SWIFTTRACK+: Fine-Grained and Robust Fast Hand Motion Tracking Using Acoustic Signal

Yongzhao Zhang, *Member, IEEE*, Hao Pan, *Member, IEEE*, Dian Ding, *Member, IEEE*, Yue Pan, *Student Member, IEEE*, Yi-Chao Chen, *Member, IEEE*, Lili Qiu, *Fellow, IEEE*, Guangtao Xue, *Member, IEEE*, Ting Chen, *Member, IEEE*, and Xiaosong Zhang, *Member, IEEE*

***Abstract*— Acoustic tracking technology, leveraging the ubiquitous presence of speakers and microphones in commercial off-the-shelf (COTS) mobile devices, has become a versatile tool across various applications. However, current phase-based acoustic tracking methods encounter significant limitations in tracking fast movements, thereby restricting their practical utility. This paper identifies three practical challenges to enable fast hand motion tracking using acoustic signals: 1) high mobility, 2) low signal-to-noise ratio (SNR), and 3) variations in hardware frequency response. The high mobility introduces Doppler shift and phase ambiguity which is the primary cause of failure in fast movement tracking, while the latter two factors can further impair the tracking performance in practical scenarios involving high mobility. To address the high mobility issue, we effectively compensate the Doppler shift in the Channel Impulse Response (CIR) for better selection of channel taps and then propose a novel phase derivative approach to mitigate the phase ambiguity. To enhance the real-world robustness, we integrate multiple algorithms including an SNR enhancement algorithm inspired by time-domain beamforming and a hardware frequency response compensation approach that addresses both amplitude and phase distortions. Additionally, an LSTM-based distance reconstruction algorithm is further implemented to correct residual phase noise. Implemented on Android platforms under the name SWIFTTRACK+, our system demonstrates superior performance in tracking fast movements. Through extensive evaluations, SWIFTTRACK+ proves its efficacy across diverse scenarios, significantly broadening the scope and reliability of acoustic tracking applications.**

***Index Terms*— Acoustic motion tracking, acoustic sensing, fast motion tracking, SNR enhancement.**

Received 5 December 2023; revised 30 September 2024; accepted 10 November 2024; approved by IEEE TRANSACTIONS ON NETWORKING Editor M. Zhang. Date of publication 11 December 2024; date of current version 18 April 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 6240071246; in part by Sichuan Science and Technology Program under Grant 2024NSFTD0031, Grant 2024NSFSC0004, and Grant 24ZNSFSC0038; and in part by the Key Research and Development Program of Sichuan Province under Grant 2022YFSY0005 and Grant 2023ZHJY0006. This work was partially presented at IEEE/ACM IPSN 2023 [DOI: 10.1145/3583120.3586954]. *(Corresponding author: Ting Chen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Shanghai Jiao Tong University.

Yongzhao Zhang, Ting Chen, and Xiaosong Zhang are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China (e-mail: zhangyongzhao@uestc.edu.cn; brokendragon@uestc.edu.cn; johnsonzxs@uestc.edu.cn).

Hao Pan, Dian Ding, Yue Pan, Yi-Chao Chen, and Guangtao Xue are with the College of Computer Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: panh09@sjtu.edu.cn; dingdian94@sjtu.edu.cn; kyino-pan@sjtu.edu.cn; yichao@sjtu.edu.cn; gt_xue@sjtu.edu.cn).

Lili Qiu is with Computer Science Department, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: lili@cs.utexas.edu).

Digital Object Identifier 10.1109/TNET.2024.3504517

I. INTRODUCTION

DEVICE-FREE acoustic tracking technologies hold the potential to revolutionize a myriad of applications. These include motion-based gaming, Augmented Reality (AR), Virtual Reality (VR), touchless user interfaces (UIs) for the Internet of Things (IoT), and also applications in health monitoring like breathing and heart-rate tracking, as well as indoor localization.

The recent advancements in this field have led to a plethora of device-free acoustic tracking algorithms. These can be categorically divided into four primary types: time-of-arrival (TOA) based [2], [3], [4], [5], Doppler based [6], [7], [8], Frequency-Modulated Continuous-Wave (FMCW) based [9], [10], and phase based [11], [12], [13], [14], [15]. Among these, the resolution of TOA, Doppler, and FMCW based approaches is generally constrained by the sampling rate and bandwidth. In contrast, phase-based tracking stands out due to its higher resolution capabilities; for example, a mere $1mm$ movement can result in a 0.74 radian phase change in device-free tracking at $20KHz$.

However, the direct use of signal phase at each sample is often challenged by multipath effects. To address this, Strata [12] introduces a novel method that estimates the channel impulse response (CIR), by calculating the channel coefficient for each channel tap to separate the multipath signals and utilizing the phase of a suitable channel tap for motion tracking in each signal frame. This approach has since inspired several notable extensions in the field, such as [13], [16], [17], [18], and [19].

Limitations of Existing Work: Though phase measurement in acoustic tracking offers high accuracy, our extensive experiments reveal a fundamental limitation: current phase-based tracking methods struggle with fast movements, significantly constraining their applicability. As illustrated in Fig.1(a), systems like Strata [12] are effective under slow motion. However, as depicted in Fig.1(b), these systems exhibit high errors or even inverse directionality when movement speed exceeds a

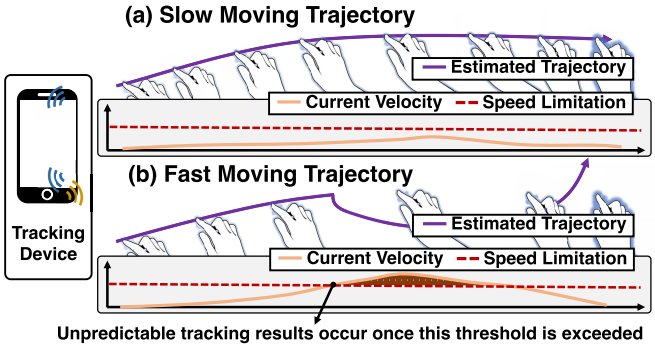


Fig. 1. An example of the rapid motion problem. (a) When user's finger is moving with speeds below the limit, the phase-based tracking schemes can estimate the trajectory accurately. (b) Once the moving speed exceeds the limit, tracking accuracy degrades significantly.

specific threshold, such as $0.8m/s$. This issue is not unique to Strata; other methods like LLAP [11] and VSkin [13] also face performance decline at higher speeds, functioning only below $0.25m/s$ and $0.12m/s$, as reported respectively. This poses a significant problem as human movement typically ranges between $1.5 - 2m/s$, with peak speeds around $2.7m/s$ as reported by [20], [21], and [22]. Currently, no acoustic-based tracking method effectively supports speeds beyond $1m/s$ in a device-free sensing system.

Delving deeper into the fast movement issue, we observe it intersects with other challenges in acoustic sensing, notably low SNR and uneven hardware frequency response. The rapid attenuation of reflected signals over distance results in low SNR beyond $1m$ in commercial off-the-shelf (COTS) smartphone setups. Tracking performance deteriorates significantly beyond this range due to phase error accumulation [23]. Existing methods [14], [24], while not ideally suited for high mobility scenarios, often rely on microphone arrays for spatial beamforming, a feature not typically available in COTS mobile phones. Furthermore, uneven frequency response of speakers and microphones distorts the received signal. Previous studies (*e.g.*, [25], [26]) have identified the heterogeneous nature of inaudible acoustic frequency response and suggested compensating for amplitude. However, our findings highlight that discrepancies in both amplitude and phase contribute to this issue. These distortions in the received signals substantially impact tracking accuracy.

Gaming applications, particularly in AR/VR or on mobile devices, often necessitate the capability to monitor human arm motions (moving speed up to $2.7m/s$) in the device-free manner. Existing acoustic sensing systems, however, struggle to support fine-grained tracking at higher speeds of movement (only effective for speed lower than $0.8m/s$). This paper aims to enable the tracking of fast motions of human arms using acoustic signals on COTS mobile devices, thus extending the scope of acoustic sensing in practical applications. To realize this goal, we address two primary challenges: (i) unraveling the root causes of limitations in fast motion tracking and devising effective solutions, and (ii) proposing a series of techniques, including SNR enhancement, uneven hardware frequency response compensation and LSTM-based distance recovery, to enhance the robustness and reliability of tracking under fast movement conditions in real-world scenarios.

Our Approach: Our in-depth analysis of fast movement traces has led to several key observations and solutions. Firstly, rapid movement often results in a phase change exceeding π between consecutive updates, causing phase under-sampling and, consequently, phase ambiguity. To address this, we have developed a novel approach using phase derivatives, effectively circumventing the under-sampling problem. 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 (above the peak acceleration of human arms) than the velocity and hence is more robust against the high mobility issue. Secondly, fast movement introduces Doppler shift, altering the phase change. We have crafted a simple yet effective technique for the real-time estimation and compensation of Doppler shift in the CIR on mobile devices.

To enhance the robustness of fine-grained tracking in practical scenarios involving rapid movements, we've designed a series of algorithms: (i) For mitigating low SNR issues, we employ a strategy inspired by time-domain beamforming [24], where we enhance SNR through the constructive addition of CIR profiles. This necessitates the accurate estimation and compensation of phase changes resulting from movement during measurement, for which we have developed a practical method based on the maximum entropy principle, even with a single microphone. (ii) Addressing hardware frequency response, we go beyond the typical approach of merely compensating for amplitude. Our systematic analysis underscores the necessity of compensating for the phase response due to its lower noise sensitivity. Thus, we propose a dual compensation method that addresses both phase and magnitude aspects of the frequency response. (iii) Lastly, to handle sudden noise in phase measurements that our SNR enhancement algorithm may not fully eliminate when the target moves at a longer distance, we introduce an LSTM-based fine-grained displacement reconstruction algorithm. This algorithm is specifically designed to filter out abrupt changes in phase derivatives, thereby further extending the operational range of fast hand motion tracking.

Implemented as SwiftTrack+ on COTS Android phones, our techniques facilitate fine-grained and robust fast hand motion tracking with acoustic signals from up to $2m$ away. Extensive evaluations validate the effectiveness of each technique. We also compare our system's performance with existing solutions like LLAP, Strata, and VSkin. Results show that for velocities ranging from $5cm/s$ to $240cm/s$, the median error in estimated absolute distance is just $0.63cm$, surpassing Strata, VSkin, and LLAP by 253%, 327%, and 1114%, respectively. Our work systematically addresses several critical practical challenges, demonstrating its benefits through real-world implementation. Note that, our solutions can be easily integrated to existing acoustic sensing systems to improve their performance under high mobility, low SNR, and heterogeneous devices.

The rest of this paper is organized as follows: Section II provides an overview of the conventional CIR-based motion tracking methods, while Section III outlines the major challenges in motion tracking. Our approach is detailed in

Section IV, where SWIFTTRACK+'s algorithmic series for precise and robust fast hand motion tracking in real-world scenarios are elaborated. The performance evaluation of our system is presented in Section V. Section VI reviews related work, followed by discussions in Section VII, and the paper concludes in Section VIII.

II. PRELIMINARY

In this section, we provide an overview of CIR-based acoustic tracking schemes as utilized in [12] and [13], which are notable for their high tracking accuracy and ability to differentiate multiple targets with distinct delays.

A. Estimating Transmission of the Reflected Signal

For channel estimation, the transmitter emits a known acoustic sequence (derived from a training sequence) within the inaudible range (e.g., above $17kHz$). There are a bunch of training sequences can be used for acoustic tracking in literature [11], [12], [13], [27]. The choice of training sequence under fast movement will be discussed in Sec. IV-B.1 and here we use $x[n]$ to denote our training sequence. To ensure the acoustic signals remain inaudible, we first transform the root sequence from the time domain to the frequency domain using a Fast Fourier Transform (FFT), then pad zeros to confine its bandwidth within the inaudible range (e.g., $17-23KHz$). Subsequently, we convert it back to the time domain via an Inverse Fast Fourier Transform (IFFT). The baseband signal $x_t[t]$ length becomes $N = N_r \times \frac{f_s}{bw}$, where N_r is the length of the root sequence $x[n]$, bw is the required bandwidth, and f_s denotes the sampling rate. The up-conversion of the baseband signal to the inaudible band is achieved by multiplying it with $exp(j2\pi f_c t)$, where f_c is the central frequency (e.g., $20KHz$ for $17-23KHz$).

The receiver, such as a built-in microphone on a smartphone, captures the reflected acoustic signal and performs down-conversion to retrieve the baseband signal ($x_r[t]$). As the receiver operates as a Linear Time-Invariant (LTI) system, the received base-band signal can be modeled as:

$$x_r[t] = \sum_i A_i(t) x_t[t - \tau_i(t)] = h[t] * x_t[t] \tag{1}$$

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 $h[n]$ represent the discrete output of $h[t]$, which can be formulated as:

$$h[n] = \sum_i A_i(n)\delta[n - \tau_i(n)] \tag{2}$$

where $\delta[n]$ is a discrete Dirac's delta function [28], effective only when $n = \tau_i(t)$. The phase change at the target tap is directly proportional to the delay change of the target, facilitating the tracking of moving objects using the CIR phase. Following existing methodologies [13], [17], we estimate the channel response $h[n]$ by correlating the received signal $x_r[n]$ with the transmitted signal $x_t[n]$.

Before estimating the target's distance, it is essential to eliminate background multipath effects, including the direct path from the speaker to the microphone and reflections from static surroundings. Various background removal algorithms exist in the literature, such as LEVD [11], DDBR [29], and direct subtraction [15]. For simplicity, we employ direct subtraction, measuring the background interference in the absence of the target and then subtracting it from the estimated channel ($h[n]$) in the presence of the target, akin to the approach in [15] and [24].

B. Achieving Fine-Grained Tracking With Phase Measurement

In the realm of acoustic sensing, precise target tracking is of paramount importance. Two distinct pieces of information time-of-flight (ToF) and phase-based measurements are instrumental in deriving the displacement of a target.

Coarse-grained tracking using time delay: The channel tap, designated as $\hat{n}$, which maximizes the magnitude of the CIR, is indicative of the signal delay. Hence, we can ascertain the delay based on $\hat{n}$. However, this estimation is subject to inaccuracies stemming from limitations in audio sampling frequency, ambient noise, and multipath interference. Consequently, this estimation serves primarily as a coarse indicator of the initial position.

Fine-grained tracking using phase measurement: In contrast, the phase of the selected tap offers more precision and facilitates a fine-grained estimation of the target's displacement. As outlined in [12] and [13], the displacement is calculated from the phase difference as follows:

$$\Delta dist = \frac{c}{f_c} \frac{\Delta Phase}{2\pi} \times \frac{1}{2} \tag{3}$$

Here, the result is halved to account for the round-trip path of the signal in a device-free setup.

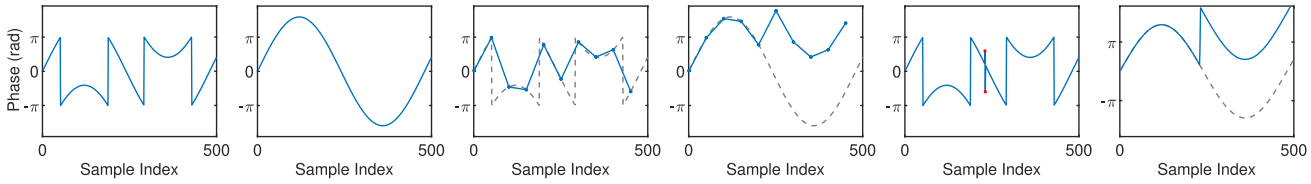
Subsequently, by integrating the coarse-grained initial position with the fine-grained displacement estimation, we can deduce the target's current position, following methodologies established in [12] and [13].

III. CHALLENGES

A. Consequences of Fast Movement

Rapid movement in acoustic tracking introduces two primary issues: Doppler shift and phase ambiguity. These are discussed in more detail below.

1) Impact of Doppler Effect: Consider a scenario where a target moves at a speed of $2m/s$ and the frequency of sound waves is $20kHz$. This movement can induce a frequency shift of approximately $233Hz$ in a round-trip. Given that the speaker emits periodical signals with a $10ms$ period, the frequency spacing will be around $100Hz$. Therefore, a $2m/s$ movement can result in a non-negligible frequency shift (*i.e.*, $233Hz$), which subsequently affects the down-conversion process, leading to distortions in the channel response $h[n]$. Such pulse distortion significantly impacts the accuracy of tap selection and phase measurement.



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

Fig. 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.

2) Impact of Phase Ambiguity: The phase measured from channel $h[n]$, as per Eq. 2, is inherently wrapped within the range $(-\pi, \pi)$. Phase unwrapping is the process of reconstructing the original phase from this wrapped value [30]. This concept is illustrated in Fig. 2(a), where the correct phase is reconstructed by removing "phase jumps". As stated by Itoh [31], this reconstruction is feasible if the smoothness condition, represented by

$$|\Delta\phi_n| \leq \pi \tag{4}$$

is satisfied. For a comprehensive understanding of the phase unwrapping algorithm, refer to [30] and [31].

As depicted in Fig. 2(b), insufficient phase samples can result in phase changes greater than π, leading to unwrapping failures. Additionally, as shown in Fig. 2(c), sudden phase changes induced by noise can also contribute to phase ambiguity. Fast movement exacerbates these issues, as it often results in an inadequate sampling rate for phase measurements, further intensifying phase ambiguity.

While advanced phase unwrapping techniques such as the Path Following Algorithm (PFA) [32], [33], [34] and Quality-Guided Algorithm (QGA) [32], [35], [36] are prevalent, they are predominantly designed for two-dimensional phase unwrapping. These methods utilize a carefully chosen unwrapping path to circumvent regions of ambiguity. In two-dimensional phase data, ambiguity is effectively detected through closed path loops [32]. However, one-dimensional phase data presents a unique challenge: closed loops are not applicable for identifying ambiguous regions, because there is only a single unwrapping path. This makes dealing with ambiguity in one-dimensional data particularly challenging, especially under the fast movement conditions in real-world implementation.

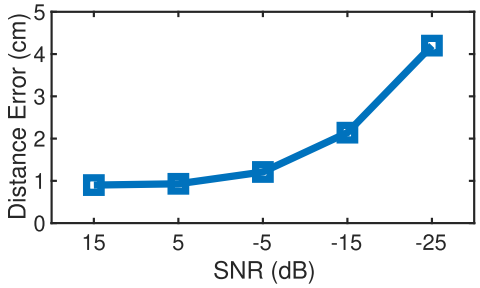
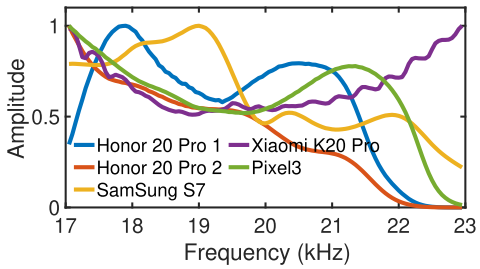
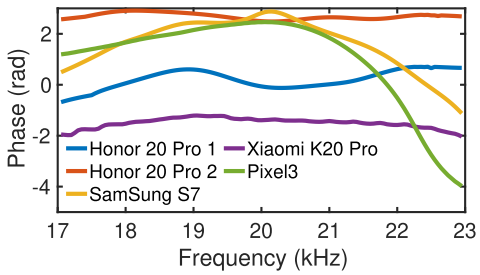


Fig. 3. Accuracy change with the decrease of SNR.



(a) Amplitude response.



(b) Phase response.

Fig. 4. Frequency response of 5 different devices.

B. Impact of Low SNR

SNR plays a critical role in tracking accuracy. Phase-based methods are particularly sensitive to SNR variations, as the error accumulates when integrating phase changes over time for distance estimation [23]. As depicted in Figure 3, a decrease in SNR from $15dB$ (equating to a $30cm$ distance between the target and the phone) to $-25dB$ (at a distance of $150cm$) significantly increases the distance estimation error in Strata from $0.93cm$ to $4.25cm$.

C. Impact of Hardware Frequency Response

In commodity mobile devices, frequencies above $15kHz$ are less audible and typically not optimized, resulting in uneven frequency responses from the speakers and microphones. This uneven response can cause substantial distortion in signals and inaccuracies in the generated channel impulse response as defined in Eq. 2. While several previous studies (*e.g.*, [26], [37]) have addressed and compensated for uneven amplitude across frequencies, our work sheds light on the uneven phase across different frequencies. As illustrated in Figure 4, there is a significant variation in phase across frequencies. Such pronounced phase changes can introduce tracking errors, underscoring the importance of measuring and compensating for both the uneven phase and amplitude.

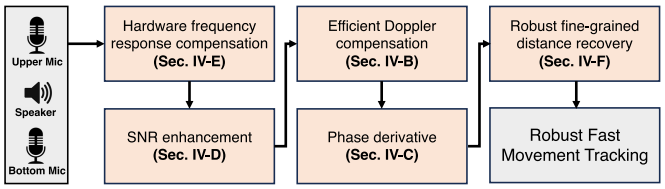


Fig. 5. Signal processing pipeline of SWIFTTRACK+.

IV. OUR APPROACH

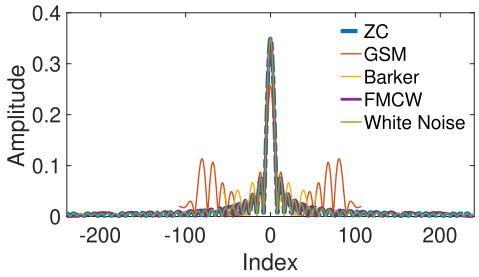
This section firstly provides an overview of the pipeline of SWIFTTRACK+ design, then elaborates on our strategies to address the challenges of fast movement, low SNR, and uneven frequency response in acoustic tracking.

A. System Overview

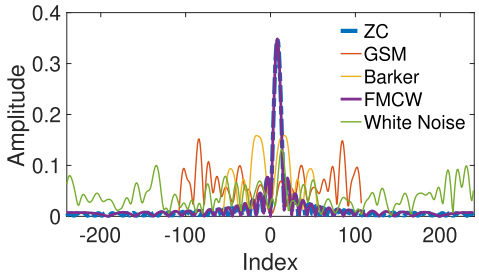
SWIFTTRACK+ utilizes inaudible acoustic signals in the range of $17kHz$ to $23kHz$ for fine-grained, device-free hand motion tracking. Surpassing previous works, SWIFTTRACK+ fundamentally enables tracking of fast movements and enhances robustness in real-world applications through a series of novel algorithms. Fig. 5 illustrates the signal processing pipeline of SWIFTTRACK+. The system emits inaudible sounds from a built-in speaker of COTS mobile devices, which travel in a straight line, reflect off a moving hand, and are then captured by built-in microphones. The process begins with compensation for both amplitude and phase distortions due to hardware frequency response. This is followed by an SNR enhancement algorithm inspired by time-domain beamforming, aimed at ensuring robust tracking, particularly over longer distances. Subsequently, Doppler shift caused by movement is compensated, and corresponding channel taps are selected for time delay and phase measurement. The phase derivative method is then applied to resolve phase ambiguity associated with fast movements, enabling precise measurement of fine-grained velocity. Lastly, the finely estimated velocity and coarsely determined tap index (*i.e.*, the time delay) are fed into an LSTM-based distance estimator for robust distance recovery. To elucidate our approach in the following discussion, we begin with the core aspect of facilitating fast movement tracking - tackling high mobility challenges. This is addressed in two aspects: (i) choosing an appropriate training sequence with anti-Doppler property and compensating for Doppler shift to enhance tap selection (Sec. IV-B), and (ii) implementing our phase derivative method to circumvent phase ambiguity (Sec. IV-C). Additionally, due to the complexities of real-world deployment, such as long-distance tracking and the heterogeneity of COTS devices, we introduce algorithms for SNR enhancement (Sec. IV-D), hardware frequency response compensation (Sec. IV-E), and the LSTM-based robust distance recovery (Sec. IV-F).

B. Training Sequence and Doppler Compensation

Fast movement primarily affects acoustic tracking through the Doppler effect, which reduces the correlation between the received and transmitted signals. This poses significant challenges in designing the training sequence. We explore the choice of training sequence for scenarios involving rapid movement and discuss methodologies to effectively compensate for the Doppler effect caused by such motions.



(a) Correlation w/o Doppler effect



(b) Correlation w/ Doppler effect ($\Delta f = 100Hz$)

Fig. 6. The correlation of different training samples under severe Doppler effect.

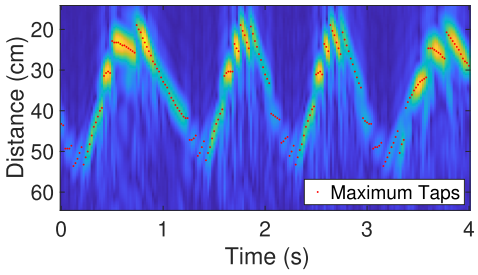
1) Training Sequence Under Fast Movement: There are many pseudo-random sequences can be used for training sequence, including GSM [38], Barker [39], and Zadoff-Chu (ZC) [40]. We opt for the ZC sequence, widely used in modern cellular systems like LTE and 5G NR [41], defined as:

$$ZC[n] = exp(-j\frac{\pi\mu n(n+1)}{N_{ZC}}) \tag{5}$$

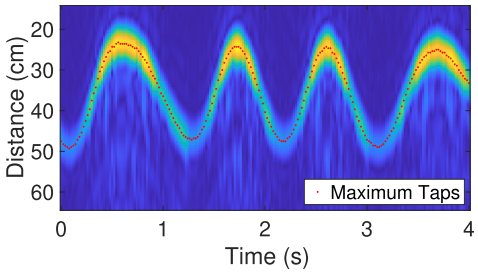
where N_{ZC} is the sequence length, $0 \leq n < N_{ZC}$, $0 < \mu < N_{ZC}$ and $gcd(N_{ZC}, \mu) = 1$. Following the procedures outlined in Sec. II-A, we convert these training sequences into inaudible sounds for transmission.

Fast movement scenarios, which induce substantial Doppler shift, can affect the performance of different transmitting signals. Fig. 6 presents a simulation comparing the correlation performance of the aforementioned training sequences along with Frequency-Modulated Continuous-Wave (FMCW) [42], [43] and white noise [14]. While Fig. 6(a) indicates all five transmitting signals exhibit favorable correlation properties in the absence of Doppler shift, Fig. 6(b) reveals that in the presence of a $100Hz$ Doppler shift (equivalent to approximately $85m/s$ moving speed in a device-free system), only the ZC and FMCW sequences still maintain a high correlation peak, while the others showing negligible correlation. This is possibly due to the fact that the phase change of ZC and FMCW signals in the frequency domain is quadratic [44], while the frequency shift caused by the Doppler effect introduces only a linear phase change; thus the correlation peak is still preserved.

This indicates that high-confidence channel tap selection is possible using ZC and FMCW signals, even under fast movement conditions. In SWIFTTRACK+, we use the ZC signal, but our proposed algorithms for enabling fast movement tracking



(a) Pulse distortion caused by motion



(b) After Doppler compensation

Fig. 7. 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.

are also applicable to systems employing FMCW signals, since both ZC and FMCW signals can be used for CIR estimation and FMCW signals can also be transformed into signals with flat spectrum [42], like ZC signals.

2) Efficient Doppler Compensation for Tracking: Although selecting high-confidence channel taps (with sufficient power) is achievable, we observe that motions still significantly distort the CIR profiles in practice. This distortion is evident in Fig. 7(a), which was generated by tracking a hand-sized object moving between $20cm$ and $55cm$. Such distortions can lead to substantial errors in both CIR estimation and tap selection. To mitigate these effects, compensating for the Doppler shift is necessary to reduce the distortions in the CIR. Drawing inspiration from frequency synchronization techniques commonly used in digital communication systems [45], [46], we process the baseband complex signal by exploring different Doppler shifts and choosing the one that maximizes the peak value of the CIR profiles. The effectiveness of this approach is illustrated in Fig. 7(b), where the CIR profile and tap selection appear significantly smoother, indicating better accuracy. However, while this method is simple and effective, it is not sufficiently efficient for implementation on mobile devices.

Previous works like Strata, LLAP, and VSkin have shown accurate tracking performance when the target's velocity is below a certain maximum supportable velocity, v_{max}, corresponding to a Doppler shift of Δf_{vmax}. This suggests that Doppler shifts below Δf_{vmax} only introduce negligible errors. Therefore, for tracking purposes, our Doppler compensation can be relatively coarse. We aim to ensure the frequency shift between the compensated and template frames is smaller than Δf_{vmax}. Consequently, we search for the velocity using a step size empirically set to $10cm/s$. Given that the peak acceleration of human arms is approximately $30m/s^2$ [47], the maximum velocity change between two consecutive frames is about $30cm/s$. Accounting for 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 previous velocity estimate and the initial velocity is assumed to be 0 due to the target starting from a stationary position. Recent studies indicate that time domain cross-correlation is memory efficient, but frequency domain cross-correlation is more time efficient [48], [49]. Therefore, we utilize frequency domain cross-correlation [50] to expedite the cross-correlation step when enumerating potential Doppler shifts. Additionally, we employ the FFTW library [51] to enhance the efficiency of FFT computations.

C. Enable Fine-Grained Fast Hand Motion Tracking

To facilitate fine-grained fast hand motion tracking, measuring the phase change of each channel tap is crucial. However, even with smoother tap selection, ambiguity between consecutive channel taps persists. This section first examines the impact of fast movement on phase and the limitations of conventional solutions. We then introduce our core solutions to overcome phase ambiguity in fast motions. Lastly, we discuss real-world implementation issues, motivating further improvements by addressing other acoustic tracking challenges (*i.e.*, enhancing SNR in Sec. IV-D, adjusting for hardware frequency response in Sec. IV-E, and robust distance recovery in Sec. IV-F).

1) Why Fast Movement Causes Phase Ambiguity: The small wavelength of inaudible acoustic signals, while enabling precise motion tracking, also increases the likelihood of phase wrapping. A displacement as small as $1.73cm$ at $20kHz$ can result in a 2π phase change, leading to phase wrap-around. Accurate displacement determination thus requires phase unwrapping, which requires the smoothness condition in Eq. 4 and is crucial for avoiding phase ambiguity. The phase measurement $\phi_d[t]$ relates to the distance $d[t]$ as: $\phi_d[t] = -\frac{2\pi f_c}{c} d[t] \times 2 = -\frac{4\pi f_c}{c} d[t]$. To satisfy the smoothness condition, we derive:

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

With $c = 343m/s$, $f_c = 20kHz$, and $T = 10ms$, the maximum velocity should satisfy: $|v[t]| \leq 0.43m/s$. Since hand movements can reach up to $2.7\ m/s$ [47], phase changes may easily exceed π, causing significant errors in distance estimation. The red line in Fig. 8 illustrates the fast hand motion tracking using the phase of the channel taps after Doppler compensation shown in Fig. 7. As we can see, even though the CIR profile appears smooth after Doppler compensation, the phase measurements remain ambiguous due to the fast motions. **In a nutshell, fast movement makes phase change faster and the sampling rate (we accumulate**

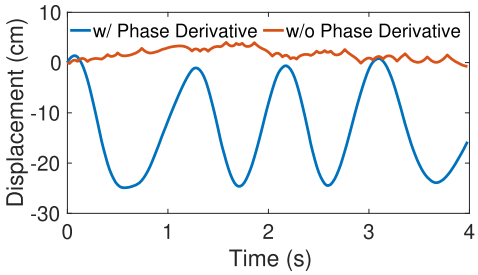


Fig. 8. Phase derivative can effectively resolve phase ambiguity and recover the fine-grained trajectory in fast hand motion tracking.

one phase sample at each frame) is not sufficient for reconstructing the true motions. This is a more acute issue in acoustic sensing, where speeds exceeding $0.43m/s$ are common, unlike in radar systems where RF signals propagate at much higher speeds.

2) Conventional Solutions to the Under-Sampling Issue: To accommodate rapid movements in acoustic tracking, reducing the frame length used for channel estimation is a common strategy. For instance, to support velocities up to $2.7m/s$, the frame length might be decreased to as short as $1.25ms$ according to Eq. 6. However, this reduction negatively impacts both detection range and SNR [52], [53]. Another method involves increasing the sampling frequency while keeping the frame duration constant, which is typically achieved by fractionally sliding the window. This technique, used in previous works like [13] and [54], becomes ineffective when motion exceeds the maximum supported speed as observed in our tests. Therefore, our goal is to circumvent phase ambiguity in situations of fast motion while maintaining practical frame lengths (e.g., $10ms$) and channel estimation intervals (e.g., $10ms$).

3) Applying Phase Derivative to Circumvent Phase Under-Sampling: Our core observation is that the phase quotient of two consecutive taps can be utilized to estimate motion:

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

Unwrapping the phase of $p'[t]$, denoted by $\phi_v[t]$, can derive the fine-grained velocity estimation. Interestingly, using the phase derivative for tracking avoids phase ambiguity until much higher thresholds, which offers a viable solution for tackling fast motion issue:

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

This indicates no phase ambiguity as long as the acceleration of our moving arms stays below $42.3\ m/s^2$ over a $10ms$ frame. Given peak human hand acceleration is about $30m/s^2$ [47], phase derivative tracking should not incur phase ambiguity. The blue line in Fig. 8 shows the tracking results by applying the phase derivative method, which successfully recovers the fine-grained trajectory of the fast-moving hand. **In essence, applying phase unwrapping to the first-order phase derivative effectively circumvents the under-sampling issue caused by fast human movement.** We assume the target starts from a static position (*i.e.*, zero initial velocity and acceleration). Additionally, this method could extend to higher-order phase derivatives for even greater mobility. Our phase derivative method differs from conventional path following algorithms (PFA) and quality-guided algorithms (QGA) in two aspects: (i) it is applicable to one-dimensional phase data, and (ii) it directly applies basic unwrapping algorithms to the phase derivative without requiring ambiguity detection.

4) Issues in Real-World Implementation: While the phase derivative approach is theoretically sound and effective in simulations, real-world implementation presents additional issues: (i) significant sudden phase changes (caused by low SNR or signal distortion), as exemplified in Fig. 2(c), may lead to phase unwrapping failure, resulting in errors when converting fine-grained velocity into distance, (ii) low SNR conditions may impair the effectiveness of the Doppler compensation algorithm. These phenomena are largely due to the rapid attenuation of acoustic signals and distortion in hardware frequency response, which are common challenges in motion tracking. Therefore, we address these issues in Sec. IV-D and Sec. IV-E to enhance the robustness of fast hand motion tracking. These algorithms are also applicable to other acoustic tracking systems requiring SNR enhancement and hardware frequency compensation.

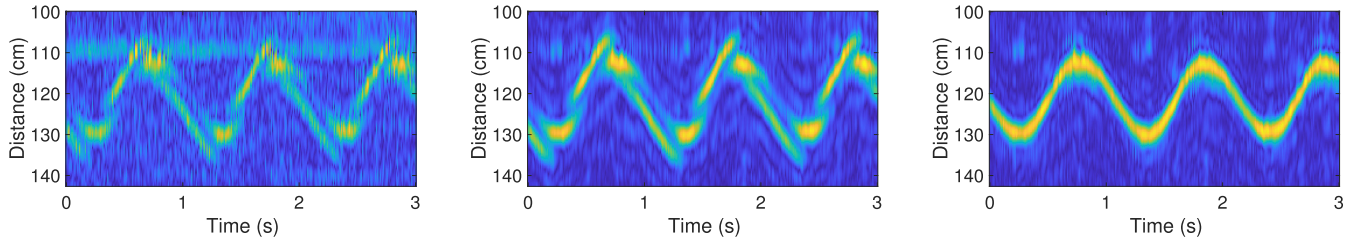
D. SNR Enhancement of CIR Profiles

The SNR in acoustic tracking can be adversely affected by the rapid attenuation of the reflected signal and the limited power output of mobile device speakers. This often results in noisy phase measurements of the taps corresponding to the target, as illustrated in Fig. 9(a), leading to errors in Doppler estimation and phase measurement. The CIR is modeled as:

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

where $w[n]$ represents noise, $\hat{h}[n]$ is the measured CIR, and $g[n]$ is the distortion caused by Doppler shift. Our objective is to minimize the impact of $w[n]$ in $\hat{h}[n]$. Our approach is inspired by the principle of antenna beamforming, where signals received at different times can be constructively combined after compensating for phase differences caused by varying positions [24]. This method is effective even with a single microphone, as CIRs are summed over time for each microphone.

To compensate for phase changes between consecutive frames, we note that CIRs are typically rotated by a phase



(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.

Fig. 9. Time domain beamforming effectively improves the SNR of the CIR profile.

caused by motion. This relationship can be expressed as:

$$h[n] = h[n-1]e^{j\phi} \tag{10}$$

where ϕ is the phase change due to motion. This model is valid for a single moving object. For multiple well-separated moving objects, each object influences a subset of taps, and the equation holds for the taps adjacent to each object. We estimate the next CIR using the Exponential Weighted Moving Average (EWMA) [55] as follows:

$$z[n] = (1-K) \times z[n-1]e^{j\hat{\phi}} + K \times \tilde{h}[n] \tag{11}$$

where $z[n]$ is the SNR-enhanced CIR, $\tilde{h}[n]$ is the measurement, and $z[0] = \tilde{h}[0]$. The weighting factor K is empirically set to 0.3. $\hat{\phi}$ is the best phase estimate aligning $z[n-1]e^{j\hat{\phi}}$ with $\tilde{h}[n]$.

To determine $\hat{\phi}$, we employ the maximum entropy principle [56] in the following optimization model:

$$\underset{\phi}{\operatorname{argmax}}\, H(normalized(|\tilde{h}[n] - \hat{z}[n]| \cdot |\hat{z}[n]|)) \tag{12}$$

where $H(\cdot)$ represents the Shannon entropy and $\hat{z}[n] = z[n-1]e^{j\hat{\phi}}$ is the predicted next channel. The term $|\tilde{h}[n] - \hat{z}[n]|$ is the amplitude of the error between the measured and predicted CIR, spread across all taps. The dot product between $|\tilde{h}[n] - \hat{z}[n]|$ and $|\hat{z}[n]|$ prioritizes minimizing errors closest to the target. The maximum entropy principle suggests that without prior knowledge of a distribution, the most uniform distribution (maximum entropy) should be chosen. Hence, we maximize the Shannon entropy of this error function to evenly distribute weighted residual errors across each tap. An iterative gradient descent algorithm is used to solve this optimization problem.

An example of an SNR-enhanced profile is shown in Fig. 9(b), where noise is effectively suppressed. Again, we assume the object starts from a static state with zero initial phase change (*i.e.*, $\hat{\phi}_0 = 0$). Additionally, as shown in Fig. 9(b), our algorithm can suppress residual multipath and background noise, enhancing tracking accuracy for long-range detection.

Given the short frame duration (*i.e.*, $10ms$), two consecutive CIR profiles are likely affected by a similar Doppler distortion $g[n]$. Thus, Eq. 10 holds even with the presence of $g[n]$, indicating that our SNR enhancement algorithm is also effective under Doppler shift. After obtaining the SNR-enhanced CIR profile, $z[n,t]$, we need to convert the CIR profiles back to a baseband signal to compensate for the Doppler distortion $g[n]$. Inspired by frequency domain cross-correlation, the baseband signal from a CIR profile is approximated as follows:

$$\hat{rx}[n,t] \approx iFFT(FFT(z[n,t]) \cdot FFT(zc[n])) \tag{13}$$

The Doppler shift in $\hat{rx}[n,t]$ is then compensated, and the resulting signal is converted back to a CIR profile, as depicted in Fig. 9(c).

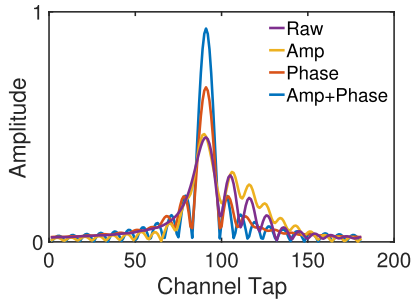


Fig. 10. Different hardware frequency compensation schemes.

E. Compensating Hardware Frequency Response

Speakers and microphones commonly exhibit an uneven frequency response. The received signal r is influenced by both the wireless channel and the hardware frequency response, described as: $ZC_R[n] = h[n] * h_h * ZC_T[n]$, where h_h represents the hardware frequency response. It's important to note that h_h is a complex number, indicating that the hardware frequency response affects both the amplitude and phase of the received signal.

Prior studies primarily focus on the amplitude aspect of the hardware frequency response (*e.g.*, [26], [37]). However, our findings highlight the significance of the phase of the frequency response. In the absence of hardware distortion, the phase response should be linear with respect to frequency, but hardware distortion results in a nonlinear phase response [57].

Fig. 10 demonstrates the effectiveness of different frequency compensation schemes within the $17-23KHz$ range on a Google Pixel3: (i) compensating for phase response alone, $\angle h_h$, (ii) compensating for amplitude response alone, $|h_h|$, and (iii) compensating for both phase and amplitude responses, h_h. The purple line indicates the raw CIR profile. Without compensation, two issues are evident: (i) the diminished energy of the main lobe, which leads to reduced SNR,

and (ii) the increased energy of side lobes, potentially causing more distortions and complicating the estimation of Doppler shifts and tap selection. Amplitude response compensation alone, as adopted in previous works [26], [37], yields some improvement. However, compensating for the phase response not only narrows the main lobe but also reduces the side lobes. Furthermore, compensating for both phase and amplitude results in the highest peak value. Consequently, we address hardware distortions by compensating for both the phase and amplitude responses.

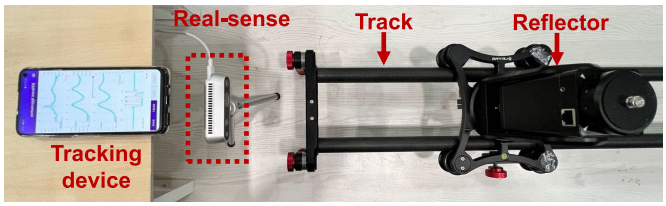
In Section V, we extend this evaluation to various compensation schemes on different phones and frequencies, demonstrating that their relative effectiveness depends on the specific hardware and frequency range. Nonetheless, compensating for the phase of the frequency response is universally crucial in all scenarios.

F. Robust Position Estimation

By applying the phase derivative method, we effectively circumvent phase ambiguity issues arising from fast motions. However, this approach yields unwrapped phases that represent the target's velocity. This section discusses how to robustly reconstruct the position of our target from this recovered velocity.

1) LSTM-Based Robust Fine-Grained Distance Reconstruction: A straightforward method to reconstruct fine-grained displacement is integrating the estimated velocity derived from the phase derivative. However, as discussed in Sec. IV-C, sudden noise can cause abrupt changes or 'jumps' at some samples, leading to errors after phase unwrapping. If we integrate the velocity over time, these errors can accumulate. Although we have proposed SNR enhancement and hardware frequency response compensation algorithms in Sec. IV-D and Sec. IV-E, respectively, sudden noise instances may still occasionally occur when the object moves at a longer distance. To address this issue, we incorporate both the fine-grained velocity estimation (which has less noise but occasional longer-term errors) and the coarse-grained distance estimation, *i.e.*, the time delay, from channel taps (which exhibits more noise but no longer-term errors) into a simple LSTM model, which consists of two LSTM layers (with input size 16 and hidden size 64) and conventional layers at the input and output. Since the time interval between two consecutive estimations is $10ms$, the time delay by using the LSTM model is $80ms$ (we have fine-grained velocity and coarse-grained distance at each estimation, respectively), which is acceptable for user interactions. This model combines these two types of data to achieve robust distance recovery even with the existence of significant noise in the phase derivative, which further extends the working range of SWIFTTRACK+.

2) 2D Positioning: Given the known relative positions of the speaker and two microphones on a phone, we can calculate the path length from the speaker to one microphone. Consequently, the potential locations of our target should lie on an ellipse in 2D space, with the speaker and microphone serving as foci, as per the principles of acoustic triangulation [2], [11], [12]. Utilizing two microphones, the target can be pinpointed at the intersections of two ellipses. Typically, there are two intersection points, but we select the one located in front of the tracking device, which is a common area for user interaction, such as tracking hand movements.



(a) Track scenarios

(b) Real-life scenarios

Fig. 11. Experiment setup.

V. EVALUATION

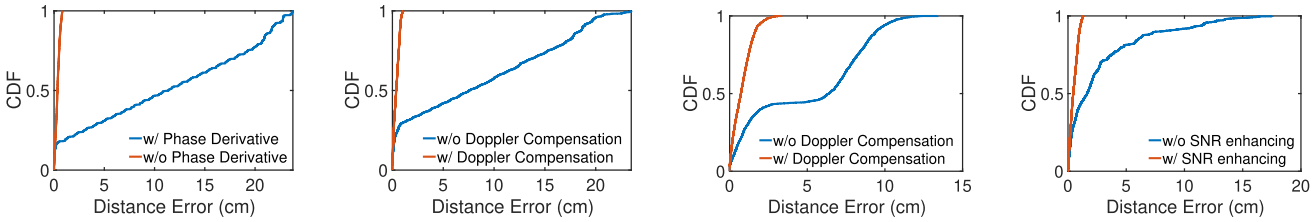
A. Experiment Setup

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 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 and push it with human arms, as shown in Fig. 11(a). We placed a Real-Sense D435i [58] 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 [59] 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. 11(b). We attach a red marker to the center of the user's palm, which facilitates the Real-Sense to get ground-truth positions.

B. Micro-Benchmark

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



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

Fig. 12. Performance of (a) phase derivative scheme, doppler compensation scheme for (b) phase-based and (c) tap-based, (d) SNR enhancement.

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. 12(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.

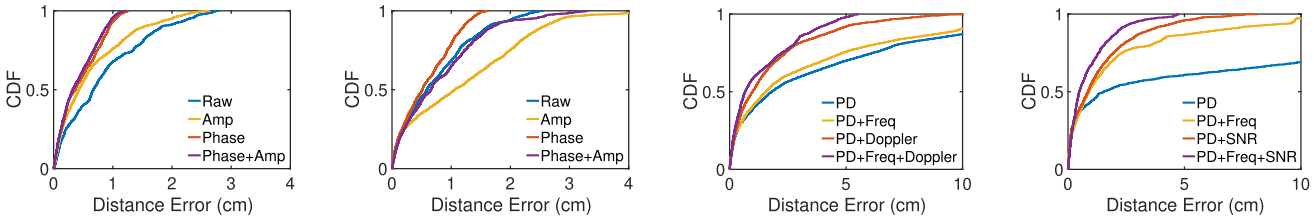
2) Doppler Shift Compensation: We evaluate the impact of pulse distortions on tracking performance. Since the pulse distortion is related to the moving speed and high moving speed will cause an insufficient phase sampling rate, we apply the circular shift-based up-sampling scheme to decouple the pulse distortion and sampling interval. Then the tracking performance is mainly determined by the pulse distortion. Fig. 12(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. 12(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.

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 Fig. 12(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 $0.91cm$).

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. 13 (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. 13(a). We use $17kHz$ to $23kHz$ in the second experiment for Fig. 13(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) Robust Fine-Grained Distance Recovery: We evaluate the LSTM-based robust fine-grained distance recovery scheme with the presence of residual sudden noise in the phase derivatives (*e.g.*, tracking a moving hand in the distance from $150cm$ to $200cm$), where the errors at some points of the unwrapped phase (represent the estimated velocity) will accumulate over time if we directly integrate the estimated velocity, making the reconstructed distance severely drift. Previous works [2], [11], [12] also adopted a fusion approach which combines the coarse-grained distance and the fine-grained displacement to achieve better performance by simply using a weighting factor. Fig. 14 compares the performance of our LSTM-based approach and the conventional weighting factor-based approach with the presence of sudden noise change when we recover the fine-grained velocity to the distance of the target, which shows that the LSTM-based approach can significantly enhance the tail performance when there exist sudden change in the phase derivate, reducing the 90^{th} percentile error by 70% from $8.50cm$ to $2.51cm$. The failure of the weighting factor-based approach implies that it cannot effectively tackle the significant error accumulation when integration is required.

6) 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 experiments under different



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

Fig. 13. 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 ($120cm$).

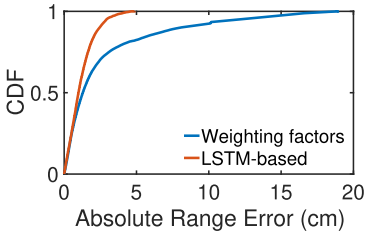


Fig. 14. Robust distance reconstruction with the presence of residual sudden noise.

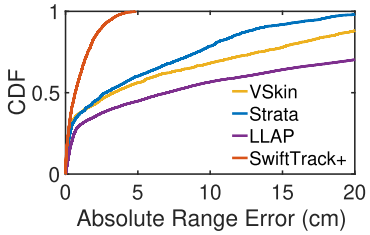


Fig. 15. Overall comparison of ranging errors.

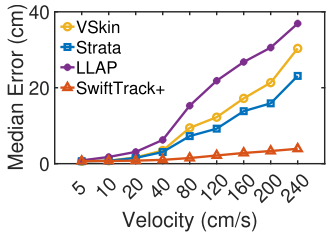


Fig. 16. Impact of velocities.

conditions to evaluate their performance. The results are shown in Fig. 13 (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. 13(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. 13(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%.

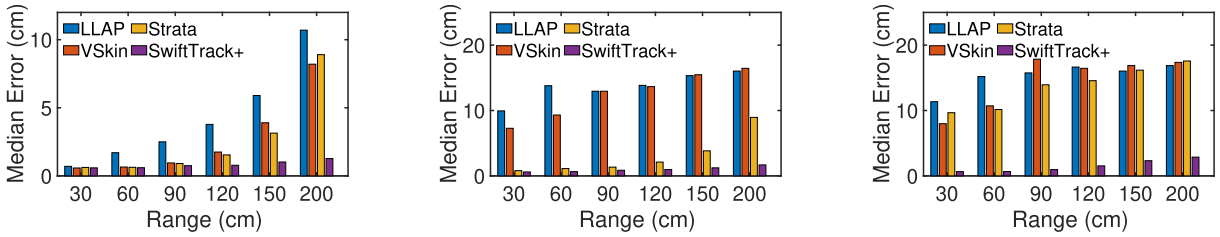
C. Overall Comparison

In this section, we compare the performance of our system with the following previous work: Strata [12], VSkin [13], and LLAP [11]. 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. 15 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.

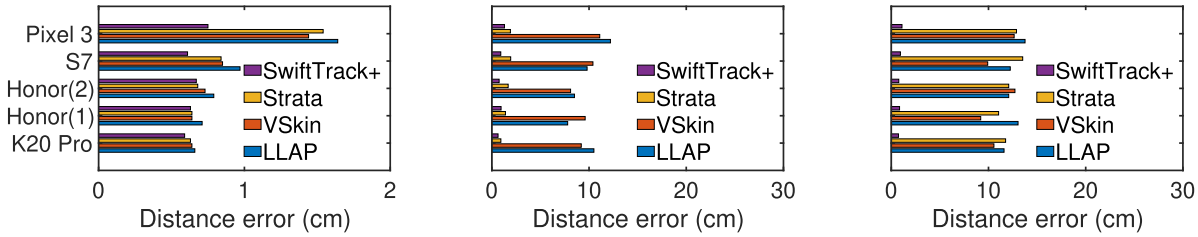
1) Impact of Velocities: We further evaluate SWIFTTRACK+ under various scenarios. We first compare the tracking errors under various velocities from $5cm/s$ to $240cm/s$ to cover the speed range of human hand motion. Since human arm's movement is unpredictable in real scenarios and the errors caused by fast motions may accumulate over time, we calculate the mean error of each trace and the maximum velocity. Fig. 16 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 ($3.69cm$).

2) Impact of Distances: We test the system performance at various distances from $30cm$ to $200cm$ 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. 17. We can make the following two observations. First,



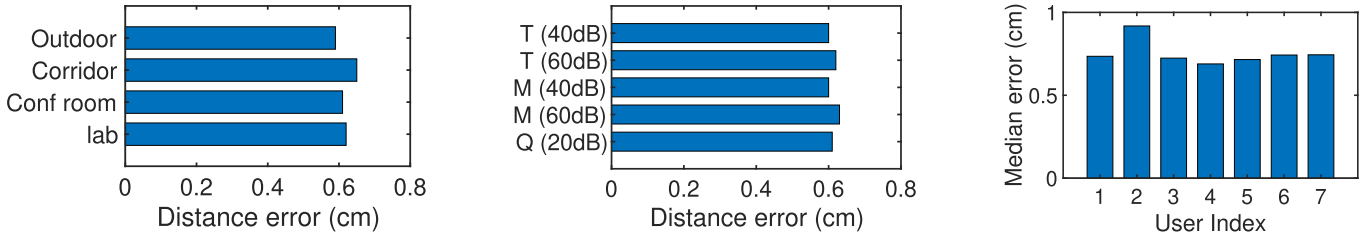
(a) Low speed. (b) Medium speed. (c) High speed.

Fig. 17. Impact of distances at low speed range ($\leq 20cm/s$), medium speed range ($20cm/s \sim 80cm/s$) and high speed range ($\geq 80cm/s$).



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

Fig. 18. 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

Fig. 19. Impact of (a) locations, (b) ambient noise, and (c) users.

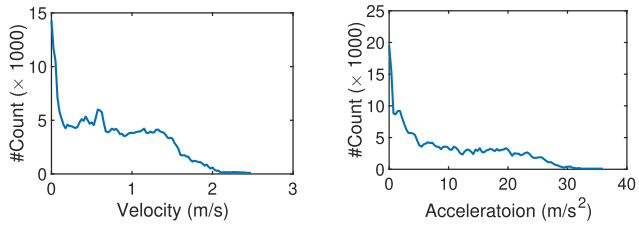
when the speed is high, only SWIFTTRACK+ works well at various distances, while the other methods suffer from fast motion. Second, when the speed is low or medium, thanks to the SNR enhancement, frequency response compensation schemes, and LSTM-based distance recovery, SWIFTTRACK+ can still out-perform the other methods.

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. 18. 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.

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. 19(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.

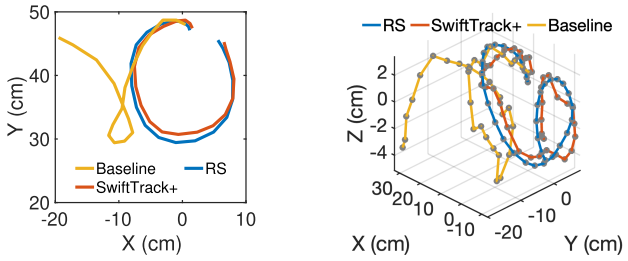
5) Impact of Ambient Sounds: We evaluate SWIFTTRACK+ with different ambient sounds. Specifically, we conduct experiments in three scenarios: (i) a quiet environment (labeled as "Q" in Fig. 19(b)), (ii) an environment with people talking ("T"), and (iii) an environment with music playing ("M"). We place 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. 19(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.

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. The results are shown in Fig. 19(c) and we can make two observations. First, though the speed may change significantly



(a) Speed Distribution (b) Acceleration Distribution

Fig. 20. Statistics of collected data.



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

Fig. 21. 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.

during each test (i.e., including significant acceleration and deceleration since users are not required to move theirs hands at a constant speed), the median error over the seven participants is $0.74cm$, which is just slightly higher than that of a hand-sized reflector ($0.63cm$). The slightly higher error is more likely because arms introduce additional reflection and arms' movement may differ from that of the hand. Second, the tracking accuracy across different users is similar, while user 2 has slightly worse performance possibly due to that she has a smaller hand size (female) with less reflective area.

D. Motion Statistics

We measure the speed distribution and acceleration distribution during the experiments. The results are shown in Fig. 20. 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 indicats the phase derivative algorithm can handle fast human motion tracking.

E. Drawing Samples

We build a draw-in-the-air interface based on SWIFTTRACK+ to show its drawing capability in 2D and 3D spaces according to experiment setup in Strata [12] and FMTrack [15]. 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 [60]. We draw circles quickly with our hands in the air (the 3D example) or on the desk (the 2D example). The maximum velocity is $132cm/s$. Fig. 21 gives the drawing examples with fast motion. We can see that the baseline deviates significantly from the reference trajectories while SWIFTTRACK+ follows.

TABLE I
PROCESSING TIME OF EACH PART

Doppler Shift	SNR enhancement	Others	All
4.37ms	1.26ms	1.21ms	6.84ms

F. 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 I. 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 $6.84ms$. Therefore SWIFTTRACK+ can process each $10ms$ frame in real-time.

VI. RELATED WORKS

Our work intersects with tracking technologies in the following aspects: (i) device-free acoustic tracking, (ii) device-based acoustic tracking, (iii) RF-based tracking, and (iv) phase unwrapping techniques.

Device-free (Contactless) Acoustic Tracking: Device-free acoustic tracking utilizes the speakers and microphones commonly found on COTS mobile devices and achieves high accuracy. FingerIO [2] employs the Time-of-Arrival (TOA) of reflected signals for tracking on mobile devices, using the cyclic suffix property of OFDM symbols. However, its range resolution is limited by the microphone's sampling rate, and phase changes in the frequency domain are noise-sensitive. LLAP [11] uses continuous waves (CW) for tracking, less susceptible to noise but prone to multipath interferences. Strata [12] overcomes this using channel impulse response and appropriate channel tap phase. RTrack [24] combines signal processing with machine learning to extend sensing range, and FMTrack [15] achieves multi-target tracking through iterative parameter optimization.

Device-based Acoustic Tracking: In contrast, device-based tracking involves users holding a device, offering higher SNR and reduced multipath interference effects. AAMouse [61] turns a mobile device into an air mouse by combining Doppler shifts across frequencies. CAT [43] and Rabbit [25] leverage chirp signals for distance estimation, while SoundTrack [62] locates a specialized finger ring in 3D space using CW signal phase information. Millisonic [42] tracks multiple devices using the phase of time-domain mixed chirp signals.

RF-based Tracking: RF signals, including WiFi, RFID, and mmWave, are also used for localization and tracking. ArrayTrack [63] utilizes signal phase for tracking with a median error of $23cm$ using 16 antennas. RF-IDraw [64] and WiDraw [65] achieve centimeter-level accuracy. High-frequency systems like mTrack [29], Soli [66], and mmVib [67] improve performance but require custom hardware.

Phase Unwrapping Techniques: Phase unwrapping is widely used in fields like optical interferometry [68], [69], profilometry [70], SAR [71], [72], and MRI [73], [74], [75]. Key challenges include identifying singularities in two-dimensional phase data and selecting unwrapping paths to avoid these areas, using methods like the residual theorem [32], [33], [34] and quality maps [32], [35], [36]. However, for one-dimensional phase data in tracking systems, identifying and bypassing singularities is more challenging.

VII. Discussion

SwiftTrack+ enhances tracking accuracy and robustness under fast movements and can be easily integrated into other systems to improve SNR and tackle uneven frequency response. Despite the promising results in tracking fast hand motions, several challenges remain for future work: (i) extending the SNR enhancement to multiple channels to further improve performance by leveraging spatial diversity, and (ii) enabling fast movement tracking for various objects by utilizing higher-order phase derivatives to accommodate larger accelerations and decelerations.

VIII. Conclusion

In this work, we enable fine-grained and robust fast movement tracking using acoustic signals on COTS mobile devices for human arms tracking. To achieve this, 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. (iv) LSTM-based distance recovery algorithm combines the history information of phase measurement and coarse-grained time delay to robustly reconstruct the distance of the target, largely extending the working range of SwiftTrack+. 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 median error of $0.63cm$.

Acknowledgment

The authors sincerely thank the reviewers and editors for their time, effort, and valuable feedback, which have greatly contributed to improving this work.

References

[1] Y. Zhang et al., "Addressing practical challenges in acoustic sensing to enable fast motion tracking," in *Proc. 22nd Int. Conf. Inf. Process. Sensor Netw.*, May 2023, pp. 82–95.

[2] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "FingerIO: Using active sonar for fine-grained finger tracking," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2016, pp. 1515–1525.

[3] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-FMCW based contactless respiration detection using acoustic signal," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–20, Jan. 2018.

[4] C. Peng, G. Shen, and Y. Zhang, "BeepBeep: A high accuracy acoustic ranging system using COTS mobile devices," in *Proc. 5th ACM Conf. Embedded Netw. Sensor Syst.* New York, NY, USA: Association for Computing Machinery, 2007, pp. 1–14, doi: 10.1145/1322263.1322265.

[5] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda, "Swordfight: Enabling a new class of phone-to-phone action games on commodity phones," in *Proc. 10th Int. Conf. Mobile Syst., Appl., Services*. New York, NY, USA: Association for Computing Machinery, 2012, pp. 1–14, doi: 10.1145/2307636.2307638.

[6] S. Gupta, D. Morris, S. Patel, and D. Tan, "SoundWave: Using the Doppler effect to sense gestures," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2012, pp. 1911–1914.

[7] M. T. I. Aumi, S. Gupta, M. Goel, E. Larson, and S. Patel, "DopLink: Using the Doppler effect for multi-device interaction," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.* New York, NY, USA: Association for Computing Machinery, Sep. 2013, pp. 583–586, doi: 10.1145/2493432.2493515.

[8] K.-Y. Chen, D. Ashbrook, M. Goel, S.-H. Lee, and S. Patel, "AirLink: Sharing files between multiple devices using in-air gestures," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.* New York, NY, USA: Association for Computing Machinery, Sep. 2014, pp. 565–569, doi: 10.1145/2632048.2632090.

[9] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*. New York, NY, USA: Association for Computing Machinery, May 2015, pp. 45–57, doi: 10.1145/2742647.2742674.

[10] Y. Zhuang, Y. Wang, Y. Yan, X. Xu, and Y. Shi, "ReflecTrack: Enabling 3D acoustic position tracking using commodity dual-microphone smartphones," in *Proc. 34th Annu. ACM Symp. User Interface Softw. Technol.* New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 1050–1062, doi: 10.1145/3472749.3474805.

[11] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 82–94.

[12] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, "Strata: Fine-grained acoustic-based device-free tracking," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*. New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 15–28, doi: 10.1145/3081333.3081356.

[13] K. Sun, T. Zhao, W. Wang, and L. Xie, "VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 591–605.

[14] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in *Proc. 25th ACM Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–16.

[15] D. Li, J. Liu, S. I. Lee, and J. Xiong, "FM-track: Pushing the limits of contactless multi-target tracking using acoustic signals," in *Proc. 18th Conf. Embedded Networked Sensor Syst.*, 2020, pp. 150–163.

[16] Y. Fu, S. Wang, L. Zhong, L. Chen, J. Ren, and Y. Zhang, "SVoice: Enabling voice communication in silence via acoustic sensing on commodity devices," in *Proc. 20th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2022, pp. 622–636.

[17] Y. Zhang et al., "Endophasia: Utilizing acoustic-based imaging for issuing contact-free silent speech commands," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–26, Mar. 2020.

[18] Y. Zhang et al., "Acoustic sensing and communication using metasurface," in *Proc. 20th USENIX Symp. Networked Syst. Design Implement.*, 2023, pp. 1359–1374.

[19] Y. Zhang, Y.-C. Chen, H. Wang, and X. Jin, "CELIP: Ultrasonic-based lip reading with channel estimation approach for virtual reality systems," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. ACM Int. Symp. Wearable Comput.*, Sep. 2021, pp. 580–585.

[20] E. Langbehn, T. Eichler, S. Ghose, K. von Luck, G. Bruder, and F. Steinicke, "Evaluation of an omnidirectional walking-in-place user interface with virtual locomotion speed scaled by forward leaning angle," in *Proc. GI Workshop Virtual Augmented Reality (GI VR/AR)*, 2015, pp. 149–160.

[21] M. P. J. Habgood, D. Moore, D. Wilson, and S. Alapont, "Rapid, continuous movement between nodes as an accessible virtual reality locomotion technique," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2018, pp. 371–378.

[22] K. M. Stanney and P. Hash, "Locus of user-initiated control in virtual environments: Influences on cybersickness," *Presence*, vol. 7, no. 5, pp. 447–459, Oct. 1998.

[23] G. E. Spoorthi, R. K. S. S. Gorthi, and S. Gorthi, "PhaseNet 2.0: Phase unwrapping of noisy data based on deep learning approach," *IEEE Trans. Image Process.*, vol. 29, pp. 4862–4872, 2020.

[24] W. Mao, M. Wang, W. Sun, L. Qiu, S. Pradhan, and Y.-C. Chen, "RNN-based room scale hand motion tracking," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2019, pp. 1–16.

[25] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, "Indoor follow me drone," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2017, pp. 345–358.

[26] W. Mao, W. Sun, M. Wang, and L. Qiu, "DeepRange: Acoustic ranging via deep learning," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 4, pp. 1–23, Dec. 2020.

[27] X. Song et al., "SpiroSonic: Monitoring human lung function via acoustic sensing on commodity smartphones," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, 2020, pp. 1–14.

[28] A. V. Oppenheim, *Discrete-Time Signal Processing*. London, U.K.: Pearson, 1999.

[29] T. Wei and X. Zhang, "mTrack: High-precision passive tracking using millimeter wave radios," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, 2015, pp. 117–129.

[30] M. Gdeisat and F. Lilley, "One-dimensional phase unwrapping problem," *Signal*, vol. 4, p. 6, Jan. 2011.

[31] K. Itoh, "Analysis of the phase unwrapping algorithm," *Appl. Opt.*, vol. 21, no. 14, p. 2470, 1982.

[32] M. Gdeisat, "Two-dimensional phase unwrapping problem," in *Proc. LJMU*, Jun. 2019, pp. 1–32.

[33] J. M. Huntley, "Noise-immune phase unwrapping algorithm," *Appl. Opt.*, vol. 28, no. 16, pp. 3268–3270, 1989.

[34] J. Huntley and J. Buckland, "Characterization of sources of 2π phase discontinuity in speckle interferograms," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 12, no. 9, pp. 1990–1996, 1995.

[35] Q. Kemao, "Two-dimensional windowed Fourier transform for fringe pattern analysis: Principles, applications and implementations," *Opt. Lasers Eng.*, vol. 45, no. 2, pp. 304–317, Feb. 2007.

[36] G. Liu, R. Wang, Y. Deng, R. Chen, Y. Shao, and Z. Yuan, "A new quality map for 2-D phase unwrapping based on gray level co-occurrence matrix," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 444–448, Feb. 2014.

[37] W. Mao, M. Wang, and L. Qiu, "AIM: Acoustic imaging on a mobile," in *Proc. 16th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2018, pp. 468–481.

[38] G. Gu and G. Peng, "The survey of GSM wireless communication system," in *Proc. Int. Conf. Comput. Inf. Appl.*, Dec. 2010, pp. 121–124.

[39] S. W. Golomb and R. A. Scholtz, "Generalized Barker sequences," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 533–537, Oct. 1965.

[40] B. M. Popovic, "Generalized chirp-like polyphase sequences with optimum correlation properties," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1406–1409, Jul. 1992.

[41] J. G. Andrews, "A primer on zadoff Chu sequences," 2022, *arXiv:2211.05702*.

[42] A. Wang and S. Gollakota, "MilliSonic: Pushing the limits of acoustic motion tracking," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–11.

[43] W. Mao, J. He, and L. Qiu, "CAT: High-precision acoustic motion tracking," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 69–81.

[44] J. B. Mead and A. L. Pazmany, "Quadratic phase coding for high duty cycle radar operation," *J. Atmos. Ocean. Technol.*, vol. 36, no. 6, pp. 957–969, Jun. 2019.

[45] U. Mengali, *Synchronization Techniques for Digital Receivers*. Berlin, Germany: Springer, 2013.

[46] S. S. I. On, *Digital Communications*. New York, NY, USA: Van Nostrand Reinhold, 1985.

[47] K. M. DeGoede, J. A. Ashton-Miller, J. M. Liao, and N. B. Alexander, "How quickly can healthy adults move their hands to intercept an approaching object? Age and gender effects," *J. Gerontol. Ser. A, Biol. Sci. Med. Sci.*, vol. 56, no. 9, pp. 584–588, Sep. 2001.

[48] N. S. Senobari et al., "Super-efficient cross-correlation (SEC-C): A fast matched filtering code suitable for desktop computers," *Seismological Res. Lett.*, vol. 90, no. 1, pp. 322–334, Jan. 2019.

[49] C. J. Chamberlain et al., "EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python," *Seismological Res. Lett.*, vol. 89, no. 1, pp. 173–181, Jan. 2018.

[50] G. Helffrich, "Extended-time multitaper frequency domain cross-correlation receiver-function estimation," *Bull. Seismological Soc. Amer.*, vol. 96, no. 1, pp. 344–347, Feb. 2006.

[51] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 3, May 1998, pp. 1381–1384.

[52] M.-H. Ka and A. I. Baskakov, "Selection of pulse repetition frequency in high-precision oceanographic radar altimeters," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 345–348, Jul. 2007.

[53] M. Vespe, G. Jones, and C. J. Baker, "Lessons for radar," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 65–75, Jan. 2009.

[54] M. H. AlSharif, M. Saad, M. Siala, T. Ballal, H. Boujemaa, and T. Y. Al-Naffouri, "Zadoff-Chu coded ultrasonic signal for accurate range estimation," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 1250–1254.

[55] H. J. Stuart, "The exponentially weighted moving average," *J. Quality Technol.*, vol. 18, no. 4, pp. 203–210, 1986.

[56] S. Guiasu and A. Shenitzer, "The principle of maximum entropy," *Math. Intell.*, vol. 7, no. 1, pp. 42–48, 1985.

[57] J. R. Wait, "Distortion of pulsed signals when the group delay is a nonlinear function of frequency," *Proc. IEEE*, vol. 58, no. 8, pp. 1292–1294, 1970.

[58] Intel. (2022). *Intel® RealsenseT Depth Camera D435i*. [Online]. Available: https://store.intelrealsense.com/buy-intel-realsense-depth-camera-d435i.html

[59] A. Grunnet-Jepsen, J. N. Sweetser, and J. Woodfill, "Best-known-methods for tuning Intel® RealSense™ D400 depth cameras for best performance," Intel Corp., Satan Clara, CA, USA, Tech. Rep., 2018. [Online]. Available: https://scholar.google.com.hk/scholar?hl=en&as_sdt=0%2C5&q=Best-known-methods+for+tuning+Intel%C2%AE+RealSense TM+D400+depth+cameras+for+best+1270+performance%2C&btnG=

[60] (Feb. 2023). *Real-Sense Hand Tracking Tutorial*. [Online]. Available: https://www.intel.com/content/dam/develop/external/us/en/documents/hand-tracking-843462.pdf

[61] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 15–29, doi: 10.1145/2742647.2742662.

[62] C. Zhang et al., "SoundTrak: Continuous 3D tracking of a finger using active acoustics," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 1–25, Jun. 2017, doi: 10.1145/3090095.

[63] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in *Proc. 10th USENIX Symp. Networked Syst. Design Implement.*, 2013, pp. 71–84.

[64] J. Wang, D. Vasisht, and D. Katabi, "RF-IDraw: Virtual touch screen in the air using RF signals," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 235–246, 2014.

[65] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "WiDraw: Enabling hands-free drawing in the air on commodity WiFi devices," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 77–89.

[66] J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.

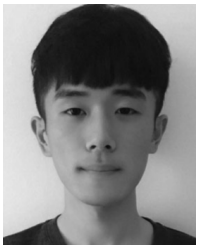
[67] C. Jiang, J. Guo, Y. He, M. Jin, S. Li, and Y. Liu, "mmVib: Micrometer-level vibration measurement with mmWave radar," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.* New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 1–13, doi: 10.1145/3372224.3419202.

[68] S. M. Pandit, N. Jordache, and G. A. Joshi, "Data-dependent systems methodology for noise-insensitive phase unwrapping in laser interferometric surface characterization," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 10, pp. 2584–2592, 1994.

[69] C. R. Mercer and G. Beheim, "Fiber-optic projected-fringe digital interferometry," in *Proc. Fall Conf. Hologram Interferometry Speckle Metrology*, Jan. 1990, pp. 1–15, Paper 103252.

[70] L. Han et al., "Vibration detection and motion compensation for multi-frequency phase-shifting-based 3D sensors," *Sensors*, vol. 19, no. 6, p. 1368, Mar. 2019.
[71] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar*, vol. 11. New York, NY, USA: Wiley, 1991.
[72] P. A. Rosen et al., "Synthetic aperture radar interferometry," *Proc. IEEE*, vol. 88, no. 3, pp. 333–382, Mar. 2000.
[73] P. C. Lauterbur, "Image formation by induced local interactions: Examples employing nuclear magnetic resonance," *Nature*, vol. 242, no. 5394, pp. 190–191, Mar. 1973.
[74] B. Quesson, J. A. de Zwart, and C. T. W. Moonen, "Magnetic resonance temperature imaging for guidance of thermotherapy," *J. Magn. Reson. Imag.*, vol. 12, no. 4, pp. 525–533, 2000.
[75] A. Rauscher, M. Barth, J. R. Reichenbach, R. Stollberger, and E. Moser, "Automated unwrapping of MR phase images applied to BOLD MR-venography at 3 Tesla," *J. Magn. Reson. Imag., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 18, no. 2, pp. 175–180, Aug. 2003.

Yongzhao Zhang (Member, IEEE) received the B.S. degree from the University of Electronic Science and Technology of China, China, in 2018, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2023. He is currently an Assistant Professor with the School of Computer Science and Engineering, University of Electronic Science and Technology of China. His research interests include wireless networks and security, mobile sensing, and underwater Internet of Things (IoT).



Hao Pan (Member, IEEE) received the bachelor's degree from the Yingcai Honors College, University of Electronic Science and Technology of China (UESTC), in 2016, and the Ph.D. degree in computer science from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2022. His research interests include networked systems and span the areas of wireless communication and sensing, human–computer interaction, and computer vision.

Dian Ding (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China, in 2023. He is currently a Post-Doctoral Researcher with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. He has published 15 papers in international conferences and journals, including ACM MobiCom, UIST, UbiComp, SenSys, IEEE INFOCOM, and IEEE TRANSACTIONS ON MOBILE COMPUTING. His research interests include mobile computing, human–computer interaction, and blockchain. His work has been recognized with nominations for the WCAI 2021 Young Outstanding Paper Award and the ACM UbiComp 2024 MIMSVAI Workshop Best Paper Award.

Yue Pan (Student Member, IEEE) received the B.S. degree from Beijing University of Posts and Telecommunications in 2023. He is currently pursuing the master's degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include sensor networks and deep learning.

Yi-Chao Chen (Member, IEEE) received the B.S. and M.S. degrees from the Department of Computer Science and Information Engineering, National Taiwan University, in 2004 and 2006, respectively, and the Ph.D. degree in computer science from The University of Texas at Austin in 2015. In 2018, he joined the Department of Computer Science and Engineering, Shanghai Jiao Tong University, as a tenure-track Assistant Professor. Prior to joining SJTU, he spent a year as a Researcher at the Huawei Future Network Theory Laboratory, Hong Kong, and then worked as the Co-Founder at Hauoli LLC. His research interests focus on networked systems and span the areas of wireless networking, network measurement and analytics, and mobile computing.

Lili Qiu (Fellow, IEEE) received the Ph.D. degree in computer science from Cornell University, Ithaca, NY, USA, in 2001. She is currently a Professor with the Department of Computer Sciences, The University of Texas (UT) at Austin, Austin, TX, USA. Her research interests include internet and wireless networking, with special focuses on wireless network performance and management. She was a recipient of the U.S. NSF Career Award in 2006 and the Google Faculty Research Award. She is an ACM Distinguished Scientist.

Guangtao Xue (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China, in 2004. He is currently a Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include vehicular ad hoc networks, wireless networks, mobile computing, and distributed computing. He is a member of the IEEE Communication Society.

Ting Chen (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2013. He is currently a Professor with the School of Computer Science and Engineering, UESTC. He has authored or co-authored tens of high quality papers in prestigious conferences and journals. His research interests include block-chain, smart contract, and software security. His work was a recipient of several best paper awards, including INFOCOM 2018 Best Paper Award.

Xiaosong Zhang (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), in 2011. He is currently a Professor with the School of Computer Science and Engineering, UESTC. He is also the Director with the Cybersecurity Institute, UESTC. He has authored or co-authored more than 100 papers about cybersecurity, block-chain, and AI security. His research interests include network security, AI security, and blockchain security. He was a recipient of the First Prize of the National Science and Technology Progress Award in 2019 and the Second Prize of the National Science and Technology Progress Award in 2021.