of ring becomes large due to signal's small scale fading caused by multiple paths. The small scale fading causes a rapid change of a_0 in the frame. Thus, we can use the inter radius of ring to estimate the impact of multiple paths. ■

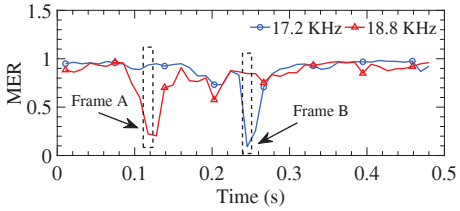
Due to the path loss and frequency selective fading, the received signals at different frequencies have different attenuation, so we can't compare the width of ring between them directly. Then, we define a metric to evaluate the impact of multiple paths. We design a parameter σ_j that we refer to as a *Multipath Effect Ratio (MER)* for the measured distance at frequency $f_j, j \in \{1,2,...,k\}$. σ_j captures the impact of multipath effects in the frame. We define σ_j as

$$\begin{aligned} D &= f(R_{\alpha,j}, R_{\beta,j}) \\ \gamma_{mean} &= g_{mean}(D) \\ \gamma_{inter} &= g_{inter}(D) \\ \sigma_j &= \frac{\gamma_{inter}}{\gamma_{mean}}. \end{aligned} \tag{8}$$

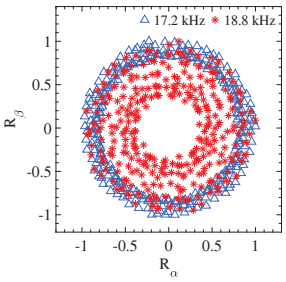
Where $R_{\alpha,j}$ and $R_{\beta,j}$ are the R_α and R_β in each frame at frequency j. D is the signal's trace diagram deriving from $R_{\alpha,j}$ and $R_{\beta,j}$. g_{mean} and g_{inter} calculate the mean radius γ_{mean} and inter radius γ_{inter} of the ring in D. The larger σ_j means that the inter radius is close to the outer radius, and indicates less impact of multipath effects. Finally, the frequency with the largest *MER* is selected to estimate the moving distance in each frame.



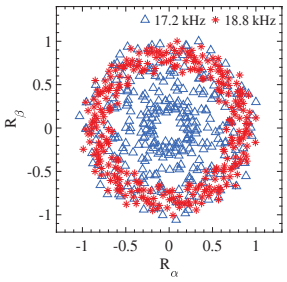
(a) The distance measured by the signal at different frequencies and the estimated distance using MECF



(b) The *MER* of the signal at different frequencies are used to evaluate the impact of multipath effects in real-time.



(c) The normalized traces of Frame *A*



(d) The normalized traces of Frame *B*

Fig. 3: In order to mitigate multipath effects, we propose a novel method to determine the impact of multiple paths on the signal at different frequencies, and select the signal with the largest *MER* to estimate the moving distance in each frame.

An example to illustrate MECF is shown in Figure 3. A speaker continuously transmits acoustic signals at 17.2 KHz and 18.8 KHz at the same time. Then, we move an Android smartphone away uniformly from the speaker at the distance of 1 meter. The smartphone obtains the signals frame by frame, and each frame has 512 sampling points with 48 KHz sampling rate. PAMT calculates the *MER* in each frame, as shown in Figure 3b. During the movement, we measure the moving distance using the method in Section III-F at each frequency, and the moving distance at each frequency should be the same in each frame if there are no multipath effects.

However, due to multipath effects, the measured distances are obviously different in Frame A of the signal at 18.8 KHz, and in Frame B of the signal at 17.2 KHz, as shown in Figure 3a. In Frame A, the *MER* of the signals at 17.2 KHz and 18.8 KHz corresponding to $\{0.9, 0.53\}$. Thus, we select the signal at 17.2 KHz to estimate the moving distance due to the larger *MER* in the frame. In Frame B, the *MER* of the signals at 17.2 KHz, 18.8 KHz corresponding to $\{0.24, 0.78\}$ so that the signal at 18.8 KHz is chosen to estimate the distance. We compare our estimation with the ground truth distance, and the result indicates that MECF could evaluate and mitigate the impact of multipath effects.

F. Robust Phase Based Ranging

After mitigating the multipath effects, we use the phase change of the selected signal in each frame to determine the LoS path length change. Due to the selected signal with largest *MER*, the measurement is robust to multipath effects. When the receiver moves close/away, the phase of acquired signal would increase/decrease. As the phase of the signal increases by 2π, the path length would decrease by one wavelength of the sound wave. We can use phase changes of the signal with the largest *MER* to estimate the distance change, and calculate relative moving distance Δd_{frame} in the frame

$$\begin{aligned} \Delta d_{frame} &= -\frac{\Delta\theta + 2\pi k}{2\pi}\lambda \\ d &= \sum \Delta d_{frame}. \end{aligned} \tag{9}$$

Where λ represents sound length of the selected signal. $\Delta\theta$ denotes the wrapped phase change relative to initial phase, and *k* is an integer which denotes the number of phase wraps during the movement. When the phase varies from π to $-\pi$ / from $-\pi$ to π, *k* increases/decreases by 1. We can use the distance change to calculate the moving speed and direction. Multipath effects have the least impact on the selected signal so that the measurement is robust to multipath fading.

IV. Motion Tracking Algorithm

In this section, we first use a calibration scheme to obtain a reference position. Then, we combine the initial position with the fine-grained distance change to enable 2-D motion tracking.

A. Estimating Reference Position

The phase based algorithm in Section III-F only measures the relative distance change, which is not sufficient for motion tracking. We cannot determine actual position only using the relative distance change due to the lack of the initial position. One way is to measure using some additional tools, such as ruler, hand-hold distance finder. However, those methods are cumbersome and error-prone. In this subsection, we propose a calibration method to estimate reference position. We choose two of the speakers assigned with different frequencies to estimate the initial position. Without loss of generality, we assume that two speakers A and B are placed along an x-axis. The coordinates of A and B are given, as shown in Figure 4.

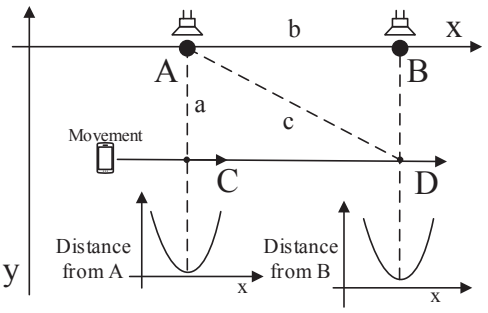


Fig. 4: Estimating reference position by moving a smartphone parallel to the x-axis.

We let a user move a smartphone parallel to the x-axis with unknown distance a. As the mobile is moving close/away each speaker, the distance between the smartphone and the speakers reduces/increases. When smartphone moves to the C/D position, the relative distance between A/B and the smartphone has minimum value. So we can determine the position of C and D on the moving path when the distances become minimum. We could calculate the difference d_{ac} of the relative distance between the smartphone and speaker A at C and D. Due to the property of a rectangular triangle, $a^2 + b^2 = c^2$ and $c = d_{ac} + a$, the distance a can be calculated by

$$a = \frac{b^2 - d_{ac}^2}{2d_{ac}} \tag{10}$$

Where b denotes the absolute distance between A and B, and c denotes the absolute distance between A and D. Thus, the coordinates of C and D can be obtained, and can be used as the reference position. To improve the accuracy, we can sweep the smartphone along the moving path multiple times, and use the mean positions as the estimation for C and D.

B. Tracking Motion by Computing Real-time Position

In order to track the device's motion, we should obtain the real-time position of the devices.

The range measurement between the smartphone and the *i*-th (i = 1,2,3,...,N) speaker is denoted as $\hat{d}_i$, where N is the number of speakers. Let $[x, y]^T$ be the unknown coordinate of the smartphone, and let $[x_i, y_i]^T$ be the known coordinate *i*-th speaker. The known coordinate of the reference position introduced in the previous subsection is $[x_R, y_R]^T$.

The error-free distance change between the smartphone and *i*-th speaker relative to the distance at reference position is calculated as

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} - R_i \tag{11}$$

where $R_i = \sqrt{(x_R - x_i)^2 + (y_R - y_i)^2}$. The measurements of the range differences are modeled by

$$\hat{d}_i = d_i + \varepsilon_i, \qquad i = 1, 2, 3, ..., N \tag{12}$$

where ε_i is the measurement error of $\hat{d}_i$.

The LS error function is then defined as the difference between the measured and true values

$$\mathbf{e} = \hat{\mathbf{d}} - \mathbf{d} \tag{13}$$

where $\hat{\mathbf{d}} = \left[\hat{d}_1, \hat{d}_2, ..., \hat{d}_N\right]^T$ and $\mathbf{d} = [d_1, d_2, ..., d_N]^T$.

We define vector $\mathbf{\Lambda} \triangleq \left[x \;\; y \;\; x^2 + y^2\right]^T$. Then, we can rewrite Equation 13 in matrix form as

$$\mathbf{e} = \mathbf{A}\mathbf{\Lambda} - \mathbf{b} \tag{14}$$

where

$$\mathbf{A} = [v_1, v_2, ..., v_i, ..., v_N]^T, v_i = \left[-2x_i \quad -2y_i \quad 1\right],$$
$$\mathbf{b} = [b_1, b_2, ..., b_i, ..., b_N]^T, b_i = \left(\hat{d}_i + R_i\right)^2 - x_i^2 - y_i^2,$$
$$i = 1,2,...,N.$$

According to LS algorithm, the solution of minimizing $\mathbf{e}$ is given by

$$\mathbf{\Lambda} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{15}$$

The unknown coordinate of the smartphone can be expressed as

$$[x, y]^T = \left[\mathbf{\Lambda}\left(1\right), \mathbf{\Lambda}\left(2\right)\right]^T. \tag{16}$$

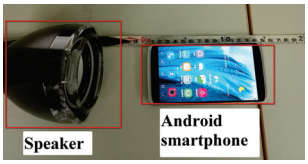
Note that the smartphone's coordinate is updated in each sampling time, and we can track the smartphone's motion using the real-time coordinate.

V. IMPLEMENTATION

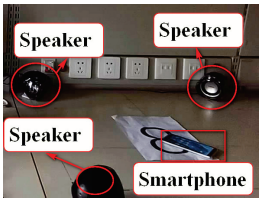
We implement PAMT on a standard Android platform. We conduct experiments on a GIGABYTE Z77X-UD3H desktop with Intel I7 CPU and 8 GB RAM. The desktop has a VIA sound card which could support at most 8 speakers. Some PHILIPS SPA311/93 speakers ($8 each) connect to the desktop and transmit audio signals at different frequencies. PHILIPS SPA311/93 can transmit acoustic signals up to 23 KHz in practical usage. We develop PAMT on a ZTE B2015 mobile phone with Andriod 5.1 Operating System, and a speaker connected with the desktop, none of them are installed additional hardware. We use the speakers to transmit inaudible acoustic signals, each at certain frequency. Then, we use the bottom microphone of the smartphone to record the sound wave with the sampling rate of 48 KHz, which is supported by most smartphones. Inspired by the task scheduling in fog networks [28] [29], we offload the computation of ranging and tracking to the desktop in order to extend batter lifetime.

VI. EVALUATION

For 1-D ranging, we first evaluate the average movement distance error, the impact of multipath effects reduced by MECF. Then, we compare the performance of different multipath mitigation methods, and the performance of MECF using different number of carriers. A speaker connected with the desktop is implemented as an audio source. The Android smartphone is used as a mobile receiver. We move the smartphone close/away the audio source, and record the distance change during the movement. Since the distance between them changes over time, we use a rule along the moving path that collects the ground truth data, as shown in Figure 5a.



(a) Testbed setup for 1-D ranging



(b) Testbed setup for the performance evaluation of motion tracking.

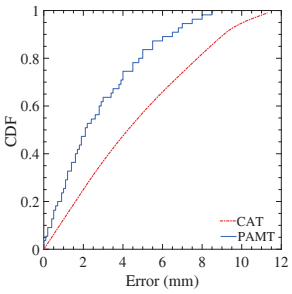
Fig. 5: Experimental Setup

For 2-D tracking, we use a three-speaker system as shown in Figure 5b. The separation between adjacent speakers is 70 centimeters. Each speaker is allocated 0.8 KHz bandwidth with 200 Hz interval. The three-speaker system occupies 17.2-19.4 KHz, which are virtually inaudible to most people. The remaining band-width (from 19.4 KHz to 23 KHz) retains for future works, such as assigning to extra speakers to improve tracking performance. We evaluate the performance of motion tracking by evaluating the median error between the trajectories tracked by PAMT versus the ground truth trajectories recorded by a camera. The ranging, tracking, and visualization, are both done on-line in real-time.

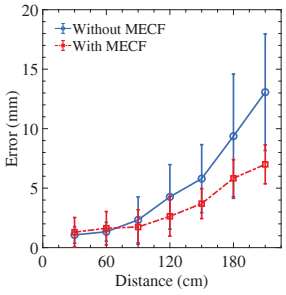
A. Experimental Results

a) The accuracy of 1-D ranging: PAMT achieves an average movement distance error of 2 mm when the smartphone moves for 40 cm at a distance of 1 m. In our experiments, we use PAMT to measure the distance changes between a smartphone and a speaker, and use a rule to collect ground truth data. The smartphone's initial position is 1 m away from the speaker, and moves away from the speaker for a distance of 40 cm. Figure 6a plots the Cumulative distribution function (CDF) of the 1-D relative distance measurement error for 100 measurements. The median error is 2 mm, and 90-th percentile error is 6 mm. We also compare our approach with CAT, a moving distance measurement approach based on FMCW [7]. Results show that our approach outperforms CAT in terms of distance measurement accuracy by 50% on average.

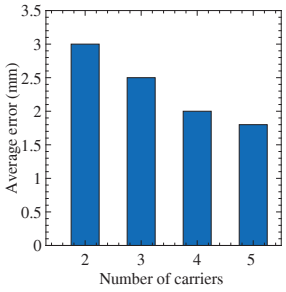
b) Impact of multipath effects: In this experiment, we examine the impact of multipath effects to our measurement. Particularly, we conduct following experiments: we use the speaker to play the sinusoidal audio signals at four frequencies. We first estimate the moving distance using the signal at one of the frequencies without using MECF, and using MECF with four carriers, respectively. Then, we calculate the measurement error at different distances. The comparisons of the results (95% confidence interval) from the estimation without/with MECF are shown in Figure 6b. The median error increases as the distance increases because the multipath effects are strengthened as the transmission distance increases. Nevertheless, we observer that MECF reduces median error from



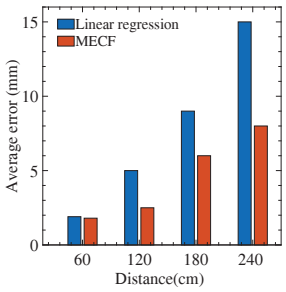
(a) CDF for 1-D ranging



(b) The impact of multipath effects reduced by MECF

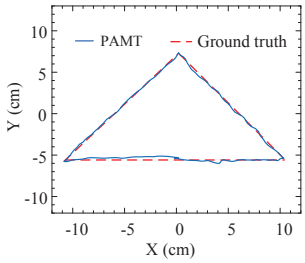


(c) Median error of different number of carriers used in MECF

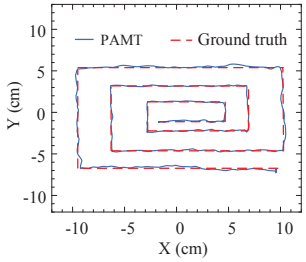


(d) Median error of using different multipath effect mitigration methods

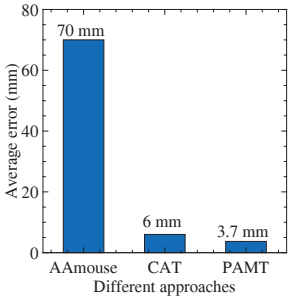
Fig. 6: The accuracy of 1-D ranging, and the performance of MECF



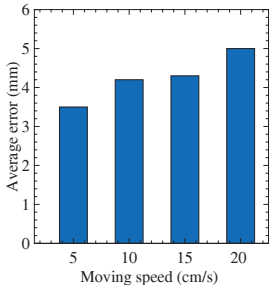
(a) A trajectory of triangle



(b) A trajectory of loop back



(c) Median error for motion tracking using different approaches



(d) Median error at different moving speed

Fig. 7: Tracking performance.

9.5 mm to 5.5 mm at the distance of 180 cm, and is very effective for variance reduction. Overall, MECF can improve the ranging accuracy and the robustness against effects of multipath. We also evaluate the influence of the number of carriers at a distance of 1 meter, as shown in Figure 6c. The carriers are transmitted by the same speaker with the same amplitude. We can see that error reduces as the number increases. When four carriers are used, the error reduces to 2 mm. The results show the robustness and feasibility of MECF under multipath effects.

c) The performance of MECF: We compare the performance of MECF with the linear regression based method proposed in [17]. A speaker transmits four carriers, and we calculate the median error of ranging at different distances, as shown in Figure 6d. MECF and linear regression method have similar performance at 60 cm because multipath effects have limited impact at the distance. As the distance increases, the performance of linear regression reduces significantly because it requires wider bandwidth to ensure good regression results, while MECF has more stable and robust performance. We observe that when the distance is 2.4 meters, the linear regression based method has a median error of 15 mm, and cannot mitigate multipath effects effectively. The results indicate PAMT can achieve good ranging performance with a limited bandwidth of acoustic signals (*e.g.*, 0.8 KHz), which ensures great robustness for practical usage on mobile devices.

d) The accuracy of motion tracking: We use a PAMT-enabled smartphone to draw a triangle, and a loop back in 2-D space. We use a camera to collect the ground-truth trajectories of the smartphone. We compare the measured trajectories versus the original trajectories to quantify the accuracy. Figure 7a and 7b show the sample results of the estimated trajectories and ground truth trajectories. We compare the median error of tracking with CAT and a Doppler effect based method [15] (denoted by AAMouse), as shown in Figure 7c. The median error of PAMT is 3.7 mm, while that of CAT and AAmouse are 6 mm and 7 cm respectably. The results show that the tracking accuracy of PAMT out-performs CAT and AAMouse significantly.

e) The impact of moving speed: We move a smartphone along straight line for 20 cm at different speed, and track the motion of the smartphone using the three-speaker system. The median errors of the trajectories are shown in Figure 7d. We observe that the median errors are in the range of 3-5 mm at different speed. As the speed increases, the median error increases slightly, because the fast changing phase caused by high speed movement leads to frequency shift. The frequency shift may be seen as an interference from near frequencies, and restrained by band-pass filter, which would reduce tracking accuracy. Nevertheless, the results indicate PAMT could accurately track the motion at different speeds, which are not easily detected by Doppler effect based methods on mobile devices.

VII. LIMITATIONS

In this section, we discuss some limitations of our current system and potential directions for future work. First, some limitations of hardware and system bring extra latency to track the real time position. The minimum of buffer size in Android platform is 512 samples which would introduce at least 10

milliseconds delay. Further, in order to extend battery life, the smartphone offloads its computation to a desktop using Wi-Fi network with UDP protocol, which also brings in a few milliseconds delay. In future work, we plan to balance the computation and energy consumption, and make an optimal solution for the trade-off between latency and battery lifetime.

Second, the total bandwidth of sound is limited on smartphone. In PAMT, each speaker is allocated 0.8 KHz bandwidth separated by 200 Hz guard band to avoid carrier interference. The guard band and the band-pass filter used in our system limit the maximum movement speed (about 30-40 cm/s) of devices that could be accuracy tracked. In future work, we plan to reduce the limitation of speed by improving bandwidth. For example, we plan to use some commercial microphones to replace the speakers as receiver, and a smartphone as source, so that different degrees of freedom can share the total bandwidth.

VIII. Conclusions

This paper presents the system design of PAMT, which provides a fine-grained mobile interaction solution for commercial mobile devices. We propose a phase based scheme to measure the distance change between the source and receiver using multiple signals at single frequency. And we also propose a novel method to select the signal suffering from the least multipath fading and calculate moving distance, then improve the performance of our approach in multipath fading environments. In this way, we implement a prototype of PAMT using a commercial smartphone and some speakers. The prototype could position and track a smartphone with mm-level accuracy. We conduct systematic evaluation based on the prototype. Experiment results validated our idea as well as the system design.

References

[1] K. Joshi, S. Hong, and S. Katti, "Using sound source localization in a home environment," *In Proceedings of USENIX NSDI'13*, vol. 4, pp. 241–254, USENIX Association, 2013.

[2] J. Xiong, K. Sundaresan, and K. Jamieson, "Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization," *In Proceedings of MobiCom'15*, pp. 537–549, ACM, 2015.

[3] S. Zhu and X. Zhang, "Enabling high-precision visible light localization in today's buildings," *In Proceedings of MobiSys'17*, pp. 96–108, ACM, 2015.

[4] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: Light-weight indoor positioning with visible light," *In Proceedings of MobiSys'15*, pp. 317–330, ACM, 2015.

[5] O. Abari, H. Hassanieh, M. Rodreguiz, and D. Katabi, "Poster: A millimeter wave software defined radio platform with phased arrays," *In Proceedings of MobiCom'16*, pp. 419–420, ACM, 2015.

[6] A. Olivier, G. Bielsa, I. Tejado, M. Zorzi, J. Widmer, and P. Casari, "Lightweight indoor localization for 60-ghz millimeter wave systems," *In Proceedings of SECON'16*, pp. 1–9, IEEE, June 2016.

[7] W. Mao, J. He, and L. Qiu, "Cat: High-precision acoustic motion tracking," *In Proceedings of MobiCom'16*, pp. 69–81, ACM, 2016.

[8] C. Zhang, Q. Xue, A. Waghmare, S. Jain, Y. Pu, S. Hersek, K. Lyons, K. A. Cunefare, O. T. Inan, and G. D. Abowd, "Soundtrak: Continuous 3d tracking of a finger using active acoustics," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 30:1–30:25, June 2017.

[9] Y. Zhang, J. Wang, W. Wang, Z. Wang, and Y. Liu, "Vernier: Accurate and fast acoustic motion tracking using mobile devices," *In Proceedings of INFOCOM'18*, IEEE, 2018.

[10] C. Peng, G. Shen, and Y. Zhang, "Beepbeep: A high-accuracy acousticbased system for ranging and localization using cots devices," *ACM Trans. Embed. Comput. Syst.*, vol. 11, pp. 4:–4:29, Apr. 2012.

[11] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," *In Proceedings of MobiCom'00*, pp. 32–43, ACM, 2000.

[12] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. P. Martin, "Detecting driver phone use leveraging car speakers," *In Proceedings of MobiCom'11*, pp. 97–108, ACM, 2011.

[13] Z. Sun, R. Bose, and P. Zhang, "Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing," *GetMobile: Mobile Comp. and Comm.*, vol. 18, pp. 11–14, Jan. 2015.

[14] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, and Y. Liu, "Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones," *In Proceedings of INFOCOM'14*, IEEE, Apr. 2014.

[15] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," *In Proceedings of MobiSys'15*,pp. 15–29, ACM, 2015.

[16] K.-Y. Chen, D. Ashbrook, M. Goel, S.-H. Lee, and S. Patel, "Airlink: Sharing files between multiple devices using in-air gestures," *In Proceedings of UbiComp'14*, pp. 565–569, ACM, 2014.

[17] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," *In Proceedings of MobiCom'16*, pp. 82–94, 2016.

[18] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, "Strata: Fineg-rained acoustic-based device-free tracking," *In Proceedings of MobiSys'17*, pp. 15–28, ACM, 2017.

[19] R. Gao, H. Wang, D. Wu, K. Niu, E. Yi, and D. Zhang, "A model based decimeter-scale device-free localization system using cots wi-fi devices," *In Proceedings of UbiComp'17*, pp.241–244, ACM, 2017.

[20] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, and H. Mei, "Indotrack: Device-free indoor human tracking with commodity wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 72:1–72:22, Sept. 2017.

[21] M. Youssef and A. Agrawala, "The horus wlan location determination system," *In Proceedings of MobiSys'05*, pp. 205218, ACM, 2005.

[22] S. Chung and I. Rhee, "vtrack: Virtual trackpad interface using mm-level sound source localization for mobile interaction," *In Proceedings of UbiComp'16*, pp. 41–44, ACM, 2016.

[23] R. Jia, M. Jin, Z. Chen, and C. J. Spanos, "Soundloc: Accurate roomlevel indoor localization using acoustic signatures," *In Proceedings of CASE'15*, pp. 186193, IEEE, Aug 2015.

[24] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," *In Proceedings of MobiSys'11*, pp. 155–168, ACM, 2011.

[25] Y. Shu, C. Bo, G. Shen, C. Zhao, L. Li, and F. Zhao, "Magicol: Indoor localization using pervasive magnetic field and opportunistic wifi sensing," *IEEE J. Sel. Areas Commun.*, vol. 33, pp. 1443–1457, 2015.

[26] M. T. I. Aumi, S. Gupta, M. Goel, E. Larson, and S. Patel, "Doplink: Using the doppler effect for multi-device interaction," *In Proceedings of UbiComp'13*, pp. 583–586, ACM, 2013.

[27] D. Lindgren, G. Hendeby, and F. Gustafsson, "Distributed localization using acoustic doppler," *Signal Process.*, vol. 107, pp. 43–53, 2015.

[28] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo and M. Zhou, "MEETS: Maximal Energy Efficient Task Scheduling in Homogeneous Fog Networks," in IEEE Internet of Things Journal, vol. 5, no. 5, pp. 4076-4087, Oct. 2018.

[29] Y. Yang, "Multi-tier Computing Networks for Intelligent IoT," Nature Electronics, vol.2, pp. 4-5, Jan. 2019.

[30] A. Pigazo, V. M. Moreno, and E. J. Estbanez, "A recursive park transformation to improve the performance of synchronous reference frame controllers in shunt active power filters," *IEEE Trans. Power Electron.*, vol. 24, pp. 2065–2075, Sept 2009.

[31] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "BatTracker: High Precision Infrastructure-free Mobile Device Tracking in Indoor Environments," *In Proceedings of SenSys '17*, Rasit Eskicioglu (Ed.). ACM, New York, NY, USA, Article 13, 14 pages.

[32] J. Gjengset, J. Xiong, G. McPhillips, and K. Jamieson, "Phaser: enabling phased array signal processing on commodity WiFi access points," *In Proceedings of MobiCom '14*. ACM, New York, NY, USA, 153-164.