Addressing Practical Challenges in Acoustic Sensing To Enable Fast Motion Tracking

Yongzhao Zhang
Shanghai Jiao Tong University
Shanghai, China
zhangyongzhao@sjtu.edu.cn

Hao Pan
Shanghai Jiao Tong University
Shanghai, China
panh09@sjtu.edu.cn

Yi-Chao Chen
Shanghai Jiao Tong University
Shanghai, China
yichao@sjtu.edu.cn

Lili Qiu
Microsoft Research Asia
Shanghai, China & UT Austin
Austin, USA
liliqiu@microsoft.com

Yu Lu
Shanghai Jiao Tong University
Shanghai, China
yulu01@sjtu.edu.cn

Guangtao Xue
Shanghai Jiao Tong University
Shanghai, China
xue-gt@cs.sjtu.edu.cn

Jiadi Yu
Shanghai Jiao Tong University
Shanghai, China
jiadiyu@sjtu.edu.cn

Feng Lyu
Central South University
Changsha, China
fenglyu@csu.edu.cn

Haonan Wang
Shanghai Jiao Tong University
Shanghai, China
wanghaonan@sjtu.edu.cn

ABSTRACT

Motivated by many potential applications that could be enabled by acoustic motion tracking, in this paper we systematically examine the factors that limit the accuracy of acoustic tracking in practical scenarios. We identify three main challenges: (i) high mobility, (ii) low SNR, and (iii) hardware frequency response. We further show that the last two issues may exacerbate the performance issue under high mobility. We develop effective approaches to address the issues. In particular, to address high mobility, we tackle phase wrap-around using the derivative of the phase; we further estimate the Doppler shift under diverse scenarios and compensate the Doppler in channel impulse response (CIR). To address low SNR, we use a novel approach to estimate the phase shift between consecutive time intervals to effectively support time-domain beamforming and increase SNR. To tackle the uneven frequency response, we show that it is important to estimate and compensate the phase as well as the amplitude of the frequency response. Our extensive evaluation shows that each of our techniques is effective and putting them together significantly enhances the accuracy of acoustic motion tracking in general scenarios.

Yongzhao Zhang and Hao Pan did this work as interns at Microsoft Research Asia and Yi-Chao Chen did this work as a visiting researcher at Microsoft Research Asia.

IPSN '23, May 09–12, 2023, San Antonio, TX, USA

ACM ISBN 979-8-4007-0118-4/23/05…$15.00
https://doi.org/10.1145/3583120.3586954

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**.

KEYWORDS

acoustic motion tracking, fast motion tracking, SNR enhancement

ACM Reference Format:
Yongzhao Zhang, Hao Pan, Yi-Chao Chen, Lili Qiu, Yu Lu, Guangtao Xue, Jiadi Yu, Feng Lyu, and Haonan Wang. 2023. Addressing Practical Challenges in Acoustic Sensing To Enable Fast Motion Tracking. In *The 22nd International Conference on Information Processing in Sensor Networks (IPSN '23), May 09–12, 2023, San Antonio, TX, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3583120.3586954

1 INTRODUCTION

Background: Device free acoustic tracking technologies can enable a wide variety of interesting applications, such as motion based gaming, Augmented Reality (AR), Virtual Reality (VR), touchless user interface (UI) for IoT, breathing and heart-rate monitoring, and indoor localization.

Recently there have been a range of interesting device-free acoustic tracking algorithms proposed. They can be broadly classified into four categories: time-of-arrival (TOA) based [46, 51, 68, 77], Doppler based [4, 6, 17], FMCW based [45, 78], and phase based [32, 60, 66, 69, 74]. Among them, the resolution of the first three approaches is limited by sampling rate and bandwidth. In comparison, phase-based tracking is attractive since the phase has a much higher resolution (*e.g.*, $1mm$ movement results in 0.74 radian phase change in device-free tracking at $20KHz$).

Directly using the phase of received signal is vulnerable under multipath. In order to deal with multipath, Strata [74] proposes a novel technique that estimates the channel impulse response (CIR), which gives the channel coefficient of each channel tap, and uses the phase of an appropriate channel tap for motion tracking. Since then, there have been several interesting extensions (*e.g.*, [34, 60, 76]).

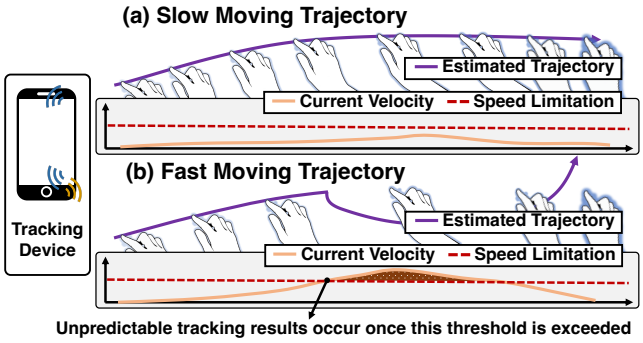


Figure 1: An example of the rapid motion problem. (a) When user's finger is moving with speeds below the limit, the phase-based tracking schemes can estimate the trajectory accurately. (b) Once the moving speed exceeds the limit, tracking accuracy degrades significantly.

Limitations of existing work: In this work, through extensive measurements we identify several limitations of existing phase-based motion tracking: (i) fast movement of target, (ii) low SNR, and (iii) frequency response of hardware.

The first issue is that current phase based tracking methods only works under slow motion. Specifically, as shown in Fig. 1(a), Strata [74] performs well most time, but incurs high errors and sometimes even moves in an opposite direction when the movement speed exceeds a limit (*e.g.*, $0.8m/s$) as shown in Fig. 1(b). Like Strata, other existing phase based tracking schemes, such as LLAP [69] and VSkin [60], have similar performance issues under high speeds: LLAP and VSkin can work only below $0.25m/s$ and $0.12m/s$, respectively. It is quite common to have motion exceed these speeds. [18, 30, 59] report a typical speed of human interaction locomotion ranges between $1.5-2m/s$ and the highest moving speed is around $2.7m/s$ [8]. To the best of our knowledge, no prior works on acoustic-based tracking can effectively support speed beyond $1m/s$.

The second issue is that the reflected signals decays rapidly over distance and leads to low SNR beyond $1m$ in a smartphone setup. Due to the error accumulation of phase [58], the tracking performance degrades significantly once the device moves outside $1m$. Previous works [41, 66] cannot be directly applied to the high mobility scenarios and usually require a microphone array for spatial beamforming (infeasible for COTS mobile phones).

Another issue is that the uneven frequency response of the speaker and microphone will distort the received signal. While the previous works (*e.g.*, [39, 42]) have reported that the frequency response at the inaudible acoustic frequency is highly heterogeneous and propose to compensate the amplitude of the frequency response, we find that the uneven frequency response involves both uneven amplitude and uneven phase, both of which lead to significant distortion to the received signals and cause degradation in tracking accuracy.

Our approach: We first closely examine the traces under fast movement and have the following observations: (i) Under fast movement, the phase change exceeds 2π between two consecutive updates, which results in ambiguity in the phase change. We propose a novel scheme based on the phase derivative to avoid the under-sampling issue. We formally show that the ambiguity of phase derivative is determined by the acceleration instead of the velocity and that the target's acceleration wraps around at a much higher threshold (well below the human's acceleration) than the velocity and hence is more robust against the under-sampling issue. (ii) Fast movement also introduces Doppler shift, which distorts the phase change. We develop a simple yet effective method to estimate and compensate for the Doppler shift in CIR on mobile devices in real-time.

To address the low SNR, inspired by time domain beamforming [41], we constructively add up CIR profiles to enhance SNR. However, due to movement, we must compensate for the phase change incurred by the movement between the two measurement periods before adding up. We develop a practical method to estimate and compensate for the phase change based on the maximum entropy with only one microphone.

Next we examine the impact of the hardware frequency response. A natural way to handle the frequency response is to measure the speaker's frequency response and compensate for the frequency response at either the transmitter or receiver side by multiplying the inverse of the frequency response so that the resulting amplitude is flat. This type of amplitude-based frequency response compensation has been used earlier. Interestingly, through our systematic examination, we show that the phase of the frequency response is also important to compensate and is less sensitive to noise. Hence we propose frequency response compensation using both phase and magnitude.

We implement the above techniques (named SwiftTrack) on Android phones to enable fine-grained and robust fast motion tracking with acoustic signals at $1.5m$ away. Our extensive evaluation validates the effectiveness of each individual technique. We further compare the performance of our system with the existing works including LLAP, Strata, and VSkin. Our results show that, when the moving velocity ranges from $5cm/s$ to $240cm/s$, the median error of the estimated absolute distance is $0.63cm$, outperforming Strata, VSkin, and LLAP by 253%, 327%, and 1114%, respectively. Our work develops systematic approaches to address several important practical challenges and uses real implementation to demonstrate its benefits.

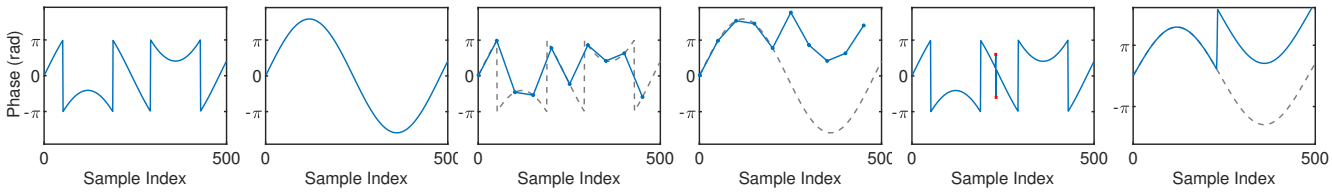
The rest of this paper is organized as follows. We give an overview to CIR-based motion tracking method in Section 2 and introduce three major challenges in motion tracking in Section 3. We describe our approach in Section 4. Note that SwiftTrack combines a series of algorithms to achieve fine-grained and robust fast motion tracking in the real-world. Then we evaluate its performance in Section 5, review related work in Section 6, discuss in Section 7, and conclude in Section 8.

2 PRELIMINARY

In this section, we provide an overview of CIR-based acoustic tracking schemes used in [60, 74] that can achieve high tracking accuracy and separate multiple targets with different delays.

2.1 Channel Estimation

The transmitter sends a known acoustic sequence for channel estimation at the inaudible band (*e.g.*, above $17kHz$). Many pseudo-random sequences (*e.g.*, GSM [15], Gold [62], Barker [13], and



(a) Phase unwrapping with no ambiguity (b) Insufficient phase samples (c) 'Fake jumps' caused by noise

Figure 2: Illustration of phase unwrapping concept and two major issues that cause the ambiguity. (a) The true phase (right) can be unwrapped losslessly from the wrapped phase (left) if there is no phase ambiguity. (b) The insufficient sampling rate of phase may cause ambiguity. (c) Noise may cause ambiguity.

Zadoff-Chu (ZC) [52]) can be used for channel estimation. We choose the ZC sequence as our training sequence since it is widely adopted in modern cellular systems, including LTE and 5G NR [3]. Note that our observations and solutions also apply to other sequences. A ZC sequence has the following expression:

$$ZC[n] = exp(-j\frac{\pi\mu n(n+1)}{N_{ZC}}) \tag{1}$$

where N_{ZC} is its sequence length, $0 \leq n < N_{ZC}$, $0 < \mu < N_{ZC}$ and $gcd(N_{ZC}, \mu) = 1$.

To make the acoustic signals inaudible, we need to fit its signal within the inaudible band. First, we perform a fast Fourier transform (FFT) to convert the time-domain root ZC sequence to the frequency domain and pad zeros to keep its bandwidth within the inaudible band (*e.g.*, $17 - 23KHz$). Then, we perform the inverse fast Fourier transform (IFFT) to convert it back to the time domain. The length of the baseband signal $ZC_T[n]$ is $N = N_{ZC} \times \frac{f_s}{BW}$, where f_s is the sampling rate. Finally, we up-convert the base-band signal to the inaudible band by multiplying it by $exp(2\pi f_c t)$, where f_c is a central frequency (*e.g.*, $20KHz$ when $17 - 23KHz$ is used).

After capturing the reflected acoustic signal by the receiver (*e.g.*, the built-in microphone on a smartphone), we perform down-conversion to obtain the base-band signal ($ZC_R[n, t]$). Because the receiver is a Linear Time-Invariant (LTI) system, the received base-band signal can be modeled as:

$$ZC_R[t] = \sum_i A_i(t) ZC_T[t - \tau_i(t)] = h[t] * ZC_T[t] \tag{2}$$

where $*$ denotes the convolution operator, i denotes the index of the propagation path, and A_i and τ_i are the channel attenuation and delay for the i-th propagation path, respectively.

Let $h[n]$ denote the discrete output of $h[t]$, which can be expressed as:

$$h[n] = \sum_i A_i(n)\delta[n - \tau_i(n)] \tag{3}$$

where $\delta[n]$ is a discrete Dirac's delta function [48], which is non-zero only when $n = \tau_i(t)$. Evidently, the phase change at the target tap is proportional to the delay change of the target, making it possible to track the moving objects using the phase of CIR. Following the existing works [60, 76], we estimate the channel response $h[n]$ by correlating the received signal $ZC_R[n]$ with the transmission signal $ZC_T[n]$.

Before estimating the distance of the target, we first remove the background multipath, including the direct path from the speaker to microphone and reflections from static surroundings. There are many background removal algorithms in literature, such as LEVD [69], DDBR [70], and direct subtraction [32]. We use the direct subtraction for simplicity. We measure the background interference when there is no target and then subtract it from the estimated channel ($h[n]$) when there is a target, similar to the methodology in [32, 41].

2.2 Distance Estimation

Coarse-grained ToF estimation: The channel tap $\hat{n}$ that maximizes the magnitude of the CIR corresponds to the delay of the signal. Therefore, we can derive the delay based on $\hat{n}$. However, due to the limited audio sampling frequency, noise, and multipath interference, this estimation is inaccurate. So we only use it to provide a coarse estimation of the initial position.

Fine-grained displacement estimation: The phase of the selected tap is more accurate and provides a fine-grained displacement estimation of the target. According to [60, 74], the displacement can be derived from the phase as follow:

$$\Delta dist = \frac{c}{f_c}\frac{\Delta Phase}{2\pi} \times \frac{1}{2} \tag{4}$$

where the result is divided by 2 because we consider the round-trip flight in a device-free system.

We further combine the coarse-grained initial position and fine-grained displacement to derive the current position [60, 74].

3 CHALLENGES

3.1 Consequences of Fast Movement

Fast movement introduces two issues: phase ambiguity and Doppler shift, which will be elaborated below.

3.1.1 Phase Unwrapping and Phase Ambiguity. Our measured phase from channel $h[n]$ in Eq. 3 is wrapped around and constrained to $(-\pi, \pi)$. Phase unwrapping refers to recovery of the original phase value from the wrapped phase value [11]. Let us consider the basic 1-dimensional phase unwrapping problem. As shown in Fig. 2(a), when the phase is wrapped around, we should reconstruct the correct phase by removing the "phase jumps". Itoh [26] shows that the original phase can be reconstructed as long as the smoothness condition is satisfied:

$$|\Delta\phi_n| \leq \pi \tag{5}$$

Refer to [11, 26] for the detailed phase unwrapping algorithm. Fig. 2(b) shows that if there are no sufficient phase samples, the true phase change might be larger than π, leading to the failure of phase

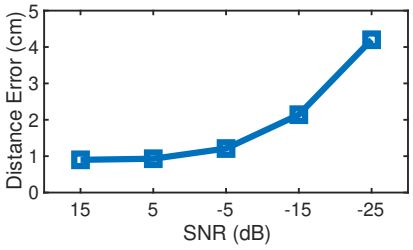
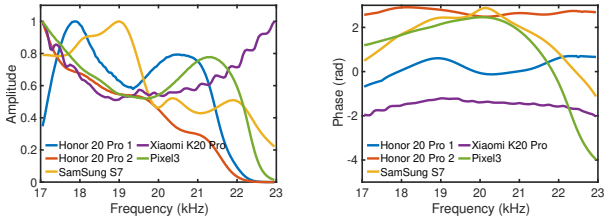


Figure 3: Accuracy change with the decrease of SNR



(a) Amplitude response. (b) Phase response.

Figure 4: Frequency response of 5 different devices.

unwrapping. Fig. 2(c) shows that the sudden phase change caused by noise may also introduce phase ambiguity. The fast movement problem will lead to an insufficient sampling rate of phase measurement, which will result in phase ambiguity. The noise may also introduce phase ambiguity due to a sudden phase change.

Many advanced phase unwrapping techniques, like Path Following Algorithm (PFA) [10, 22, 23] and Quality-Guided Algorithm (QGA) [10, 29, 35], have been widely adopted to avoid phase ambiguity. However, they are primarily two-dimensional phase unwrapping techniques, which rely on the carefully selected unwrapping path to bypass the damaged regions (i.e., the regions with ambiguity). For two-dimensional phase data, ambiguity can be effectively detected with closed path loops [10]. On the contrary, for one-dimensional phase data, we cannot determine damaged regions with closed loops and have only one unwrapping path, which is more challenging with the presence of ambiguity.

3.1.2 Doppler Effect. Assuming a target is moving at a speed of $2m/s$ and the frequency of sound waves is $20kHz$, it will cause a frequency shift of around $233Hz$ in a round-trip. If the speaker sends periodical signals with period $10ms$, the spacing of each frequency will be $100Hz$. That is, a $2m/s$ movement will cause the received signal to have non-negligible frequency shift (*i.e.*, $233Hz$), which in turn affects the down-conversion and leads to distortion in the channel response $h[n]$. Thus, this pulse distortion introduces significant uncertainty in tap selection and phase measurement.

3.2 Low SNR

It is well known that the SNR has significant impact on the tracking accuracy. For phase based methods, the error will be accumulated because we have to integrate the phase change over time to get the distance estimation [58]. Figure 3 shows that when the SNR decreases from $15dB$ (corresponding to a distance at $30cm$ between the target and the phone) to $-25dB$ (corresponding to $150cm$), the distance estimation error of Strata increases significantly from $0.93cm$ to $4.25cm$.

3.3 Hardware Frequency Response

For commodity mobile devices, the frequencies above $15kHz$ are hardly audible and are not optimized. Their speaker and microphone have uneven frequency response. The uneven response introduces significant distortion to signals and errors to the generated channel impulse response in Eq. 3. Several previous works (*e.g.*, [39, 40]) have identified and compensated for the uneven amplitude across the frequencies. Our work identifies the uneven phase across different frequencies. In particular, as shown in Figure 4, the phase changes significantly across the frequencies. Such a large change in the phase introduces tracking errors. Therefore, it is important to measure and compensate for the uneven phase as well as the uneven amplitude.

4 OUR APPROACH

In this section, we describe our approaches to address fast movement, low SNR, and uneven frequency response.

4.1 Fast Movement

Existing frame-based methods work well under slow motion, but degrade under fast movement. In this section, we show fast movement introduces two issues in motion tracking: (i) phase ambiguity and (ii) Doppler shift. Then we present our methods to address both challenges.

4.1.1 Phase Ambiguity Caused by Fast Movement. The small wavelength of inaudible acoustic signal enables fine-grained motion tracking. On the other hand, it also means that the phase can easily wrap around. A displacement of only $1.73cm$ will cause the phase at $20k$ Hz to change by 2π, which leads to the phase wrap-around. To derive the actual displacement, we should unwrap the phase. To avoid phase ambiguity, we should ensure that the phase change satisfies the smoothness condition in Eq. 5.

The phase measurement $\phi_d[t]$ is related to the distance $d[t]$ as follow: $\phi_d[t] = -\frac{2\pi f_c}{c} d[t] \times 2 = -\frac{4\pi f_c}{c} d[t]$. To satisfy the smoothness condition, we have:

$$
\begin{aligned}
|\phi_d[t] - \phi_d[t-1]| &\leq \pi \\
\frac{4\pi f_c}{c} |d[t] - d[t-1]| &\leq \pi \\
\frac{4 f_c T}{c} \left| \frac{d[t] - d[t-1]}{T} \right| &\leq 1 \\
\frac{4 f_c T}{c} |v[t]| &\leq 1 \\
|v[t]| &\leq \frac{c}{4 f_c T} \qquad (6)
\end{aligned}
$$

Considering $c = 343m/s$, $f_c = 20kHz$, and $T = 10ms$, the maximum velocity during this interval should satisfy: $|v[t]| \leq 0.43m/s$. Hand movement can reach $2.7m/s$ [8]. Therefore the phase change may easily exceed π in this case, and cause distance estimation to be off. **Fast movement may cause phase measurements to be under-sampled and introduces ambiguity in the phase**

change. Actually, this phenomenon is severe in acoustic sensing since the moving speed can easily exceed $0.43m/s$. This problem does not arise in the radar community because RF signals propagate much faster than acoustic signals.

4.1.2 Addressing Phase Under-Sampling Issue. A natural way to address the under-sampling issue is to decrease the length of the frame used for channel estimation. To support velocities up to $2.7m/s$, we can reduce the frame length to $1.25ms$. However, reducing the frame length will reduce both the detection range and SNR [28, 63]. Another way to increase the sampling frequency while keeping the frame duration is to slide the window by a portion of the frame. This approach is used in [2, 60]. However, this significantly increases the computation overhead, since fast movement also introduces Doppler shift and it is expensive to estimate Doppler shift. Estimating Doppler shifts multiple times within a frame is very expensive. For example, we try to estimate the Doppler shift every $1.25ms$ during each frame on the 5 COTS devices. The average processing time is around $22.5ms$ but the frame duration is only $10ms$, which means it cannot be processed in real-time on mobile devices. Therefore, we seek to avoid phase wrap-around while using a reasonable frame length (*e.g.*, $10ms$) and channel estimation interval (*e.g.*, $10ms$).

4.1.3 Use Phase Derivative to Avoid Under-Sampling Issue. We observed that the quotient between the phases of two consecutive taps can be also used to estimate motion, which is:

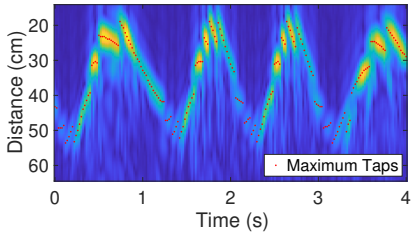
$$\begin{aligned} p'[t] &= \frac{p[t]}{p[t-1]} \\ &= \frac{|p[t]|}{|p[t-1]|} e^{-j\frac{2\pi f_c}{c}(d(t)-d(t-1))} \\ &= |p'[t]| e^{-j\frac{2\pi f_c T}{c} v(t)} \end{aligned} \tag{7}$$

where $\phi_v[t]$ denotes the phase of $p'[t]$. We can see that computing the quotient of two complex numbers will result in differentiating the phase.

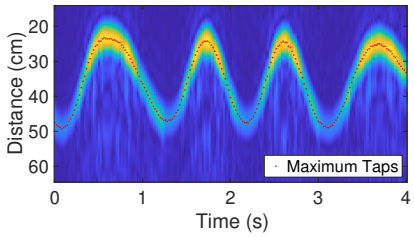
Interestingly, if we track motion using acceleration, phase will not wrap around until a much higher threshold. Specifically, we have

$$\begin{aligned} |\phi_v[t] - \phi_v[t-1]| &\leq \pi \\ \frac{4\pi f_c T}{c} |v[t] - v[t-1]| &\leq \pi \\ \frac{4 f_c T^2}{c} \left| \frac{v[t] - v[t-1]}{T} \right| &\leq 1 \\ \frac{4 f_c T^2}{c} |a[t]| &\leq 1 \\ |a[t]| &\leq \frac{c}{4 f_c T^2} \end{aligned} \tag{8}$$

This shows that there is no phase ambiguity as long as the acceleration is lower than $42.3m/s^2$ during a $10ms$ frame. The peak acceleration of human hand motion can only achieve around $30m/s^2$ [8], which is below $42.3m/s^2$. This suggests that using phase derivative for motion tracking will not incur phase ambiguity. **In a nutshell, performing phase unwrapping on the first-order phase derivative will avoid the under-sampling issue caused by fast movement for human mobility.** Note that, we can simply assume that the target starts moving from a static position (*i.e.*, with zero initial velocity and acceleration). Moreover, this approach can be potentially applied to a higher-order derivative of phase to support even higher mobility. For example, we can estimate the second-order derivative when the acceleration exceeds the threshold.



(a) Pulse distortion caused by motion



(b) After Doppler compensation

Figure 5: CIR estimations are severely distorted by motion due to the significant Doppler effect. The red dots are the selected taps according to the maximum channel taps.

4.1.4 Doppler Shift. Another main issue introduced by the fast movement is Doppler shift. If it were not considered, it would distort phase changes during the passband to baseband conversion and lead to significant errors. If we knew the ground truth velocity and compensated the resulting Doppler shift, tracking accuracy can remain high, assuming a high enough sampling rate. The major challenge is to determine how much Doppler shift to compensate.

The impact of rapid motion on the pulse of CIR estimation can be modeled by the following equation:

$$\hat{h}[n] = g[n] * h[n] \tag{9}$$

where $h[n]$ is the true CIR, $\hat{h}[n]$ is the estimated CIR, and $*$ denotes the convolution operator. The term $g[n]$ is the impact of motion. The profile in Fig. 5(a) is generated by a hand-sized object moving in the range from $20cm$ to $55cm$. The reflected signals affected by the rapid motion are very likely to severely distort the CIR profiles. Therefore fast movement makes the tap selection challenging and introduces errors in the phase measurement.

To address the issue, we develop an algorithm to compensate for the Doppler shift, as show in Alg. 1. This algorithm processes the baseband complex signal by enumerating different Doppler shifts and choosing the one that maximizes the peak value of CIR profiles. The basic idea of our compensation scheme is developed based on the widely used frequency synchronization technique in digital communication systems [43, 47]. The Doppler shift caused by the maximum supported velocity in Eq. 6 can be computed as follows:

$$F^s = \frac{2 \times v_{max}}{c} f_c = \frac{1}{2T} = \frac{1}{2}\Delta f \tag{10}$$

where $v_{max} = \frac{c}{4f_cT}$ represents the maximum velocity supported by the system without introducing phase ambiguity in Eq. 6. Let us consider the round-trip motion and $\Delta f = \frac{1}{T}$ is the frequency spacing of periodical signals [48]. **It tells us that the fast movement will not affect the phase measurement of the peak channel taps if the Doppler shift is less than half of the frequency spacing.** This observation shows that our Doppler compensation results can be relatively coarse. We only need to ensure a frequency shift between the compensated and template frames to be smaller than half of the frequency spacing. Therefore, we can search for the velocity using a step size empirically set to $10cm/s$. Furthermore, since the peak acceleration of human arms is around $30m/s^2$, we know that the maximum velocity change between two consecutive frames is around $30cm/s$. Due to the round-trip propagation in device-free tracking, we set the search range to $(v_e - 60cm/s, v_e + 60cm/s)$, where v_e is the last estimation and the initial velocity is 0 because we assume the target moves from a stationary position. Finally, we use the frequency domain cross-correlation [20] to speed up the cross-correlation step. Recent studies [5, 57] show that the time domain cross correlation is memory efficient, but the frequency domain cross correlation is time efficient. Furthermore, we use the FFTW library [9] to speed up the FFT computation. Our algorithm for Doppler compensation is shown in Alg. 1.

4.1.5 Summary. Putting together, we first mitigate pulse distortions by compensating the Doppler shift, then apply the phase derivative to calculate the current velocity and estimate the distance using integration. Applying Doppler compensation to COTS mobile devices introduces considerable overhead. We deal with this issue in three aspects: (i) Alg. 1 is only executed once in each frame since phase derivative provides enough phase samples even under fast motions, (ii) we observe that the frequency shift estimation can be very coarse for tracking purpose, (iii) we use the frequency domain cross-correlation to speed up the computation. Our phase derivative approach differs from the conventional path following algorithms (PFA) and quality-guided algorithms (QGA) in two aspects: (i) our approach can be applied to one-dimension phase data, and (ii) we directly apply the basic unwrapping algorithm to the phase derivative without ambiguity detections.

During the implementation of the above algorithms, we further find that: (i) a significant sudden phase change (caused by insufficient SNR or signal distortion), as shown in Fig. 2(c), may also cause the failure of phase unwrapping, producing errors when integrating velocity to distance, (ii) low SNR conditions may cause the failure of the Doppler compensation algorithm. We observe that these two phenomena can be attributed to fast attenuation of acoustic signals and the distortion of hardware frequency response, which are also two practical challenges for motion tracking, not only for the fast movement but also for the general scenarios. Thus, we address these two practical challenges in Sec. 4.2 and Sec. 4.3 to achieve more robust fast motion tracking.

Algorithm 1: Doppler Compensation Algorithm

Input: Baseband signal, $rx[n,t]$, $n = 0 \dots N-1$, $t = 0 \dots T$
Output: Compensated CIR, $h[n]$

Initialize $v_e[0]$: initial velocity, $k[0]$: initial tap number, $zc[n]$: template ZC sequence.
for $t \leftarrow 1$ *to* T **do**
  /*Search for the velocity with a step size of $10cm/s$*/
  for $v \leftarrow v_e[t-1] - 60cm/s$ *to* $v_e[t-1] + 60cm/s$ **do**
    /*Find the correlation peak.*/
    $R = corr(rx[n,t] \cdot e^{j2\pi(\frac{v}{c}f_c)\frac{n}{f_s}}, zc[n])$
    $R_{pks}[v] \leftarrow max(|R[taps\ around\ k[t-1]]|)$
  end
  /* Compensate Doppler with the best v.*/
  $v_e[t] \leftarrow \text{argmax}(R_{pks}[v])$
  $h[n] = corr(rx[n] \cdot e^{j2\pi(\frac{v_e[t]}{c}f_c)\frac{n}{f_s}}, zc[n])$
  /*Update taps*/
  $k[t] \leftarrow \text{argmax}(|h[taps\ around\ k[t-1], t]|)$
end
return;

4.2 Enhancing SNR

The SNR can be low due to fast attenuation of reflected signal and limited power of the speaker on a mobile device. In this case, the taps corresponding to the target are prone to noise, leading to the noisy phase measurement, as shown in Fig. 6(a), which causes errors in both Doppler estimation and phase measurement.

Specifically, the CIR is modeled as follows:

$$\hat{h}[n] = g[n] * h[n] + w[n] \tag{11}$$

where $w[n]$ represents the noise, $\hat{h}[n]$ is the CIR measurement, and $g[n]$ is the distortion introduced by Doppler shift. Our goal is to reduce $w[n]$ in $\hat{h}[n]$.

The intuition behind our scheme is that the CIRs between two consecutive frames are rotated by a certain phase, which is caused by motion. If we can compensate the phase change, the CIRs can be added up constructively and strengthened. This idea is inspired by antenna beamforming where signals received at different timestamps can be added up to improve SNR after compensating the phase difference caused by their different positions [41]. Our time-domain beamforming works even when there is one microphone because the CIRs are summed up across time from each microphone.

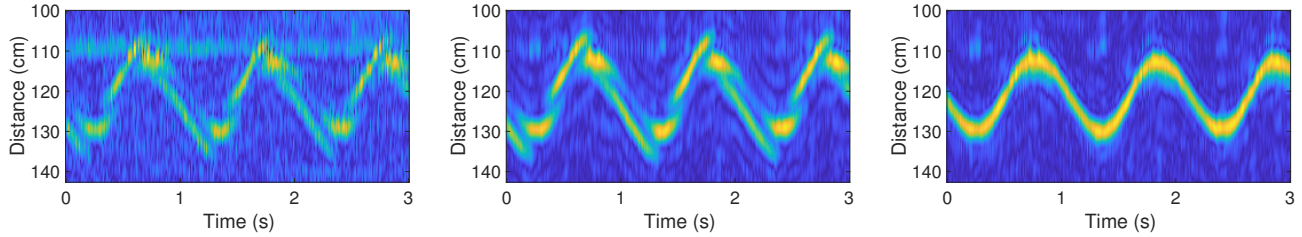
The remaining issue is how to determine the phase change during the interval to compensate. We observe the relationship between two consecutive CIRs can be modeled by:

$$h[n] = h[n-1]e^{j\phi} \tag{12}$$

where ϕ is the phase change caused by motion. This equation holds when there is one moving object. However, when there are multiple well-separated moving objects, each object affects a subset of taps and the equation holds for the set of the taps next to each object. Then we estimate the next CIR based on the measurement in the current frame using exponential weighted moving average (EWMA) [21] as follow:

$$z[n] = (1-K) \times z[n-1]e^{j\hat{\phi}} + K \times \tilde{h}[n] \tag{13}$$

where $z[n]$ is the SNR-enhanced CIR, $\tilde{h}[n]$ is the measurement, and $z[0] = \tilde{h}[0]$. K is a weighting factor, which is empirically set to be



(a) Background residuals will affect the performance of long-range tracking (b) Apply time domain beamforming to improve SNR (c) Compensate the Doppler shift after improving SNR.

Figure 6: Time domain beamforming effective improves the SNR of the CIR profile

0.3. $\hat{\phi}$ is the best phase estimation that makes $z[n-1]e^{j\hat{\phi}}$ closest to $\tilde{h}[n]$.

To find $\hat{\phi}$, we develop the following optimization model based on the maximum entropy principle [16]:

$$\underset{\phi}{\operatorname{argmax}}\, H(normalized(|\tilde{h}[n] - \hat{z}[n]| \cdot |\hat{z}[n]|)) \tag{14}$$

where $H(\cdot)$ is the Shannon entropy and $\hat{z}[n] = z[n-1]e^{j\hat{\phi}}$ is the prediction of the next channel. The term $|\tilde{h}[n] - \hat{z}[n]|$ is the amplitude of the complex error between the measured CIR and the predicted CIR, and is a vector across all taps. We perform dot-product between $|\tilde{h}[n] - \hat{z}[n]|$ and $|\hat{z}[n]|$ because $|\hat{z}[n]|$ is largest at the tap next to the target and it is more important to minimize the component in $|\tilde{h}[n] - \hat{z}[n]|$ that is closest to the target. The maximum entropy principle states that if we do not have the prior knowledge of a distribution, the best way is to assign a distribution which has the maximum entropy. Hence, we maximize the Shannon entropy of the error function to distribute the weighted residual errors evenly across each tap. We employ the iterative gradient descend algorithm to solve this optimization problem. An example of the SNR-enhanced profile is shown in Fig. 6(b), where the noise is effectively suppressed.

Note that we use the previous best estimation as the current initial point, and it is natural to assume the object starts moving from a stationary state with a zero initial phase change (*i.e.*, $\hat{\phi}_0 = 0$). Moreover, as shown in Figure 6(b), our SNR enhancing algorithm can suppress the residual multipath of the background noise, which may have considerable impact on the tracking accuracy for long-range detection.

Algorithm 2: SNR Enhancing Algorithm

Input: Baseband signal, $rx[n,t]$, $n = 0 \ldots N-1$, $t = 0 \ldots T$
Output: SNR-Improved CIR, $z[n,t]$;
Initialize $\hat{\phi}[0]$: initial phase change, zc: ZC sequence.
for $t \leftarrow 1$ *to* T **do**
  $\tilde{h}[n] \leftarrow$ Compute the CIR profile.
  $\hat{\phi}[t] \leftarrow$ Solve the optimization problem in Eq. 14.
  $z[n] \leftarrow$ Combine the predicted CIR and measured CIR.
end
return;

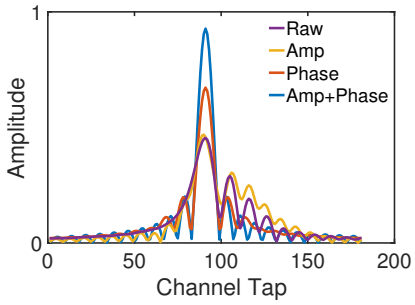


Figure 7: Different hardware frequency compensation schemes.

Two consecutive CIR profiles are likely to be affected by a similar Doppler distortion $g[n]$, given the short frame duration in between (*i.e.*, $10ms$). Thus, Eq. 12 holds with the presence of $g[n]$, which means our SNR enhancement algorithm also works under Doppler shift. After obtaining the SNR-enhanced CIR profile, $z[n,t]$, we have to convert the CIR profiles to baseband signal in order to compensate the Doppler distortion $g[n]$ according to Alg 1. Inspired by the frequency domain cross correlation, we find that we can approximately compute the baseband signal from a CIR profile as follows:

$$\hat{rx}[n,t] \approx iFFT(FFT(z[n,t]) \cdot FFT(zc[n])) \tag{15}$$

Then we can compensate the Doppler shift in $\hat{rx}[n,t]$ and convert it back to a CIR profile as shown in Fig. 6(c).

4.3 Hardware Frequency Response

It is common for speakers and microphones to have an uneven frequency response. Essentially, the received signal r is affected by both wireless channel and hardware frequency response as follow: $ZC_R[n] = h[n] * h_h * ZC_T[n]$, where h_h is the hardware frequency response. Note that h_h is a complex number, which implies that the hardware frequency response affects both the amplitude and phase of the received signal.

Existing works (*e.g.*, [39, 40]) focus on the amplitude of the hardware frequency response. We find that the phase of the frequency response is also important. Without hardware distortion, the phase response should be linear in terms of frequency. Hardware distortion causes a nonlinear phase response [64].

Fig. 7 compares the performance of different frequency compensation schemes at $17-23KHz$ on Google Pixel3: (i) compensate phase response alone, $\angle h_h$, (ii) compensate amplitude response alone $|h_h|$, and (iii) compensate both phase and amplitude responses, h_h. The purple line represents the raw CIR profile. Without compensation, we observe (i) the energy of the main lobe decreases, leading to lower SNR and (ii) the energy of side lobe increases, which may introduce more distortions and makes it difficult to correctly estimate the Doppler and select the tap. If we compensate the amplitude response alone as adopted by previous works [39, 40], we see some improvement. However, if we compensate the phase response, the main lobe becomes narrower and the side lobes also become smaller. Moreover, if we compensate both of them, we observe the highest peak value. Therefore, we compensate for the phase response as well as the amplitude response to tackle hardware distortions. We further evaluate various compensation schemes on different phones and different frequencies in Section 5 and show that their relative performance depend on the hardware and frequencies. Nevertheless, it is important to compensate for the phase of the frequency response in all cases.

4.4 Final Algorithm

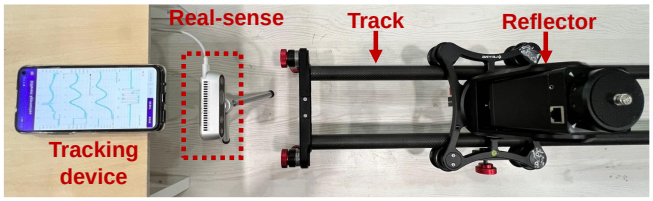
Basically, the phase derivative approach is sufficient to solve phase ambiguity issue caused by fast movement in theory. However, in real world, the phase gradient may be corrupted by noise, Doppler distortion, and hardware imperfection. Therefore, we develop and combine a series of algorithms to address the fast movement of human motions. During the process, we find that our single-channel SNR enhancement algorithm and hardware frequency response compensation can also be easily integrated into other acoustic tracking systems to improve the SNR of received signals.

The final algorithm is shown in Alg. 3. We first compensate both the amplitude and phase of hardware frequency response, followed by SNR enhancement and Doppler compensation. We enhance SNR so that we can more accurately estimate the Doppler shift. Note that SNR enhancement is applied to CIR while Doppler compensation is applied to the raw signal, so we need to perform baseband to CIR twice. Finally, we apply phase derivative to avoid the phase ambiguity and measure the fine-grained motion.

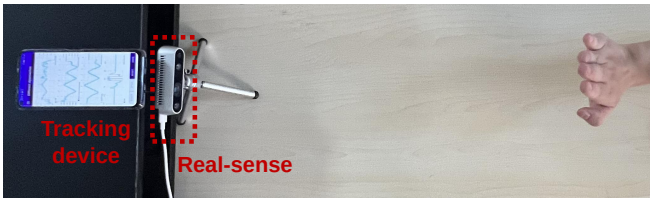
Algorithm 3: Final algorithm to handle fast movement

Input: Baseband signal, $rx[n,t]$, $n = 0 \dots N-1$, $t = 0 \dots T$
Output: Fine-grained velocity, $v[t]$

```
for ti ← 0 to T do
    rx[n,t] ← Compensate frequency response.
    h[n] ← SNR enhancement, Alg. 2.
    CIR to baseband signal, Eq. 15.
    h[n] ← Doppler compensation, Alg. 1.
    Use phase derivative to compute v[t] and d[t]
end
return;
```



(a) Track scenarios



(b) Real-life scenarios

Figure 8: Experiment setup

4.5 Position Estimation

In this section, we describe how to estimate the 1D position and 2D position of our target.

4.5.1 1D Postion Estimation. We derive two estimates of 1D position: an absolute distance based on the largest magnitudes in CIR profile, denoted as $d^p[t]$ and a relative distance derived from the fine-grained velocity estimation using the first-order phase derivative, denoted as $v^p[t]$. Then we combine these two estimations as follows:

$$d[t] = \beta d^p[t] + (1-\beta)(v^p[t]T + d[t-1]) \tag{16}$$

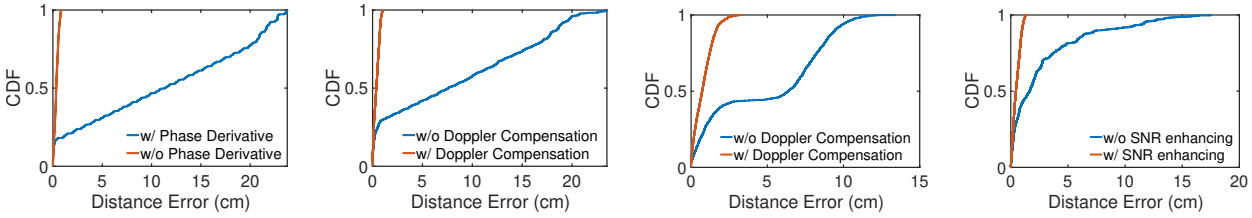
where β is a weighting factor in the range of $[0, 1]$, which is set to be 0.1, because we rely more on the fine-grained velocity estimation.

4.5.2 2D Position Estimation. Given the phone form factor, we know the relative positions of the speaker and two microphones. We can compute the path length from the speaker to one microphone. Therefore, the possible locations of our targets should be on the ellipse in the 2D space whose foci are the speaker and microphone. Using two microphones, the target is located at the intersections of two eclipses [46, 69, 74]. There are two intersections of these ellipses, and we select the one in front of the tracking device, which is a common region for the user's hand.

5 EVALUATION

5.1 Experiment Setup

To extensively evaluate the performance of our schemes, we develop an Android app SwiftTrack and test it on the following five COTS devices: SamSung S7, Google Pixel 3, Xiaomi K20 Pro, Honor 20 Pro 1, and Honor 20 Pro 2. Our app performs entire signal processing locally on the smartphones in real-time. The speaker volume was set at 80% of the maximum and the microphone sampling frequency was fixed at 48 kHz. We set the ZC sequence frame length $T = 10ms$, corresponding to a maximum unambiguous range of around $1.7m$. This operating range is large enough for the interaction of human



(a) Phase derivative (b) Doppler compensation (c) Doppler compensation (d) SNR enhancement

Figure 9: Performance of (a) phase derivative scheme, Doppler compensation scheme for (b) phase-based and (c) tap-based, (d) SNR enhancement.

hand movement, and signals outside this range are negligible. Given the $4kHz$ bandwidth and $10ms \times 48kHz = 480$ samples in each frame, we set $N_{zc} = 39$ and the root of ZC sequence $\mu = (N_{zc} - 1)/2 = 19$.

To better control the moving speed, we mounted a hand-sized reflector on a $1.6m$ long track to randomly push it back and forth with human arms, as shown in Fig. 8(a). We placed a Real-Sense D435i [25] in front of the track to get the ground truth of target movements. However, due to the smaller range of the depth error, we place the object as close to the depth sensor as possible and strictly follow the instruction [14] to tune the parameters of the depth sensor to achieve the best performance. We also evaluate our schemes with users' hands as shown in Fig. 8(b). We attach a red marker to the center of the user's palm, which facilitates the Real-Sense to get ground-truth positions.

5.2 Micro-Benchmark

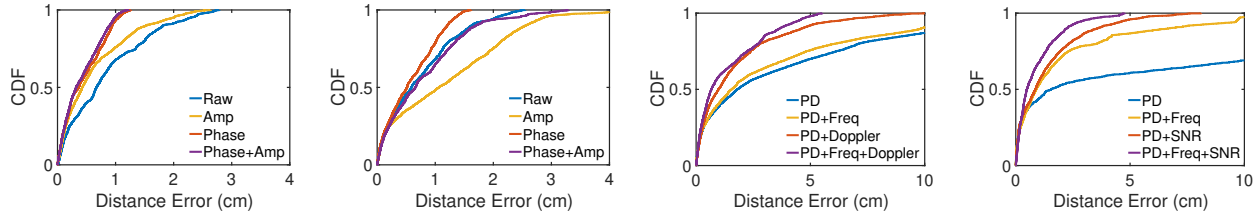
5.2.1 Impact of Phase Derivative. We first evaluate our approach using the phase derivative to address fast movement. We perform experiments using a moving track and compensate the Doppler shift using the ground truth velocities. The average velocity is $0.88m/s$; in 76% of time, the velocities are larger than $0.45m/s$, which cause phase ambiguities. Then we derive the distance from the selected taps' phase in two ways: (i) directly unwrap the measured phase without phase derivative and (ii) perform phase unwrapping on the first-order phase derivative and integrate the estimated velocity to get the final displacement. The results are shown in Fig. 9(a). Due to phase ambiguities caused by the under-sampling issue, the direct phase unwrapping cannot reconstruct the actual displacement, leading to significant errors. As we can see, the first-order phase derivative efficiently solves the under-sampling issue and reduces the median error by 96% from $11.27cm$ to $0.46cm$, 90^th percentile error from $21.73cm$ to $0.72cm$, thereby supporting fast movement.

5.2.2 Doppler shift compensation. We evaluate the impact of pulse distortions on tracking performance. Since the pulse distortion is related to the moving speed and high moving speed will cause an insufficient phase sampling rate, we apply the circular shift-based up-sampling scheme to decouple the pulse distortion and sampling interval. Then the tracking performance is mainly determined by the pulse distortion. Fig. 9(b) shows that the median error is reduced from $7.67cm$ to $0.49cm$ and the 90^{th} percentile error is reduced from $18.56cm$ to $0.79cm$, with the Doppler compensation. The result implies that even if the under-sampling issue is solved, the pulse distortion will also cause the failure of tracking rapid motion. Since the distorted pulse makes tap selection challenging, we also evaluate its impact on the absolute distance measurement corresponding to the selected taps. As shown in Fig. 9(c), the median absolute error is reduced from $6.37cm$ to $0.79cm$ and the 90^{th} percentile error is reduced from $9.27cm$ to $1.53cm$, with the Doppler compensation, implying a significant influence from the pulse distortion.

5.2.3 Enhancing SNR. We evaluate the performance of our SNR enhancement algorithm when the low SNR is low. The SNR is $-3dB$ when the target is $50cm$ from the phone and reduced to $-16dB$ at $1m$. In this experiment, we move the target between $1m$ to $1.3m$ with a speed lower than $20cm/s$ to ensure the signal is mainly affected by the noise. Then we compare the tracking accuracy with and without applying our algorithm. The results are shown in Figure 9(d). We can see that our SNR enhancement algorithm reduces the median displacement error by 70% from $1.54cm$ to $0.46cm$, and more importantly it significantly improves the tail performance (reducing the 90^{th} percentile error by 87% from $6.98cm$ to $0.91cm$).

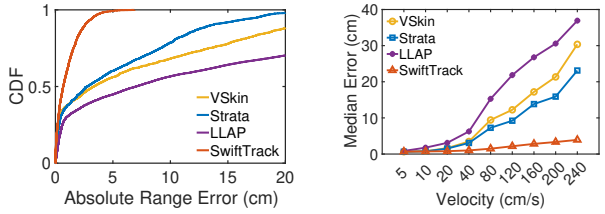
5.2.4 Hardware frequency response. The frequency response varies with devices, and different frequency responses impact the experimental results differently. To evaluate the impact of the frequency response compensation method, we compare the results in the following conditions: (i) do not compensate for frequency response (labeled as "*Raw*" in Fig. 10 (a) and (b)), (ii) compensate for amplitude response ("*Amp*"), (iii) compensate for phase response ("*Phase*"), and (iv) compensate for both phase and amplitude response ("*Phase + Amp*"). Moreover, the amplitude response may significantly suppress energy in some frequencies, as depicted in Fig. 4. Therefore, we conduct two experiments. In the first experiment, we use the frequency from $17kHz$ to $21kHz$ for Fig. 10(a). We use $17kHz$ to $23kHz$ in the second experiment for Fig. 10(b). Compensating for the amplitude response may degrade the accuracy because it will magnify the energy of noise for the frequencies above $22kHz$. In comparison, the phase compensation always has significant improvement. Thus, we use the both the phase and amplitude compensation scheme in the frequency band from $17kHz$ to $21kHz$ in our implementation.

5.2.5 Tracking rapid motion with combined schemes. We evaluate the performance of the combined schemes. Since the Doppler compensation scheme and SNR enhancement scheme can work independently to improve the tracking accuracy, we conduct two separate



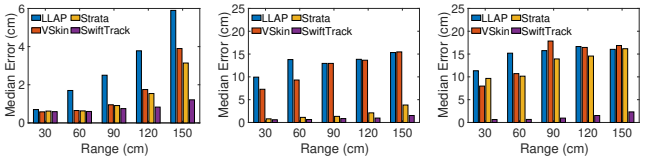
(a) $17kHz - 21kHz$ (b) $17kHz - 23kHz$ (c) Under fast motion (d) Under low SNR

Figure 10: Performance of frequency response compensation scheme on different bands, i.e., (a) $17kHz - 21kHz$ and (b) $17kHz - 23kHz$. Performance of combined schemes under different conditions, i.e., (c) the target moves fast and close to the phone ($30cm$), and (b) the target moves slowly but far from the phone ($120cm$).



(a) Ranging errors (b) Impact of velocities

Figure 11: Overall comparison of different approaches in terms of (a) ranging errors and (b) impact of velocities.



(a) Low speed. (b) Medium speed. (c) High speed.

Figure 12: Impact of distances at low speed range ($\leq 20cm/s$), medium speed range ($20cm/s \sim 80cm/s$) and high speed range ($\geq 80cm/s$).

experiments under different conditions to evaluate their performance. The results are shown in Fig. 10 (c) and (d). We compare the improvement after adding more schemes. *PD*, *Freq*, *Doppler*, and *SNR* represent phase derivative, frequency response compensation, Doppler compensation, and SNR enhancement schemes, respectively. The Doppler compensation scheme is applied when the target is close to the phone ($30cm$) and moves fast. Fig. 10(c) shows that the Doppler compensation significantly improves the tracking performance by 68% in the near field, and adding the frequency response compensation can further reduce the median error by 12%.

The SNR enhancement scheme is applied when the target is far from the phone ($120cm$) but moves slowly. In Fig. 10(d), we can see that both the SNR enhancement and frequency response compensation schemes significantly improve the accuracy by around 65% in the far field, respectively. Combining all schemes further improves the performance by 15%.

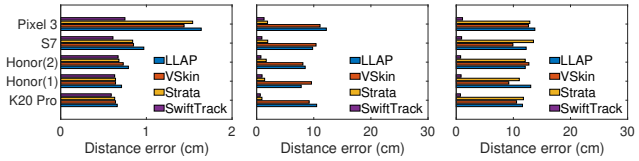
5.3 Overall Comparison

In this section, we compare the performance of our system with the the following previous work: Strata [74], VSkin [60], and LLAP [69]. For fair comparison, we use the signals with the same frame length: $T = 10ms$. Because Strata, VSkin, and our system are channel-based methods, they share the same traces. Since LLAP measures the phase of each frequency independently, we set the initial phase of each frequency to zero before playing the audio out by the speaker. We mounted a hand-sized reflector on a $1.6m$ long track to better control the maximum speed. We push this reflector in one direction for each movement while varying the maximum speed from $5cm/s$ to $240m/s$.

Fig. 11(a) shows the absolute distance estimation errors of the four schemes. SwiftTrack achieves a median error of $0.63cm$, outperforming Strata, VSkin, and LLAP by 253%, 327%, and 1114%, respectively. Moreover, the tail performance is significantly improved. The 95^{th} percentile error of SwiftTrack is reduced by 353%, 537%, and 1215%, respectively.

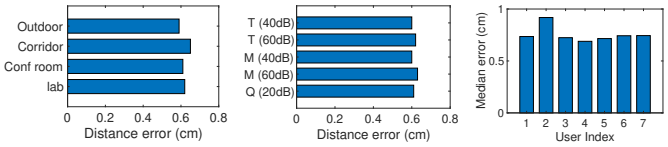
5.3.1 Impact of velocities. We further evaluate SwiftTrack under various scenarios. We first compare the tracking errors under various velocities from $5cm/s$ to $240cm/s$ to cover the speed range of human hand motion. Since human arm's movement is unpredictable in real scenarios and the errors caused by fast motions may accumulate over time, we calculate the mean error of each trace and the maximum velocity. Fig. 11(b) shows the results. When the speed is low (i.e. $\leq 10cm/s$), these four schemes have similar performance. When the speed exceeds $40cm/s$, Strata and VSkin degrade significantly. This is because for a frame length of $10ms$, the maximum supported speed is $45cm/s$ according to the Eq. 6. When the moving speed is $240cm/s$, the errors of Strata and VSkin are $23.67cm$ and $30.41cm$, respectively, while the error of SwiftTrack remains small ($3.69cm$).

5.3.2 Impact of distances. We test the system performance at various distances from $30cm$ to $1.5m$ with a step size of $30cm$. We conducted three sets of experiments under low speed ($\leq 20cm/s$), medium speed ($20cm/s \sim 80cm/s$), and high speed ($\geq 80cm/s$), respectively. The results are shown in Fig. 12. We can make the following two observations. First, when the speed is high, only SwiftTrack works well at various distances, while the other methods suffer from fast motion. Second, when the speed is low or medium, thanks to the SNR enhancement and frequency response compensation schemes, SwiftTrack can still out-perform the other methods.



(a) Low speed. (b) Medium speed. (c) High speed.

Figure 13: Impact of devices at low speed range ($\leq 20cm/s$), medium speed range ($20cm/s \sim 80cm/s$) and high speed range ($\geq 80cm/s$).



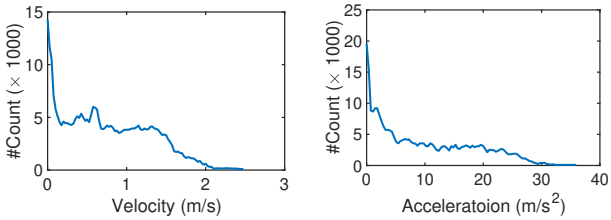
(a) Different locations (b) Ambient noise (c) Different users

Figure 14: Impact of (a) locations, (b) ambient noise, and (c) users.

5.3.3 Impact of devices. We evaluate the tracking performance on different COTS devices with low-speed ($\leq 20cm/s$), medium-speed($20cm \sim 80cm$), and high-speed ($\geq 80cm/s$) motion. The results are shown in Fig. 13. When the speed is high, the performance is mainly determined by the velocity, and only SwiftTrack works well. In the medium-speed range, both SwiftTrack and Strata work well, but SwiftTrack have better performance due to the compensation of hardware response. When the speed is low, we can see that the four methods work similarly on Honor 20 Pro 1, Honor 20 Pro 2, and Xiaomi K20 Pro; however, SwiftTrack outperforms Strata, VSkin, and LLAP by 36%-119% on Google Pixel 3 and Samsung S7 because the uneven frequency response is more pronounced on Google Pixel 3 and Samsung S7 than the other phones as shown in Fig. 4.

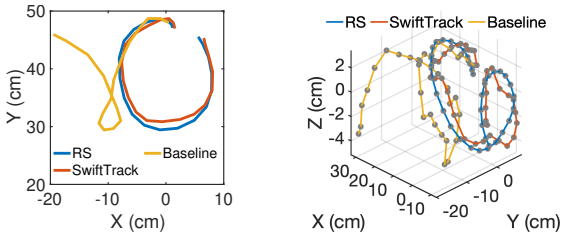
5.3.4 Impact of locations. To study the impact of different locations, we test our system at four different locations: our lab, a narrow corridor, a conference room, and an outdoor public space. The results are similar across different locations as shown in Fig. 14(a). The reasons are two-fold. First, the reflections of the environment are measured and removed by the background subtraction. Second, the slowly changed background residuals mainly affect the performance in low SNR regions, which can be effectively removed by our SNR enhancement algorithm.

5.3.5 Impact of ambient sounds. We evaluate SwiftTrack with different ambient sounds. Specifically, we conduct experiments in three scenarios: (i) a quiet environment (labeled as "Q" in Fig. 14(b)), (ii) an environment with people talking ("T"), and (iii) an environment with music playing ("M"). We place the noise source at $0.5m$ from the tracking device, and 2 different volume levels are considered for (ii) and (iii). The results of the 5 scenarios are shown in Fig. 14(b), which shows similar performance across different scenarios. The results are not surprising because we use the frequencies beyond the audible frequency range of human ears and are not affected by audible ambient sound.



(a) Speed Distribution (b) Acceleration Distribution

Figure 15: Statistics of collected data.



(a) 2D drawing sample (b) 3D drawing sample

Figure 16: 2D and 3D drawing samples. "*RS*" represents the trajectories captured by Real-Sense as reference. "$SwiftTrack$" represents the proposed method and "*Baseline*" represents disabling the phase derivative, SNR enhancement, frequency response compensation, and Doppler compensation.

5.3.6 Impact of users. We evaluate SwiftTrack in real usage scenarios. We recruit seven users to conduct experiments (with IRB approval). They are encouraged to freely move their arms back and forth. They are undergraduate and graduate students, from 21 to 27 year old, with 1 female and 6 male. As shown in Fig. 14(c), the tracking accuracy across different users is similar. The median error over the seven participants is $0.74cm$, which is slightly higher than that of a hand-sized reflector ($0.63cm$). The slightly higher error is likely because arms introduce additional reflection and arms' movement may differ from that of the hands', which are our targets.

5.4 Motion Statistics

We measure the speed distribution and acceleration distribution during the experiments. The results are shown in Fig. 15. We see the maximum speed is $2.47m/s$ and the maximum acceleration is $35.94m/s^2$. We have two observations. First, nearly half of the occurrences exceed $0.8m/s$, which is the maximum speed supported by the previous works. Second, the maximum acceleration is less than $43m/s^2$, which indicates the phase derivative algorithm can handle fast human motion tracking.

5.5 Drawing Samples

We build a draw-in-the-air interface based on SwiftTrack to show its drawing capability in 2D and 3D spaces. Since tracking in 3D space requires at least 3 mics, we use external mics for data collection. The point cloud data produced by Real-Sense are used to generate reference trajectory [1]. We draw circles quickly with our

Doppler Shift	SNR enhancement	Others	All
4.37ms	1.26ms	1.04ms	6.67ms

Table 1: Processing time of each part.

hands in the air (the 3D example) or on the desk (the 2D example). The maximum velocity is $132cm/s$. Fig. 16 gives the drawing examples with fast motion. We can see that the baseline deviates significantly from the reference trajectories while SwiftTrack follows.

5.6 System Latency

We measure the time for SwiftTrack to process each $10ms$ frame on Google Pixel 3 and report the median time after running it for 30 minutes in Table 1. Note that since the computation of the phase derivative and frequency compensation schemes are simple, their process time is close to $0ms$. The Doppler compensation scheme and SNR enhancement schemes take $4.37ms$ and $1.26ms$, respectively. The total processing time is $6.67ms$. Therefore SwiftTrack can process each $10ms$ frame in real-time.

6 RELATED WORKS

We classify related works into (i) device-free acoustic tracking, (ii) device-based acoustic tracking, (iii) RF-based tracking, and (iv) phase unwrapping techniques.

Device-free (Contactless) Acoustic Tracking: Acoustic-based tracking schemes only use speakers and microphones, which are widely equipped on most COTS mobile devices, and can achieve high tracking accuracy. FingerIO [46] provides device-free tracking on COTS mobile devices by computing the Time-of-Arrival (TOA) of the reflected signals. They utilize the cyclic suffix property of OFDM symbols to find the TOA of reflected signals. However, the sampling rate of the microphone limits the range resolution, and the slope of phase change in the frequency domain is susceptible to noise and interferences. LLAP [69] employs continuous waves (CW) to track the target. The measured phase of CW signals is less susceptible to noise and provides good range resolution. Nevertheless, its performance is highly susceptible to multipath interferences. Strata [74] addresses this issue by leveraging channel impulse response and using the phase of the appropriate channel tap for tracking. RTrack [41] proposes an interesting approach that combines signal processing with machine learning to significantly increase the sensing range. FMTrack [32] achieves multi-target tracking by iteratively searching for optimal parameters of their signal models.

Device-based acoustic tracking Unlike contactless tracking, users can hold a device to enable motion tracking. Device-based tracking enjoys higher SNR and less impact from multipath interference than device-free tracking. AAMouse [73] combines the Doppler shift measured across several frequencies to turn a mobile device into a mouse in the air. CAT [38] and Rabbit [42] utilize the property of chirp signals to estimate the distance between the transmitter and receiver. SoundTrack [75] measures the phase information of CW signals to locate a customized finger ring in 3D space. Millisonic [65] leverages the phase of mixed chirp signal in the time domain to track multiple devices.

RF-based tracking Radio Frequency (RF) signals, such as WiFi, RFID, and mmWave, have been used for localization and tracking. For example, ArrayTrack [71] measures the phase of the received signal and achieves a median error of $23cm$ using 16 antennas. RF-IDraw [67] and WiDraw [61] achieve several centimeter tracking errors. Increasing the frequency to around 60GHz can improve the tracking performance, such as mTrack [70], Soli [33], and mmVib [27], but they require customized hardware.

Phase Unwrapping Techniques Phase unwrapping techniques have been widely used in research fields, such as optical interferometry [44, 49], streak projection profilometry [19], synthetic aperture radar (SAR) [7, 55], magnetic resonance imaging (MRI) [31, 53, 54], etc. In these cases, determining the singularities of two-dimensional phase data (*i.e.*, points that do not satisfy the smoothness condition) is the key to phase unwrapping. There are two basic approaches: residual theorem [10, 22, 23] and quality map [10, 29, 35]. The singularities can be determined by coherence [24, 56], gray-level co-occurrence matrix [50], and phase derivative variance [36, 37]. After finding singularities, we can carefully select the unwrapping path to avoid those regions by using path following algorithms [12, 72]. Note, however, that for the one-dimensional phase unwrapping problem in the tracking system, it is difficult to identify and bypass singularities and to choose a different unwrapping path.

7 DISCUSSION

SwiftTrack can effectively improve tracking accuracy and robustness under fast motions and be easily integrated into other tracking systems to enhance SNR and deal with uneven frequency response. Although the results for fast human motion tracking are promising, there still exist several challenges to be addressed in the future: (i) extend the SNR enhancement scheme to multiple channels to further improve performance, and (ii) support other objects' fast movement.

8 CONCLUSION

In this work, we identify several limitations of existing phase-based acoustic motion tracking, including the phase ambiguity and Doppler shift caused by fast movement, non-uniform frequency response compensation, and low SNR. We gain the following important insights: (i) Fast movement may cause phase measurements to be under-sampled and introduce ambiguity in the phase change. Performing phase unwrapping on the first-order phase derivative can avoid the under-sampling issue caused by fast movement for human mobility. (ii) To enhance the SNR, we can add up the signals in consecutive time intervals after compensating for the phase shift during these time intervals. We can find the best phase shift to compensate by maximizing the entropy of the error function between our measurement and estimation. (iii) Hardware frequency response varies across devices. We find that compensating the phase response is more reliable than compensating the amplitude response since the latter may significantly increase the noise. We develop effective solutions for each issue and experimentally demonstrate their effectiveness using Android implementation. We evaluate SwiftTrack with the velocity ranging from $5cm/s$ to $240cm/s$ and observe a a median error of $0.63cm$.

ACKNOWLEDGMENTS

We thank the anonymous reviewers whose suggestions helped improve and clarify the work. This work is supported by NSFC (62072306,61936015) and Program of Shanghai Academic Research Leader (20XD1402100).

REFERENCES

[1] 2023.2. Real-Sense Hand Tracking Tutorial. https://www.intel.com/content/dam/develop/external/us/en/documents/hand-tracking-843462.pdf.
[2] Mohammed H AlSharif, Mohamed Saad, Mohamed Siala, Tarig Ballal, Hatem Boujemaa, and Tareq Y Al-Naffouri. 2017. Zadoff-Chu coded ultrasonic signal for accurate range estimation. In *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 1250–1254.
[3] Jeffrey G Andrews. 2022. A Primer on Zadoff Chu Sequences. *arXiv preprint arXiv:2211.05702* (2022).
[4] Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. 2013. DopLink: Using the Doppler Effect for Multi-Device Interaction. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 583–586. https://doi.org/10.1145/2493432.2493515
[5] Calum J Chamberlain, Chet J Hopp, Carolin M Boese, Emily Warren-Smith, Derrick Chambers, Shanna X Chu, Konstantinos Michailos, and John Townend. 2018. EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python. *Seismological Research Letters* 89, 1 (2018), 173–181.
[6] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. 2014. AirLink: Sharing Files between Multiple Devices Using in-Air Gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 565–569. https://doi.org/10.1145/2632048.2632090
[7] John C Curlander and Robert N McDonough. 1991. *Synthetic aperture radar*. Vol. 11. Wiley, New York.
[8] Kurt M DeGoede, James A Ashton-Miller, Jimmy M Liao, and Neil B Alexander. 2001. How quickly can healthy adults move their hands to intercept an approaching object? Age and gender effects. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 56, 9 (2001), M584–M588.
[9] Matteo Frigo and Steven G Johnson. 1998. FFTW: An adaptive software architecture for the FFT. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, Vol. 3. IEEE, 1381–1384.
[10] Munther Gdeisat. 2019. Two-Dimensional Phase Unwrapping Problem. (06 2019).
[11] Munther Gdeisat and Francis Lilley. 2011. One-dimensional phase unwrapping problem. *signal* 4 (2011), 6.
[12] Richard M Goldstein, Howard A Zebker, and Charles L Werner. 1988. Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio science* 23, 4 (1988), 713–720.
[13] S Golomb and R Scholtz. 1965. Generalized barker sequences. *IEEE Transactions on Information theory* 11, 4 (1965), 533–537.
[14] Anders Grunnet-Jepsen, John N Sweetser, and John Woodfill. 2018. Best-known-methods for tuning intel® realsense™ d400 depth cameras for best performance. *Intel Corporation: Satan Clara, CA, USA* 1 (2018).
[15] Guifen Gu and Guili Peng. 2010. The survey of GSM wireless communication system. In *2010 international conference on computer and information application*. IEEE, 121–124.
[16] Silviu Guiasu and Abe Shenitzer. 1985. The principle of maximum entropy. *The mathematical intelligencer* 7, 1 (1985), 42–48.
[17] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Soundwave: using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1911–1914.
[18] MP Jacob Habgood, David Moore, David Wilson, and Sergio Alapont. 2018. Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 371–378.
[19] Liya Han, Zhongwei Li, Kai Zhong, Xu Cheng, Hua Luo, Gang Liu, Junyun Shang, Congjun Wang, and Yusheng Shi. 2019. Vibration detection and motion compensation for multi-frequency phase-shifting-based 3d sensors. *Sensors* 19, 6 (2019), 1368.
[20] George Helffrich. 2006. Extended-time multitaper frequency domain cross-correlation receiver-function estimation. *Bulletin of the Seismological Society of America* 96, 1 (2006), 344–347.
[21] J Stuart Hunter. 1986. The exponentially weighted moving average. *Journal of quality technology* 18, 4 (1986), 203–210.
[22] JM Huntley. 1989. Noise-immune phase unwrapping algorithm. *Applied optics* 28, 16 (1989), 3268–3270.
[23] JM Huntley and JR Buckland. 1995. Characterization of sources of 2π phase discontinuity in speckle interferograms. *JOSA A* 12, 9 (1995), 1990–1996.
[24] Rubén Iglesias, Jordi J Mallorqui, Dani Monells, Carlos López-Martínez, Xavier Fabregas, Albert Aguasca, Josep A Gili, and Jordi Corominas. 2015. PSI deformation map retrieval by means of temporal sublook coherence on reduced sets of SAR images. *Remote Sensing* 7, 1 (2015), 530–563.
[25] Intel. 2022. Intel® RealSense™ Depth Camera D435i. Retrieved June, 21, 2022 from https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d435i.html
[26] Kazuyoshi Itoh. 1982. Analysis of the phase unwrapping algorithm. *Applied optics* 21, 14 (1982), 2470–2470.
[27] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. 2020. MmVib: Micrometer-Level Vibration Measurement with Mmwave Radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking* (London, United Kingdom) *(MobiCom '20)*. Association for Computing Machinery, New York, NY, USA, Article 45, 13 pages. https://doi.org/10.1145/3372224.3419202
[28] Min-Ho Ka and Aleksandr I Baskakov. 2007. Selection of pulse repetition frequency in high-precision oceanographic radar altimeters. *IEEE Geoscience and Remote Sensing Letters* 4, 3 (2007), 345–348.
[29] Qian Kemao. 2007. Two-dimensional windowed Fourier transform for fringe pattern analysis: principles, applications and implementations. *Optics and Lasers in Engineering* 45, 2 (2007), 304–317.
[30] Eike Langbehn, Tobias Eichler, Sobin Ghose, Kai von Luck, Gerd Bruder, and Frank Steinicke. 2015. Evaluation of an omnidirectional walking-in-place user interface with virtual locomotion speed scaled by forward leaning angle. In *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. 149–160.
[31] Paul C Lauterbur. 1973. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *nature* 242, 5394 (1973), 190–191.
[32] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 150–163.
[33] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–19.
[34] Chao Liu, Penghao Wang, Ruobing Jiang, and Yanmin Zhu. 2021. AMT: Acoustic Multi-target Tracking with Smartphone MIMO System. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 1–10.
[35] Gang Liu, Robert Wang, YunKai Deng, Runpu Chen, Yunfeng Shao, and Zhihui Yuan. 2013. A new quality map for 2-D phase unwrapping based on gray level co-occurrence matrix. *IEEE Geoscience and Remote Sensing Letters* 11, 2 (2013), 444–448.
[36] Yuangang Lu, Xiangzhao Wang, and Xuping Zhang. 2007. Weighted least-squares phase unwrapping algorithm based on derivative variance correlation map. *Optik* 118, 2 (2007), 62–66.
[37] YG Lu and XP Zhang. 2006. Minimum L0-norm two-dimensional phase unwrapping algorithm Based on the derivative variance correlation map. In *Journal of Physics: Conference Series*, Vol. 48. IOP Publishing, 057.
[38] Wenguang Mao, Jian He, and Lili Qiu. 2016. Cat: high-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 69–81.
[39] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. 2020. DeepRange: acoustic ranging via deep learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.
[40] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. Aim: Acoustic imaging on a mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 468–481.
[41] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-based room scale hand motion tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[42] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. 2017. Indoor follow me drone. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 345–358.
[43] Umberto Mengali. 2013. *Synchronization techniques for digital receivers*. Springer Science & Business Media.
[44] Carolyn R Mercer and Glenn Beheim. 1990. *Fiber-optic Projected Fringe Digital Interferometry*. National Aeronautics and Space Administration.
[45] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless Sleep Apnea Detection on Smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 45–57. https://doi.org/10.1145/2742647.2742674
[46] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1515–1525.

[47] Steve Spreads It On. 1985. *Digital communications*. Van Nostrand Reinhold, New York.

[48] Alan V Oppenheim. 1999. *Discrete-time signal processing*. Pearson Education India.

[49] SM Pandit, N Jordache, and GA Joshi. 1994. Data-dependent systems methodology for noise-insensitive phase unwrapping in laser interferometric surface characterization. *JOSA A* 11, 10 (1994), 2584–2592.

[50] Mari Partio, Bogdan Cramariuc, Moncef Gabbouj, and Ari Visa. 2002. Rock texture retrieval using gray level co-occurrence matrix. In *Proc. of 5th Nordic Signal Processing Symposium*, Vol. 75. Citeseer.

[51] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. 2007. BeepBeep: A High Accuracy Acoustic Ranging System Using COTS Mobile Devices. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems* (Sydney, Australia) *(SenSys '07)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/1322263.1322265

[52] Branislav M Popovic. 1992. Generalized chirp-like polyphase sequences with optimum correlation properties. *IEEE Transactions on Information Theory* 38, 4 (1992), 1406–1409.

[53] Bruno Quesson, Jacco A de Zwart, and Chrit TW Moonen. 2000. Magnetic resonance temperature imaging for guidance of thermotherapy. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 12, 4 (2000), 525–533.

[54] Alexander Rauscher, Markus Barth, Jürgen R Reichenbach, Rudolf Stollberger, and Ewald Moser. 2003. Automated unwrapping of MR phase images applied to BOLD MR-venography at 3 Tesla. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 18, 2 (2003), 175–180.

[55] Paul A Rosen, Scott Hensley, Ian R Joughin, Fuk K Li, Soren N Madsen, Ernesto Rodriguez, and Richard M Goldstein. 2000. Synthetic aperture radar interferometry. *Proc. IEEE* 88, 3 (2000), 333–382.

[56] M Schwabisch and D Geudtner. 1995. Improvement of phase and coherence map quality using azimuth prefiltering: Examples from ERS-1 and X-SAR. In *1995 International Geoscience and Remote Sensing Symposium, IGARSS'95. Quantitative Remote Sensing for Science and Applications*, Vol. 1. IEEE, 205–207.

[57] Nader Shakibay Senobari, Gareth J Funning, Eamonn Keogh, Yan Zhu, Chin-Chia Michael Yeh, Zachary Zimmerman, and Abdullah Mueen. 2019. Super-efficient cross-correlation (SEC-C): A fast matched filtering code suitable for desktop computers. *Seismological Research Letters* 90, 1 (2019), 322–334.

[58] GE Spoorthi, Rama Krishna Sai Subrahmanyam Gorthi, and Subrahmanyam Gorthi. 2020. PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach. *IEEE Transactions on Image Processing* 29 (2020), 4862–4872.

[59] Kay M Stanney and Phillip Hash. 1998. Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence* 7, 5 (1998), 447–459.

[60] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 591–605.

[61] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. 2015. Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 77–89.

[62] Shinichi Tamura, Shigenori Nakano, and Kozo Okazaki. 1985. Optical code-multiplex transmission by gold sequences. *Journal of lightwave technology* 3, 1 (1985), 121–127.

[63] Michele Vespe, Gareth Jones, and Chris J Baker. 2009. Lessons for radar. *IEEE Signal Processing Magazine* 26, 1 (2009), 65–75.

[64] JR Wait. 1970. Distortion of pulsed signals when the group delay is a nonlinear function of frequency. *Proc. IEEE* 58, 8 (1970), 1292–1294.

[65] Anran Wang and Shyamnath Gollakota. 2019. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[66] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[67] Jue Wang, Deepak Vasisht, and Dina Katabi. 2014. RF-IDraw: Virtual touch screen in the air using RF signals. *ACM SIGCOMM Computer Communication Review* 44, 4 (2014), 235–246.

[68] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.

[69] Wei Wang, Alex X Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 82–94.

[70] Teng Wei and Xinyu Zhang. 2015. mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 117–129.

[71] Jie Xiong and Kyle Jamieson. 2013. {ArrayTrack}: A {Fine-Grained} Indoor Location System. In *10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*. 71–84.

[72] Hanwen Yu, Yang Lan, Hyongki Lee, and Ning Cao. 2018. 2-D phase unwrapping using minimum infinity-norm. *IEEE Geoscience and Remote Sensing Letters* 15, 12 (2018), 1887–1891.

[73] Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a Mobile Device into a Mouse in the Air. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 15–29. https://doi.org/10.1145/2742647.2742662

[74] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-Grained Acoustic-Based Device-Free Tracking. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services* (Niagara Falls, New York, USA) *(MobiSys '17)*. Association for Computing Machinery, New York, NY, USA, 15–28. https://doi.org/10.1145/3081333.3081356

[75] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan, and Gregory D. Abowd. 2017. SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics. 1, 2, Article 30 (June 2017), 25 pages. https://doi.org/10.1145/3090095

[76] Yongzhao Zhang, Wei-Hsiang Huang, Chih-Yun Yang, Wen-Ping Wang, Yi-Chao Chen, Chuang-Wen You, Da-Yuan Huang, Guangtao Xue, and Jiadi Yu. 2020. Endophasia: Utilizing Acoustic-Based Imaging for Issuing Contact-Free Silent Speech Commands. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–26.

[77] Zengbin Zhang, David Chu, Xiaomeng Chen, and Thomas Moscibroda. 2012. SwordFight: Enabling a New Class of Phone-to-Phone Action Games on Commodity Phones. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services* (Low Wood Bay, Lake District, UK) *(MobiSys '12)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/2307636.2307638

[78] Yuzhou Zhuang, Yuntao Wang, Yukang Yan, Xuhai Xu, and Yuanchun Shi. 2021. *ReflecTrack: Enabling 3D Acoustic Position Tracking Using Commodity Dual-Microphone Smartphones*. Association for Computing Machinery, New York, NY, USA, 1050–1062. https://doi.org/10.1145/3472749.3474805