

# Combining IMU with Acoustics for Head Motion Tracking Leveraging Wireless Earphone

Jingyang Hu, Hongbo Jiang, *Senior Member, IEEE*, Daibo Liu, *Member, IEEE*, Zhu Xiao, *Senior Member, IEEE*, Qibo Zhang, Jiangchuan Liu, *Fellow, IEEE*, Schahram Dustdar, *Fellow, IEEE*

**Abstract**—Head motion tracking is a promising research field with vast applications in ubiquitous human-computer interaction (HCI) scenarios. Unfortunately, solutions based on vision and wireless sensing have shortcomings in user privacy and tracking range, respectively. To address these issues, we propose IA-Track, a novel head motion tracking system that combines inertial measurement units (IMU) and acoustic sensing. Our wireless earphone-based method balances flexibility, computational complexity, and tracking accuracy, requiring only an earphone with an IMU and a smartphone. However, we still face two challenges. First, wireless headsets have limited hardware resources, making acoustic Doppler effect-based method unsuitable for acoustic tracking. Second, traditional Kalman filter-based trajectory restoration methods may introduce significant cumulative errors. To tackle these challenges, we rely on IMU sensor data to recover the trajectory and use smartphones to emit “inaudible” acoustic signals that the earphone receives to adjust the IMU drift track. We conducted extensive experiments involving 50 volunteers in various potential IA-Track usage scenarios, demonstrating that our well-designed system achieves satisfactory head motion tracking performance.

**Index Terms**—Head motion tracking, Acoustic Signal, Human-machine interface.

## 1 INTRODUCTION

HEAD motion tracking has become a research field of increasing interest in recent years. Head tracking offers a more natural, direct, and efficient way of human-computer interaction, thus enhancing the user experience. Through the identification of the user’s head position and orientation, head tracking technology accurately combines AR/VR content with the user’s surrounding environment. Equipping smart glasses or earphones with head tracking technology, as exemplified by Google [1] and Apple [2], allows users to conveniently control smart devices by merely moving their heads. Given the prospect of ubiquitous human-computer interaction functions in future intelligent environments, head motion tracking technology has become a prerequisite for optimal human-computer interaction.

Recent work has made significant progress in achieving head motion tracking in vision-based systems [3]–[7] and wireless sensing-based systems [8]–[14]. Vision-based head tracking methods mainly use special equipment (VR glasses, etc.) [1], [5], [15] or use cameras [4], [6], [16]–[18] to use the gaze direction of the eyes and face orientation for head tracking. Alternatively, analyze the face model and the relative positions of facial organs (mouth corners, eye corners, nose, and pupils) [19], [20]. These methods require the camera to capture the user’s face at all times, which brings privacy issues [18], and the user’s face needs to be kept within the camera’s field of view at all times.

Wireless sensing-based systems have recently received

widespread attention because they do not require additional equipment and do not violate privacy. Sensing methods [21], [22] based on Wi-Fi CSI (Channel State Information) have attracted extensive attention due to the large-scale integrated deployment of Wi-Fi devices. However, Wi-Fi-based methods can only provide coarse-grained measurements due to their narrow Wi-Fi bandwidth. Acoustic-based methods [23], [24] require an array of acoustic emitters (an infrastructure that is not always available) to achieve cm-level tracking accuracy, which may limit their applicability to smart environments. In addition, limited by the signal strength, the tracking accuracy drops significantly [12], [25] when the distance between the speaker and the microphone exceeds 1m. As a result, existing methods face challenges in balancing computational complexity, user experience, and accuracy.

In this paper, we propose IA-Track, which extends current applications of wireless earphones in head tracking. Consider that most wireless earphones are interconnected with smartphones or laptops. The specific workflow of IA-Track is as follows. We use an IMU sensor configured on a wireless earphone to acquire accelerometer data and then use a Kalman filter to recover the trajectory. However, the traditional Kalman filter will bring severe accumulation and affect trajectory recovery. We use a stationary smartphone as an acoustic reference point and continuously emit an “inaudible” acoustic signal (16kHz) to correct IMU-driven motion tracking. As the capabilities of wireless headsets continue to evolve, our proposed wireless headset-based head-tracking solution has great potential to enable exciting new applications.

The idea of using wireless earphones for head motion tracking sounds simple, but two serious challenges remain:

- Due to limited hardware resources, the frequency

J. Hu, H. Jiang, D. Liu, Z. Xiao and Q. Zhang are with College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (emails: fbhijy@hnu.edu.cn, hongbojiang2004@gmail.com, dbliu.sky@gmail.com, zhuxiao@hnu.edu.cn, zhangqibo@hnu.edu.cn).

J. Liu is with School of Computing Science, Simon Fraser University, Burnaby, BC V5A1S6, Canada, and also with Jiangxing Intelligence Inc. (email: jcliu@cs.sfu.ca).

S. Dustdar is with the Distributed Systems Group, TU Wien, 1040 Vienna Austria. (email: dustdar@dsg.tuwien.ac.at).

offset generated by the built-in oscillator of wireless earphones is much larger than that of professional acoustic equipment. Secondly, the frequency range of acoustic signals supported is narrower, which makes it unusable in traditional Doppler effect-based ranging methods.

- Severe cumulative error caused by Kalman filtering. Traditional IMU-based schemes usually use Kalman filtering for trajectory recovery. But with the continuous update of the state of motion. The traditional Kalman filter will bring serious cumulative error so that the head movement trajectory is difficult to recover.

To address the first challenge, we abandon the traditional motion tracking scheme based on Doppler effect signals and use constant frequency signals for motion tracking. We found that when the relative position of the wireless earphone and the smartphone did not change, the sinusoidal signal received had the same duration as the signal sent by the wireless earphone. Moreover, when the wireless earphone is far from the smartphone, the signal received by the smartphone is stretched relative to the signal sent by the wireless earphone. Conversely, when the wireless earphone is close to the smartphone, the signal received by the smartphone will be compressed relative to the signal sent by the wireless earphone. Based on this finding, we used a ranging scheme based on the peaks and valleys of sinusoidal signals for head tracking used by IA-Track. We use a sliding window to detect how much the signal is compressed and stretched. We calculate the phase difference by calculating the difference between the peaks and troughs in the transmitted signal and the received signal in a sliding window and convert this phase difference to distance. Then we perform interpolation and filtering on the final signal to improve the ranging accuracy. To address the second challenge, IA-Track uses an acoustic-ranging approach to opportunistic calibration of the IMU's tracking. We perform trajectory recovery based on traditional Kalman filtering and continuously revise the filtering results by measuring the distance between the wireless earphone and the smartphone. The acoustic signal "pulls" the wireless earphone like a string attached to a kite as shown in Fig. 2.

In summary, we make the following major contributions in designing IA-Track:

- We use a pair of wireless earphones and smartphones to achieve high-accuracy head tracking. Through theoretical and experimental analysis, the IMU data at the ear is an important part of restoring the head trajectory. The design of IA-Track can effectively expand the application of wireless earphones in VR/AR.
- We address the challenge that wireless headsets cannot use conventional Doppler effect ranging due to hardware limitations. We adopt a novel signal peak/valley based constant frequency ranging scheme for acoustic ranging.
- We optimize traditional Kalman filter-based trajectory recovery methods. We propose an acoustic-based opportunity calibration method to reduce accumulated errors. Our carefully designed system

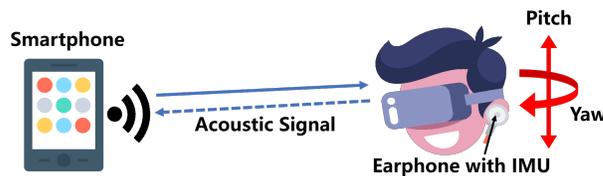


Fig. 1: Head motion tracking is performed using an IMU configured on wireless earphones, and the acoustic signal emitted by the smartphone is used to calibrate the results of the Kalman filter.

strikes a balance between computational complexity, accuracy, and equipment cost.

The rest of this paper is organized as follows. Section 2 presents the feasibility analysis of applying wireless earphone for head track. Section 3 introduces the overview of the system. Section 4 present the whole system. Section 5 present the implementation of IA-Track. Section 6 presents the system evaluation. Section 7 introduces the related work. Section 8 discussed the limitation and Section 9 conclusion of our work.

## 2 BACKGROUND

### 2.1 Earphone and IMU basics

The most crucial function of a wireless earphone is to convert electrical signals to acoustic signals. The speakers in the earphone convert electrical signals into acoustic signals. Currently, leading wireless earphones such as AirPods Pro have multiple built-in microphones to achieve noise reduction and recording. Advanced wireless Earphones integrate various sensors, such as IMU and optical sensors, for human-computer interaction. Considering that not all wireless earphones are integrated with IMU sensors, we also configured a miniature IMU sensor for the IA-Track. The IMU sensor [26] we use has a sampling rate of 400Hz and can transmit data to the smartphone with a delay of 15ms. The size of the chip is 30x30x5mm, and the weight is only 10g. We attached this IMU sensor to the Wireless earphone and Wireless headphone as shown in Fig. 2(a) and Fig. 2(b).



(a) Wireless earphone taped with a wireless IMU. (b) Wireless headphone taped with a wireless IMU.

Fig. 2: Hardware prototype used by IA-Track.

### 2.2 Doppler effect based tracking

Previous works [27], [28] use Doppler effect for tracking. When the Doppler shift occurs, the moving speed of the receiver relative to the transmitter is:

$$v = \frac{\Delta f}{f} c \quad (1)$$

For a  $f = 16\text{kHz}$  constant frequency signal. Based on the Eq. (1), the ranging error caused by a frequency offset of 1Hz is 21.3mm at 1s. Such range resolution is not sufficient for fine-grained head motion tracking. In addition, the time delay of ranging using Doppler frequency shift is also a problem.

The central process in Doppler Effect-based tracking involves the computation of the frequency shift  $\Delta f$ . This is achieved by subjecting the received acoustic signal to frequency analysis, such as the Short-Time Fourier Transform (STFT), which yields the spectral distribution of the received signal. In practical applications, STFT is performed on a sliding window. The resolution of the frequency shift  $\Delta f$  can be expressed as:

$$\hat{\Delta f} = \frac{F_s}{L_w} \quad (2)$$

where  $F_s$  represent the sampling rate and  $L_w$  is the length of sliding window. Combing Eq. (1) and Eq. (2), we dervie resolution the moving speed  $v$  as:

$$\hat{v} = \frac{\hat{\Delta f}}{f} c = \frac{F_s}{L_w f} c \quad (3)$$

From Eq. (3), We can see that the resolution of  $v$  is only related to the sliding window length  $L_w$ . A larger window provides better frequency domain resolution. However, a larger window contains more samples and results in a larger delay. Considering that the commonly used sampling rate of devices is  $F_s = 48\text{kHz}$ . Then, for a sliding window with 2048 samples, the delay of processing 1s data is  $2048/48000 = 43\text{ms}$ .

### 2.3 Frequency Offset in wireless earphones

Wireless earphones are smaller than smartphones and use a castrated hardware version. So the oscillator in the wireless earphone is not as good as in the smartphone, so the frequency offset between the transmitted signal and the actual signal will also be greater than that in the smartphone. This frequency shift has no noticeable effect on calls and music playback, but the effect is disastrous for fine-grained acoustic ranging and tracking. For traditional Doppler effect-based ranging, based on the Eq. (3), the ranging error caused by a frequency offset of 1Hz is 21.3mm at 1s. This error is intolerable for fine-grained head motion tracking.

We set comparative experiments to compare the frequency offset of smartphones and wireless earphones. We use five wireless earphones (Bose quietComfort2, Beats Fit Pro, Huawei FreeBuds 2, Samsung Buds, and Apple AirPods) and one smartphone (Huawei P40) to detect a constant 16kHz frequency offset. The results are shown in Fig. 3. We found that the frequency offset of Bose quietComfort2 and Beats Fit Pro is 0.1Hz-0.2Hz. As a comparison, the frequency offset of Huawei FreeBuds 2, Samsung Buds, and Apple AirPods are over 0.8Hz (we think it may be that the high-frequency response of these three earphones is relatively weak).

### 2.4 IMU-BASED TRACKING

We attach the IMU to the wireless earphone as shown in Fig. 4(a). The movements of the human ear can well reflect

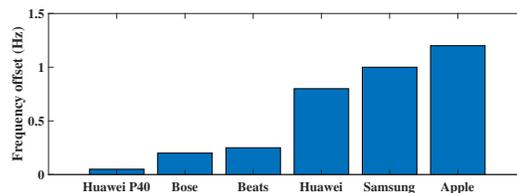


Fig. 3: Frequency offset of different devices.

the trajectory of the human head. Fig. 4(b) and Fig. 4(c), we show the three-axis acceleration data corresponding to the two head movements of yawning and nodding. Different head movements will produce different data patterns. Usually, we use Kalman filtering to process these IMU data to restore the motion trajectory. However, the traditional Kalman filter will produce serious cumulative errors that seriously affect the accuracy of the final trajectory restoration. We uses an opportunity calibration algorithm based on acoustic measurements to optimize the traditional Kalman filter results in Section 4.3.2.

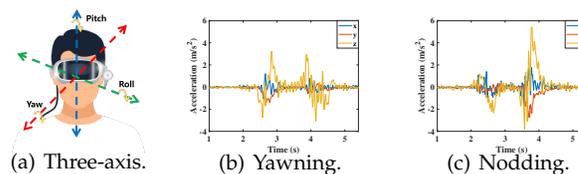


Fig. 4: The head motion acceleration data obtained by the IMU.

## 3 SYSTEM OVERVIEW

Fig. 5 shows the overall architecture of IA-Track, mainly composed of three modules. (1) Acoustic and IMU signal sensing. (2) Acoustic-based distance ranging. (3) IMU-based trajectory recovery. These three modules make up the IA-Track system.

**Acoustic and IMU signal sensing:** This module mainly connects the smartphone with wireless earphones through the sound signal. The module first accepts IMU sensor data as input. The input data comes from the IMU's three-axis accelerometer data. The smartphone then continuously transmits an inaudible high-frequency signal to the wireless earphone.

**Acoustic based distance ranging:** To accomplish this goal, we need to solve two problems. First is the preprocessing of acoustic signals. The second problem is how to measure the distance. In the case of a significant frequency deviation of the wireless headset, we abandoned the traditional Doppler effect-based distance measurement solution and used a peak/Valley-based Distance Measurement.

**IMU-based trajectory recovery:** We use a traditional Kalman filter as a basis. Then use the continuously measured distance from the previous module as an opportunity calibration. Finally, we can restore the motion trajectory of the wireless earphone.

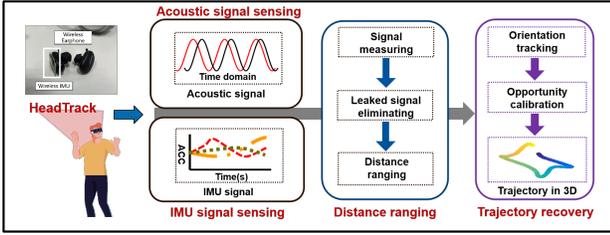


Fig. 5: System Architecture of IA-Track.

## 4 SYSTEM DESIGN

### 4.1 Acoustic Signal Preprocessing

Due to its hardware limitations, the wireless earphone will produce more significant frequency leakage in the sensing of acoustic signals, and we must first solve this problem. We define that the acoustic signal  $S_r(t)$  is composed of the airborne signal  $S_a(t)$  and the leakage signal  $S_l(t)$ . Since  $S_a(t)$  and  $S_l(t)$  share the same original signal, they have the same frequency but different phases and amplitude. Considering that only  $S_a(t)$  contains the tracking information of the target, the phase and amplitude decays are constant under a fixed frequency leakage. We can estimate frequency leakage from the measured phase and amplitude decay.

We "connect" the wireless earphone to the smartphone by emitting an acoustic signal at a fixed frequency of 16kHz as shown in Fig. 6(a). We first define the signal  $V = A \sin \phi$ ,  $A$  represents the transmitted signal amplitude,  $\phi$  represents the phase. Since we cannot know the initial values of  $\phi$  and  $A$  for running smartphones and wireless earphones, we obtain the bias of the received signal  $90^\circ$  by Hilbert transform Move  $V' = A \cos \phi$ . Through this step, we get  $A$ , which can be expressed as:

$$A = \sqrt{(A \sin \phi)^2 + (A \cos \phi)^2} \quad (4)$$

and  $\phi$  calculated as:

$$\phi = \arctan \left( \frac{A \sin \phi}{A \cos \phi} \right) \quad (5)$$

When the transmitted signal is a single-frequency sinusoidal signal, the leaked signal can be expressed as:

$$S_l(t) = A' \sin (2\pi ft + \phi') \quad (6)$$

where  $\phi'$  and  $A'$  are the phase shift and amplitude of the received signal, respectively, and the frequency  $f$  is the same as the transmitted signal. In order to measure the frequency leakage at one time, we kept the wireless earphone and the smartphone at a fixed distance in the initial laboratory, and at the same time used sound insulation material to attenuate the signal transmission in the air to reduce the strength of the signal in the air to the microphone. In this way, the signal component in the air is small and negligible. We can get  $\phi'$  and  $A'$  by comparing the phase difference between the transmitted signal and the received signal. The remaining signal after we remove the leak is shown in Fig. 6(b).

### 4.2 Phase-based Distance Measurement

In section 2.2, We analyze why acoustic ranging on wireless earphones cannot be performed using the Doppler effect.

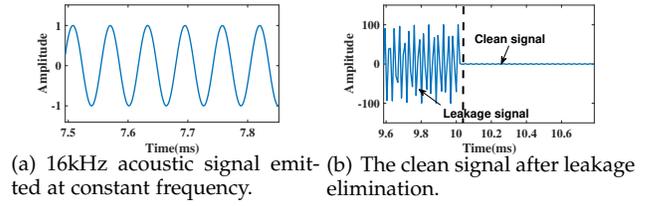


Fig. 6: Acoustic signal preprocessing.

Therefore, for the fine-grained tracking required for head motion tracking, we tend to use changes in the computed signal phase for motion distance detection. Based on [11], the change in distance  $\Delta d$  and the change in phase  $\Delta \phi$  can be expressed as:

$$\Delta d = \frac{\Delta \phi}{2\pi} \lambda. \quad (7)$$

For a constant frequency signal of 16kHz,  $\lambda = 2\text{cm}$ . For a wavelength of phase change ( $2\pi$ ), the phase resolution and range resolution are  $0.1\pi$  and 1mm, respectively. Such phase resolution and range resolution are sufficient for commonly used head motion tracking.

#### 4.2.1 Compression and stretching of acoustic signals

Since the relative movement between the signal transmitting end and the receiving end, the signal will be compressed or stretched at the receiving end relative to the original signal, we next demonstrate how the signal changes when the smartphone and the wireless earphone undergo various relative motions. When the smartphone and the wireless earphone are held relatively still. Smartphones emit sinusoidal signals, and wireless headphones receive 16kHz acoustic signals. At this time, the transmitted signal and the received signal have no compression and stretching and are exactly the same in period. But as shown in Fig. 7(a) when the wireless earphones approach the smartphone. At this time, the signal transmission time is longer than the signal receiving time. Since the received signal is contained in a smaller time window, we find that the received signal is compressed relative to the transmitted signal. Similarly, as shown in Fig. 7(b). When the wireless headset is far away from the smartphone, the signal transmission time is shorter than the signal reception time. Since the transmission time window required to receive the signal becomes larger, the reception signal is relatively shorter than the transmission time window, and the signal will be stretched.

IA-Track uses a 16kHz constant signal of constant frequency, so the phase difference between the received signal and the transmitted signal changes from time to time. We define the phase change  $\Delta \phi$  as the difference between the phase of the transmitted signal and the received signal as:

$$\Delta \phi = \Delta \phi_r - \Delta \phi_t. \quad (8)$$

We get the phase change of the transmitted signal  $\Delta \phi_t$  by multiplying the frequency of the transmitted signal by the time interval. We note that Delta  $\Delta \phi_r$  is the phase  $\Delta \phi_r = \phi_i - \phi_j$  collected at the beginning and end of each time window To get the data, we use this data to estimate the phase change of the received signal ( $\Delta \phi_r$ ). When the smartphone and the wireless earphone move relative to each

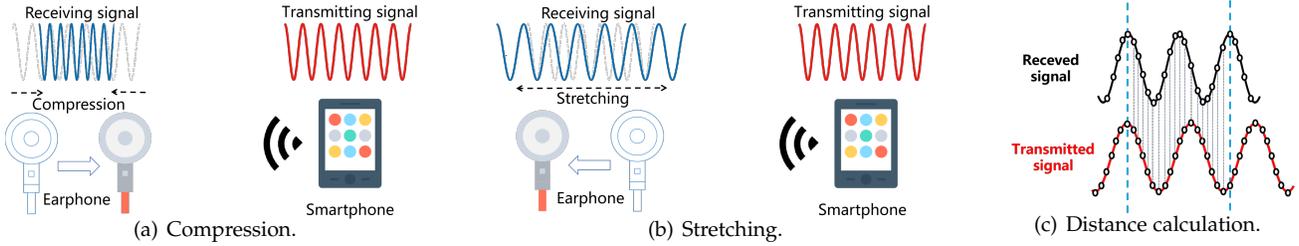


Fig. 7: IA-Track uses signal compression or stretching phenomenon for distance calculation.

other, the path between the transmitter and the receiver will also change, so the signal phase difference  $\Delta\phi_r$  on the path will also change accordingly.

Current acoustic-based tracking systems employ sliding windows to process acoustic signals in timestamps. Smartphones typically have a sampling rate of 48kHz, and we calculate the phase change by sampling points in carefully crafted timestamps. We calculate the phase change  $\Delta\phi_r$  in the starting and ending sample points in the sliding window and count the phase values in each sliding window. We transmit a constant signal at 16kHz. We calculate the phase by measuring the phase value at the sample points in each time window, then calculating the in-phase (I) and quadrature (Q) components of the signal, and then by the arctangent (I/Q). However, this method improves efficiency compared to FFT, but requires the use of Hilbert transform or mixed signal and processing through filters, such computational complexity is still optimized for smartphones and wireless earphones with limited computing power space.

We use another method based on the difference between peaks and valleys to calculate the phase difference between the transmitted signal and the received signal. For a sinusoidal signal of constant frequency, the phase change between two adjacent local extreme points (i.e., peaks and valleys) is  $\pi$ . After setting the size of the sliding window, we use the phase difference between the sample points at the beginning of the window and the end of the window to calculate. Further, we use the dynamic time window scheme. Compared with the general sliding window, we calculate the difference between the local extreme points (peak and trough) in the window instead of the difference between all sampling points. We assume that each window contains  $L$  local extreme points. Then the phase change of each window signal can be expressed as:

$$\Delta\phi_r = L\pi. \quad (9)$$

Compared with the traditional acoustic ranging scheme based on FFT and phase difference. The method based on the peak-to-valley difference has very low computational complexity and overhead. We don't need to do complicated FFTs, and Hilbert transforms on the signal. We just count the number of differences between the peaks and valleys in the sliding window. Since the smartphone and wireless earphones are in relative motion, the signal will be compressed or stretched. When the peak-to-valley difference between the transmitting signal and the receiving signal becomes smaller, the distance between the transmitting end and the receiving end increases, and when the difference

between the transmitting and receiving signals becomes larger, the transmitting end and the receiving end, The distance between the ends becomes smaller.

#### 4.2.2 Peak/Valley-based Distance Measurement

In this section, we will specifically describe how to use the peaks/valleys of a sinusoidal signal for specific ranging. When the smartphone and the wireless earphone move relative to each other, the received signal will stretch and compress relative to the transmitted signal. We can calculate the phase difference of the signal by accepting the degree of compression or stretching of the signal. This phase difference can be converted into time, and the product of time and the speed of sound is the final signal propagation distance we need.

Commonly used commercial smartphones and wireless earphones use a sampling rate of 48kHz. But at this sampling rate, the 16kHz signal in the sliding window we set has only three samples. This leads to the fact that there may not be the peaks and valleys (extreme points) we need in a sliding window. Therefore, in order to more accurately calculate the phase difference between the received signal and the transmitted signal, we use interpolation and low-pass filtering to upsample the signal once to obtain more sampling points. We use 8 times of upsampling, the phase measurement error is reduced from  $\frac{2\pi}{3}$  to  $\frac{2\pi}{3 \times 8}$ . The ranging error can be expressed as  $\hat{\Delta}d = \frac{\Delta\phi}{2\pi}\lambda = \frac{2\pi}{2\pi \times 3 \times 8}20 = 0.83\text{mm}$ , which is lower than the Doppler ranging error of 21.3mm. We finally calculate the phase change of the received signal  $\Delta\phi_r$  in the window containing  $L$  extreme points.

We assume the sampling rate and upsampling coefficient to be  $F_s$  and  $M$ , respectively. Each sliding window has  $N$  sampling points and contains  $L$  extreme points. We define the initial frequency of the emitted signal as  $f_0$ . Therefore, the phase difference  $\Delta\phi_t$  can be expressed as:

$$\Delta\phi_t = 2\pi f_0 t = 2\pi f_0 \frac{N}{F_s M}. \quad (10)$$

Then the relative displacement distance in each sliding window is:

$$\Delta d = \frac{\Delta\phi_r - \Delta\phi_t}{2\pi} \frac{c}{f_0} = c \left( \frac{L}{2f_0} - \frac{N}{F_s M} \right) \quad (11)$$

where  $c$  is the speed of sound.  $N$  is the number of sample points contained in a window. When the signal is stretched, the value of  $N$  will keep getting larger in successive windows. When the received signal is compressed, the value of  $N$  will keep getting smaller in successive windows.

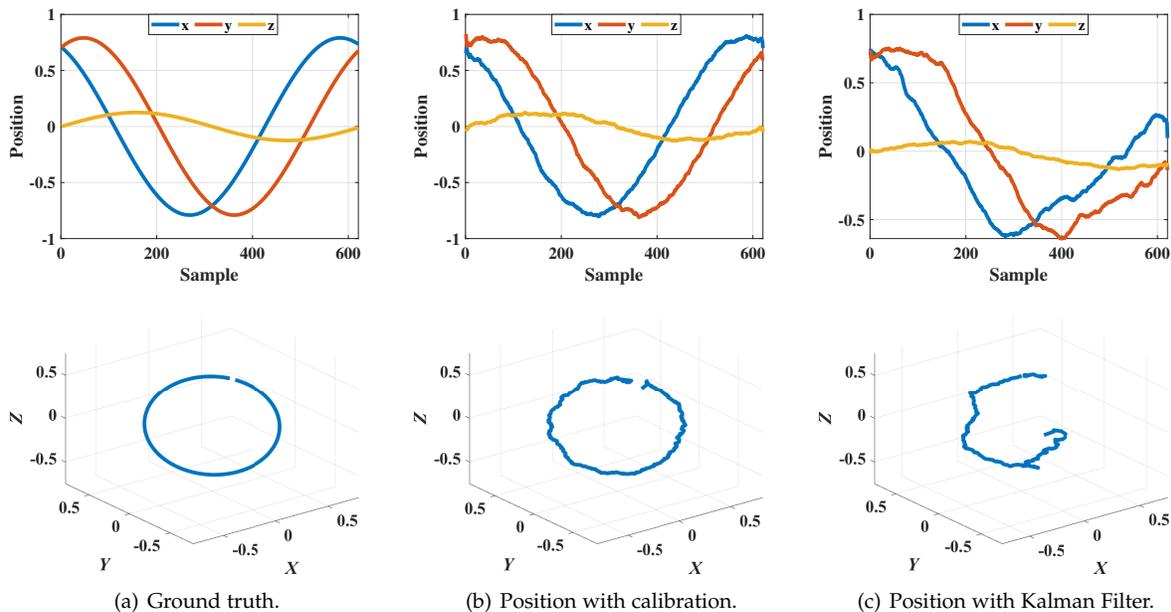


Fig. 8: Three-axis position (first row) and trace (second row) of our opportunistic calibration and Kalman filter.

As shown in Fig. 7(c), The received signal is compressed relative to the transmitted signal. The parameters are set as follows:  $c = 340m/s$ ,  $L = 4$ ,  $f_0 = 16kHz$ ,  $F_S = 48kHz$  and  $M = 4$ . In the window of Fig. 7(c), we observe that there are 14 sampling points ( $N = 19$ ). Based on Eq. (11), We can calculate the distance  $\Delta d = 340(\frac{4}{2 \times 16000} - \frac{19}{48000 \times 4}) = 9.18mm$ . Because the receiving signal is compressed, the earphone moves 9.18mm toward the smartphone.

### 4.3 IMU-based head tracking

Traditional IMU-based trajectory tracking technology is widely used in UAVs and intelligent robots, and Kalman filtering is used for attitude estimation based on accelerometers and gyroscopes. But the traditional Kalman filter cannot be used well on IA-Track. There are two main reasons. One is that Kalman filtering needs to consume a lot of computing resources, which is too large for small smart devices such as smart phones, wireless earphones and smart watches. The second point is that the traditional Kalman filter will produce a large amount of cumulative error, which is acceptable if it is only used for attitude estimation, but it will have a catastrophic effect on noise such as trajectory restoration. Therefore, we propose an acoustic-based opportunistic calibration to optimize the Kalman filtering process. We have optimized the IMU motion tracking scheme based on Kalman filter to meet the head motion tracking based on wireless earphone.

#### 4.3.1 Kalman filter basics

For a traditional motion tracking model, Based on [29], we have the true state  $s_k = [\mathbf{q}_k^T, \mathbf{v}_k^T, \mathbf{p}_k^T]^T$ , Where  $\mathbf{q}_k$  is a quaternion used to represent the direction, which consists of a real part and three imaginary parts.  $\mathbf{v}_k$  and  $\mathbf{p}_k$  represent the three-dimensional vectors of velocity and position, respectively. The relationship between true state  $s_k$  at time  $k$  and previous state  $s_{k-1}$  at time  $k-1$  can be expressed as:

$$s_k = A_k s_{k-1} + \mathbf{w}_k, \quad (12)$$

where  $\mathbf{w}_k$  and  $A_k$  are respectively the noise and state transition matrices. Moreover, at time  $k$ , an observation  $\mathbf{o}_k$  is made and we have:

$$\mathbf{o}_k = H_k s_k + \mathbf{v}_k \quad (13)$$

where  $\mathbf{v}_k$  is the state observation noise and  $H_k$  is the state observation matrix. Then, the following calculations are performed sequentially to estimation  $\hat{S}_k$  on  $\hat{s}_k$  from  $\hat{S}_{k-1}$ :

$$\begin{aligned} \hat{s}_{k|k-1} &= A_k \hat{s}_{k-1}, \mathbf{r}_k = \mathbf{o}_k - H_k \hat{s}_{k|k-1}, \\ P_{k|k-1} &= A_k P_{k-1} A_k^T + W_k, R_k = H_k P_{k|k-1} H_k^T + V_k. \end{aligned} \quad (14)$$

where  $\mathbf{r}_k$  represent the residual value,  $K_k$  is Kalman gain. To derive the estimated minimum mean square error  $\hat{s}_k$ :

$$\begin{aligned} K_k &= P_{k|k-1} H_k^T R_k^{-1}, \hat{s}_k = \hat{s}_{k|k-1} + K_k \mathbf{r}_k, \\ P_k &= (I - K_k H_k) P_{k|k-1}. \end{aligned} \quad (15)$$

However, IMU-based motion tracking may have the problem of accumulating errors, mainly for the following three reasons: i) The acceleration measured by the IMU is not completely accurate. Every measurement can be loaded with noise and drift, and especially over long runs, these errors can accumulate, leading to discrepancies between the final estimated trajectory and the true trajectory. ii) Kalman filters generally assume that the system is linear and that the errors between states and measurements are Gaussian distributed. However, there may be nonlinear effects in the real world, or the error distribution does not conform to a Gaussian distribution, and these factors may cause the cumulative error of the filter. iii) The Kalman filter requires an initial state estimate, and if the initial state is inaccurate, or if the uncertainty in the initial state is not properly modeled, then the estimated trajectory may deviate from the true value.

#### 4.3.2 Tracking through opportunistic calibrations

To solve this accumulated error, we design an acoustic-based opportunistic calibration algorithm based on the

**Algorithm 1: Opportunistic State Calibration**

**Require:**  $\hat{\mathbf{s}}_{k|k-1}, \mathbf{r}_k, t_{\text{prv}}, \varepsilon$   
**Ensure:**  $\hat{\mathbf{s}}_k, t_{\text{prv}}$

- 1:  $t \leftarrow \text{clock time} - t_{\text{prv}}$
- 2:  $\hat{\mathbf{s}}_k \leftarrow \hat{\mathbf{s}}_{k|k-1}$
- 3: **if**  $O_k < \varepsilon$  **then**
- 4:    $\hat{\mathbf{v}}_k \leftarrow \hat{\mathbf{v}}_{k|k-1} - \frac{\hat{\mathbf{p}}_{k|k-1}}{t}$
- 5:    $\hat{\mathbf{p}}_k \leftarrow 0$
- 6: **else**
- 7:    $j \leftarrow \arg \min_i \mathbf{r}_{k_i}$
- 8:   **if**  $\mathbf{r}_k^j < \varepsilon$  and  $j \in \{1, 2, 3\}$  **then**
- 9:      $\hat{\mathbf{v}}_k \leftarrow \hat{\mathbf{v}}_{k|k-1} - \frac{\hat{\mathbf{p}}_{k|k-1}^{-j}}{t}$
- 10:      $\hat{\mathbf{p}}_k \leftarrow \hat{\mathbf{p}}_{j,k|k-1}^{-j}$
- 11:   **else if**  $\mathbf{r}_k^j < \varepsilon$  and  $j \in \{4, 5, 6\}$  **then**
- 12:      $\hat{\mathbf{v}}_k \leftarrow \hat{\mathbf{v}}_{k|k-1} - \frac{\hat{\mathbf{p}}_{k|k-1}^{(j-3)}}{t}$
- 13:      $\hat{\mathbf{p}}_k \leftarrow \hat{\mathbf{p}}_{k|k-1}^{-(j-3)}$
- 14:   **end if**
- 15: **end if**
- 16:  $t_{\text{prv}} \leftarrow \text{clock time}$

configuration of wireless earphones and smartphones to correct the deviated trajectory back to the correct direction. Specifically, We establish a global coordinate system  $(x,y,z)$  with the smartphone as the origin. Basically, we redefine the observation value  $o_k$  as the radial distance through the cumulative radial displacement obtained by acoustic ranging, and correct the calculation of the residual  $\mathbf{r}_k$  as:

$$\mathbf{r}_k = \begin{bmatrix} r_k^x \\ r_k^y \\ r_k^z \\ r_k^{yz} \\ r_k^{xz} \\ r_k^{xy} \\ r_k^{xyz} \end{bmatrix} = \begin{bmatrix} |o_k - \|p_k^x\|| \\ |o_k - \|p_k^y\|| \\ |o_k - \|p_k^z\|| \\ o_k - \|[p_k^y, p_k^z]\|_2 \\ o_k - \|[p_k^x, p_k^z]\|_2 \\ o_k - \|[p_k^x, p_k^y]\|_2 \\ o_k - \|\mathbf{p}_k\|_2 \end{bmatrix} \quad (16)$$

Whenever the IMU sensor updates data, we use the steps in Eq. (14) and Eq. (15) to obtain  $\hat{\mathbf{s}}_{k|k-1}$ , and after obtaining a new observation  $o_k$ , we use Eq. (16) to calculate the residual  $\mathbf{r}_k$ . Then use Algorithm 1 (we use  $[p_k^1, p_k^2, p_k^3] = [p_k^x, p_k^y, p_k^z]$  to simplify the expression) to correct  $\hat{\mathbf{s}}_{k|k-1}$ . As shown in Fig. 8, taking the use of wireless earphones to do circular motion in three-dimensional space as an example, only using Kalman filtering will produce serious cumulative drift, as shown in Fig. 8(c). Fortunately, applying our chance calibration can largely recover the motion trajectories shown in Fig. 8(b).

**4.3.3 Head trajectory reduction**

We have previously optimized the traditional Kalman filtering process using an acoustic-based opportunity calibration strategy. Here's a complete review of the IA-Track workflow. In IA-Track the smartphone continuously transmits acoustic signals and returns from the wireless earphone. The data obtained by the wireless earphone and the IMU is returned to the smartphone via Bluetooth.

The main task of IA-Track is to restore the head movement trajectory, so the core of the system is to optimize

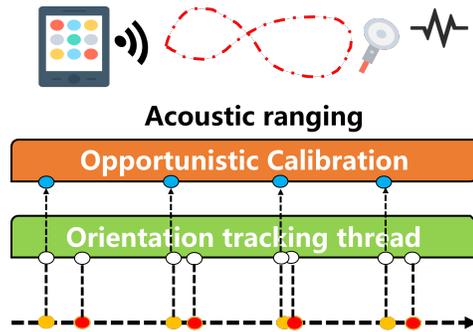


Fig. 9: Multi-threaded execution pipeline.

the trajectory recovery based on Kalman filtering through opportunity calibration. First, we acoustically track the observations needed to obtain the opportunity calibration. Then, the IMU continuously obtains the acceleration data, and IA-Track uses the smartphone as the origin of the three-dimensional coordinate system to convert the acceleration data in the IMU into the motion acceleration data of the head in the three-dimensional coordinate system. Then we continuously track the head movement direction and acceleration through the Kalman filter optimized by opportunity calibration. Since the smartphone is the origin of the entire three-dimensional coordinate system, we need to calibrate the system every time we use IA-Track. We only need to keep the wireless earphone and the smartphone relatively still for a period of time to complete the calibration. The whole process lasts about 5s.

While smartphones and wireless earphones are getting more powerful, it's not easy to assemble the required components in the IA-Track in the device. Fig. 9 shows how IA-Track coordinates multiple threads to complete the complex tracking process. There are two separate threads in the Android system that perform acoustic and IMU data recording. The smartphone first calculates the current distance from the wireless earphone. Then, the opportunistic calibration module retrieves the IMU data obtained at the previous and current time points. And we match it with the observation point, filter the acceleration value, and restore the current trajectory. The direction-tracking thread runs in parallel based on all IMU sensor readings and its output is used to adjust the acceleration.

**5 PROTOTYPE IMPLEMENTATION**

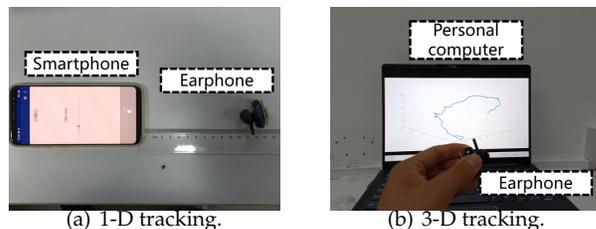


Fig. 10: Experiment setup.

We will introduce the prototype Implementation of IA-Track in this section.

**Experiment Hardware.** We have implemented head tracking on both Android smartphones and PC. On the Android platform, we developed an app that emits a sinusoidal signal at 16 kHz via connected wireless earphones and receives the signal at a sampling rate of 48 kHz via a built-in microphone. The application performs signal processing and displays the motion trajectory on the screen in real time. We used Huawei P40 and Samsung Galaxy S8 to track wireless earphones on this platform. We have selected 6 wireless earphones with different price configurations: Apple AirPods, Bose, Beats, Honor, Audio-Technica, and Samsung Galaxy Buds+. In addition, considering the popularity of personal PC in daily office and indoor environments. We set the IMU sensing rate to the highest level, so the sampling frequency is roughly 400Hz for all IMU sensors, though their samples may still arrive irregularly.

**Experiment Setup.** We recruited 50 participants (28 males and 22 females) for IA-Track, each participating in three experiments. We first conducted two benchmark Studies to test the IA-Track's acoustic ranging module and the tracking function in the handheld earphones state as shown in Fig. 10. Specifically, we set 15 locations from 10cm to 150cm with a step size of 10cm to detect the ranging error of IA-Track at different locations. Then, We verified IA-Track's ability to recover trajectories by asking participants to hold an earphone and perform motion trajectories, which were displayed on the screen and consisted of lines, circles, squares, and triangles at different scales. It is worth noting that when performing these gestures, they do not need to return to the exact starting position deliberately but only follow the gesture tracking. Finally, we asked the participants to wear earphones to perform three kinds of head movements (1) The head moves back and forth for 5 seconds; (2) The head moves left and right for 5 seconds; (3) The head moves randomly for 5 seconds. Before starting each experiment, we calibrated the IA-Track system to ensure accurate tracking results. Participants can wear the earphones according to their preferences, as our algorithm is not sensitive to their wearing habits. As a token of appreciation, each participant received a 50-dollar shopping card after completing the experiment.

**Ground truth.** For performance testing, we use VR headsets and Kinect-acquired head positions as ground truth. We then compare the ground truth with the measurements of the IA-Track system to complete the evaluation. We use wireless headphones to transmit data while running IA-Track, and then complete the data analysis on the smartphone. Due to the small amount of collected acoustic data and IMU data, it is easy to complete trajectory recovery on a smartphone.

## 6 EVALUATION

In this section, we first conducted two benchmark Studies to evaluate the acoustic ranging and 3-D tracking performance of IA-Track and then evaluate the overall head-tracking capabilities of the system. Finally, we discuss the application of IA-Track in practical scenarios.

### 6.1 Benchmark studies

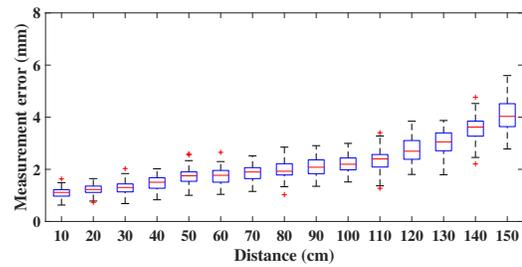


Fig. 11: Ranging accuracy interconnected with smartphone.

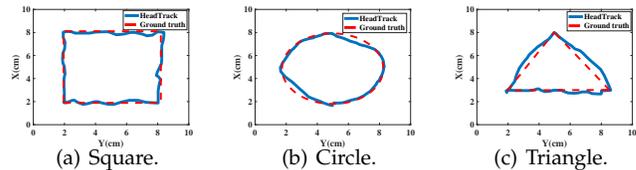


Fig. 12: Performance of track restoration.

**Ranging accuracy:** The acoustic-based ranging function is the basis for IA-Track to achieve head tracking. We evaluated the acoustic ranging accuracy of IA-Track in the range of 10cm to 150cm. The results are shown in the Fig. 11. When the distance between the smartphone and the wireless earphone is within 0.8m, the average error is less than 2.0mm, and when the distance exceeds 1.5m, the ranging error reaches 3.9mm. We recommend that users place their smartphones within 80cm of wireless earphones, which is a reasonable assumption since users need to place their smartphones close enough to themselves to interact.

**Track restoration:** We evaluate the tracking accuracy of IA-Track after adding the IMU and opportunity calibration mechanism. In this experiment, we asked volunteers to hold a wireless earphone and spend a triangle, a circle and a square on the table. The starting position of the wireless earphone is 30cm away from the Smartphone. Fig. 12(a), Fig. 12(b) and Fig. 12(c) shows examples of plot trajectories for squares, circles, and triangles. We have found that these graphs can be reproduced well for IA-Track. But after a more in-depth observation, we found that IA-Track is better than square and triangle for the restoration of circular trajectories. Because the system is using the IMU as the basis for trajectory restoration, while the square will have a sudden change of direction during the drawing, the process of this direction change may introduce errors in the way of opportunity calibration. On the other hand, the error of the circle will be small in the way of restoring the trajectory. In practical application scenarios, the motion trajectory of the head will also be closer to a circular trajectory.

### 6.2 overall performance

**Performance vary across different participants:** We assess the tracking accuracy of IA-Track across a diverse range of participants' head motions. The 50 participants were categorized into five distinct groups according to their height: 161cm-165cm, 166cm-170cm, 171cm-175cm, 176cm-180cm, and 181cm-185cm. As shown in Fig. 13, when evaluating two specific movement types backward and forward, as well

as left and right the positional tracking error of IA-Track hovers around 6cm. However, in the case of random head movements, this error margin expands to 8cm. Notably, it's important to highlight that the tracking accuracy of IA-Track remains consistently reliable across all participants, indicating its applicability across a broad spectrum of user groups.

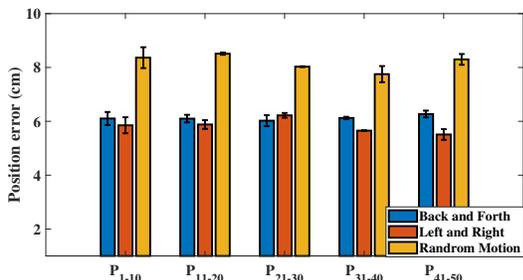


Fig. 13: Ranging Accuracy Interconnected with Smartphone.

**Comparison with other approaches:** We compare the performance of IA-Track with ArmTrack (IMU-based) [30], EchoTrack (FMCW-based) [31], and LLAP(Phase-based) [11]. As shown in Fig. 14(a), IA-Track performs best in 3-D spatial tracking, because acoustic-based methods suffer from severe multipath interference during tracking. Although enabling a microphone array can solve this problem, it will greatly increase hardware costs. Besides, IA-Track is ahead of the other three approaches in terms of delay as shown in Fig. 14(b), because IA-Track does not use complex FFT algorithms. From the perspective of performance comparison, IA-Track has achieved a balance between accuracy and computational complexity.

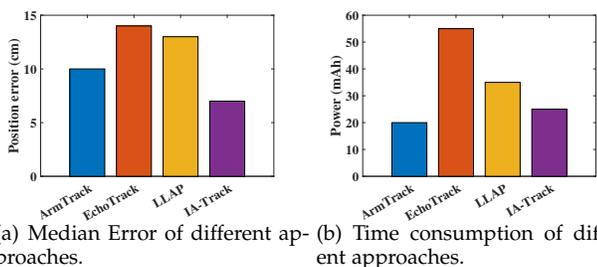


Fig. 14: Tracking performance comparison with different approaches.

### 6.3 Other factors

**Different smartphone placement:** We place the smartphone in different positions (holder, on a table, and in the hand as shown in Fig. 15(a)) to study the effect of different positions on IA-Track. We control the distance between the earphone and the smartphone at 50cm in these three positions. The final results are shown in Fig. 15(b). The average position errors of IA-Track under three different conditions are 6.5cm, 6.5cm, and 8.2cm, respectively. When a smartphone is held in the hand, the hand produces involuntary movements, thus interfering with tracking accuracy.

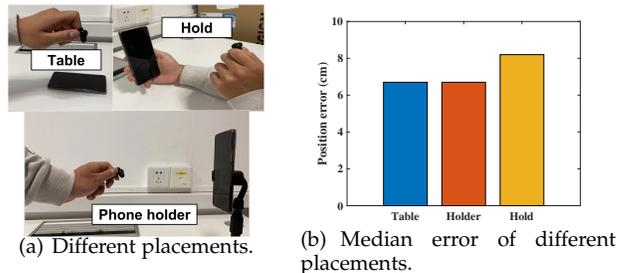


Fig. 15: The impact of different placements.

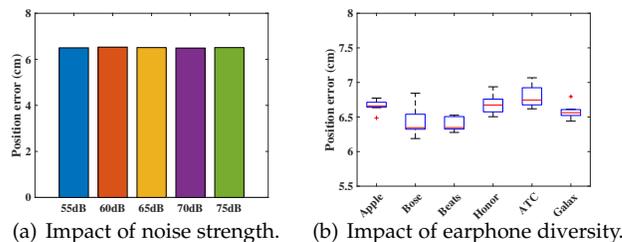


Fig. 16: Impact of noise strength and earphone diversity.

**Impact of noise strength:** Acoustic-based ranging systems all face the problem of environmental noise interference. The potential application scenarios of IA-Track are AR/VR, driving, and other scenarios with severe acoustic interference. In order to verify the anti-interference ability of IA-Track under different noise levels, we conducted experiments in a quiet conference room. The distance between the smartphone and the wireless earphone is controlled between 50cm. We used speakers next to the system to play noises with different sound pressure levels. The experimental results are shown in Fig. 16(a). It can be seen that the IA-Track can maintain stable operation under different noise levels. This is because IA-Track uses a constant 16kHz signal for acoustic ranging, and this frequency band can be effectively distinguished from acoustic signals in the "audible" range.

**Impact of earphone diversity:** Wireless earphones are the platform on which IA-Track can implement functions, so it is important to implement functions on different wireless earphones. We have selected six wireless earphones with varying configurations of price: Apple AirPods, Bose, Beats, Honor, Audio-Technica, and Samsung Galaxy Buds+. The result is shown in Fig. 16(b). We found the Boser and Beats to perform the best, possibly because the transmission ability

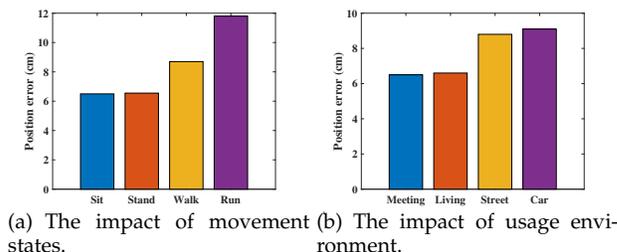


Fig. 17: Impact of other Factors.



Fig. 18: Applications in driving.

of these two types of high-frequency signals could be more robust. Therefore, if the acoustic tracking function is to be implemented on the earphone, it is necessary to improve the high-frequency transmission capability of the wireless earphone.

**The impact of different movement states:** We evaluated the performance of IA-Track in different motion states of the participants. Specifically, we set four movement states of sit, stand, walk and run. The experimental results are shown in Fig. 17(a). When the participant maintains a fixed posture such as sit and stand, the average position error does not exceed 6.5cm, but when the participant walks and runs, the position error increases, which is due to the tracking error caused by body movement. Especially when running, the movement of the smartphone will also cause the failure of the opportunity calibration.

**The impact of usage environment:** We conducted tests on the IA-Track across four distinct environments: a meeting room, a living room, a street setting, and a moving car. Within the meeting room and living room scenarios, we instructed our volunteers to maintain a stationary posture. For the street trials, participants were guided along a pre-determined path. Meanwhile, the car remained in motion at a consistent velocity during the vehicular assessment. The results, illustrated in Fig. 17(b), revealed median positional errors of 6.5cm, 6.6cm, 8.8cm, and 9.1cm for the respective environments. Although there is an observable impact on the IMU in walking and when subjected to vehicular vibrations, the effect on performance remains limited.

#### 6.4 Application in driving

IA-track has a wealth of potential application scenarios, and we test the performance of IA-track in a driving environment. In a real driving scene, the driver's head movement can reflect the driver's driving state. For example, the driver needs to frequently look to the left and to the right to observe the driving route. When the driver feels drowsy, the driver may perform actions such as nodding and yawning, and these actions can be abstracted into the trajectory of the head in three-dimensional space as shown in Fig. 18(a).

As shown in Fig. 18(b), we asked the driver to wear earphones in a smooth-driving car, and we fixed the mobile phone on the air outlet of the vehicle. The vehicle is driving on an unmanned closed road at a speed of 15km/h. We ask drivers to nod, yawn, turn left, turn right, etc. We used the same DTW algorithm as DriverSonar [13] for

benchmarking to classify the trajectories of these actions. Then we experiment. During the driving process, we will randomly instruct the driver to make him nod, yawn, turn left, turn right, etc. In contrast, we compare the accuracy and latency of IA-track with existing work DriverSonar [13] and D3-Gurad [32] on driving maneuvers. We found that the motion detection accuracy of IA-track is higher than that of DriverSonar and D3-Gurad in Fig. 18(c) because IA-track captures the complete trajectory of the driver's head motion in the time domain, while DriverSonar and D3-Gurad are trained only with incomplete motion trajectories.

## 7 RELATED WORK

We discuss the recent work on earphones, IMU-based, and acoustic-based separately below:

**Sensing based on earable devices:** As an important interface of human-computer interaction equipment, wireless headsets are expected to fundamentally promote the development of human wireless sensing applications. There are many excellent ear-worn devices in the field of ubiquitous computing. EarphoneTrack [12] has designed an acoustic headphone motion tracking system. Sabrina et al. [33] used an accelerometer in an in-ear headset to sense the user's facial expression. McGill et al. [34] discussed the impact of acoustic transparency, compared the directional tracking and acoustic noise reduction capabilities of different indoor and outdoor headphones through experiments, and put forward suggestions on how to improve the application of headphones in the field of mixed reality. EarHealth [35] uses commercial smart earphones to monitor the health of the user's ear canal, which is a low-cost, non-invasive and efficient way to monitor ear health. Ferlini et al. [36] used eSense [37] to study the value changes of gyroscopes and acceleration sensors when people's head is moving, but they did not carry out more in-depth trajectory restoration. IA-Track enables fine-grained head motion restoration.

**IMU-based sensing:** Most of the work using IMU for perception is related to trajectory restoration and authentication. The benefit of tracking the elbow is to overcome the difficulty of obtaining valid observations with a smart watch. Li et al. [38] originally used the IMU in the mobile phone to restore the moving vehicle trajectory. Li found that the traditional Kalman filter would produce severe cumulative errors and introduced chance calibration. ArmTrak [30] uses an IMU based on human elbow tracking. For the IMU-based

authentication problem, Cao et al. [39] found that different users generated different vibration characteristics when they tapped into their smartphones. Su et al. [40] used the IMU in a smartphone to restore voice calls, revealing a major privacy hazard in commercial smart devices. Taprint [41] uses unique knuckle vibration characteristics to authenticate the position of the user's hand when tapping. TouchPass [42] deploys an additional built-in motor to generate vibrations, and then analyzes the characteristics of the user's finger feedback for authentication when the user's finger touches the screen.

**Acoustic-based sensing:** Acoustic-based sensing has been greatly developed recently, and with the development of smart devices, more and more commercial smart devices can use acoustic sensors for wireless sensing. Xu et al. [43] used an acoustic sensor on a smartphone to capture changes in the angle of the driver's hand on the steering wheel. BatMapper [44] proposes a system for restoring corridor maps by emitting ultrasonic waves into the environment from a user's handheld smartphone. EchoPrint [45] provides a user authentication scheme that combines visual and acoustic features. UltraSE [46] uses ultrasound for single-channel speech enhancement in commercial equipment. DriverSonar [13] uses commercial smart devices to detect dangerous driving in a moving vehicle. BlinkListener [47] finds out the acoustic response characteristics corresponding to the blink pattern, and uses commercial smart devices to perform blink detection for the first time. CanalScan [48] uses existing smartphones to detect lingual and jaw movements by capturing sound signals from the ear canal. SpeedTalker [49] uses the phone's built-in microphone and camera to estimate the speed of the car through a combination of sound and image signals. Earecho [50] uses acoustic signals to capture unique structural features in the human ear canal for authentication.

## 8 DISCUSSION

Although IA-Track has many attractive advantages and features, it is still only a conceptual prototype and is far from practical application. At the same time, there is still room for improvement in user experience, computational complexity, and reliability. The specific limitations are as follows:

**High frequency response of wireless earphone:** The high frequency response of the wireless earphone is very limited, if the wireless earphone can have higher bandwidth and higher frequency response in the future. Then more fine-grained acoustic tracking can be done on the wireless earphone. At the same time, we also support manufacturers to expand more hardware devices into wireless earphones, and truly expand the application of wireless earphones in human-computer interaction.

**Sensing range:** This acoustic-based perception scheme is still deficient in perception range. When the sensing range exceeds 1.8m, the sensing error will increase greatly. In actual VR/AR applications, a farther perception range is required. Therefore, IA-Track is only suitable for simple human-computer interaction and a small range of head movements. Such as monitoring the driver's head movement in the driving environment.

**User experience:** Head tracking with wireless earphones is not optimal due to the ear discomfort associated with long-term wear of wireless earphones. In future work, we could actually use the same IMU combined with opportunistic calibration ideas in VR glasses or everyday glasses. The application of head tracking is not limited to VR/AR scenes but can also be used for daily head and neck health monitoring.

## 9 CONCLUSION

In this paper, we propose IA-Track. A low-cost, friendly, low-computational, ear-worn, and universal head-tracking solution. It can be easily applied in VR/AR and daily life scenarios. Based on the IMU's trajectory restoration, we connect the wireless earphone with the smartphone through an acoustic signal (a high-frequency signal with a constant frequency that exceeds the range of human hearing). We introduce opportunistic calibration into the traditional Kalman filter-based trajectory restoration method to achieve high-precision and high-efficiency trajectory restoration. It provides an up-and-coming solution for the application of wireless earphones in VR/AR environments and more life scenarios.

## REFERENCES

- [1] G. A. headset., <https://www.theverge.com/2022/1/20/22892152/google-project-iris-ar-headset-2024.>, 2022.
- [2] A. Glasses., <https://www.tomsguide.com/news/apple-glasses>, 2022.
- [3] Z. Wang, M. Xu, N. Ye, F. Xiao, R. Wang, and H. Huang, "Computer vision-assisted 3d object localization via cots rfid devices and a monocular camera," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 893–908, 2021.
- [4] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.
- [5] H. Drewes, A. De Luca, and A. Schmidt, "Eye-gaze interaction for mobile phones," in *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, 2007, pp. 364–371.
- [6] S. Mayer, G. Laput, and C. Harrison, "Enhancing mobile voice assistants with worldgaze," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–10.
- [7] M. Song, Z. Ou, E. Castellanos, T. Ylipiia, T. Kmrinen, M. Siekkinen, A. Yl-Jski, and P. Hui, "Exploring vision-based techniques for outdoor positioning systems: A feasibility study," *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, pp. 3361–3375, 2017.
- [8] H. Cheng and W. Lou, "Pd-fmcw: Push the limit of device-free acoustic sensing using phase difference in fmcw," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2022.
- [9] X. Wang, K. Sun, T. Zhao, W. Wang, and Q. Gu, "Dsw: One-shot learning scheme for device-free acoustic gesture signals," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2022.
- [10] K. Sun, T. Zhao, W. Wang, and L. Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 591–605.
- [11] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 82–94.
- [12] G. Cao, K. Yuan, J. Xiong, P. Yang, Y. Yan, H. Zhou, and X.-Y. Li, "Earphonetrack: involving earphones into the ecosystem of acoustic motion tracking," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 95–108.
- [13] H. Jiang, J. Hu, D. Liu, J. Xiong, and M. Cai, "DriverSonar: Fine-grained dangerous driving detection using active sonar," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–22, 2021.

- [14] J. Hu, H. Jiang, D. Liu, Z. Xiao, S. Dustdar, J. Liu, and G. Min, "Blinkradar: Non-intrusive driver eye-blink detection with uwb radar," in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022, pp. 1040–1050.
- [15] M. Kumar, A. Paepcke, and T. Winograd, "Eyepoint: practical pointing and selection using gaze and keyboard," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007, pp. 421–430.
- [16] M. Bâce, S. Staal, and A. Bulling, "Accurate and robust eye contact detection during everyday mobile device interactions," *arXiv preprint arXiv:1907.11115*, 2019.
- [17] D. Osokin, "Real-time 2d multi-person pose estimation on cpu: Lightweight openpose," *arXiv preprint arXiv:1811.12004*, 2018.
- [18] C. Katsini, Y. Abdrabou, G. E. Raptis, M. Khamis, and F. Alt, "The role of eye gaze in security and privacy applications: Survey and future hci research directions," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–21.
- [19] J. Y. Kaminski, D. Knaan, and A. Shavit, "Single image face orientation and gaze detection," *Machine Vision and Applications*, vol. 21, no. 1, pp. 85–98, 2009.
- [20] S. Martin, A. Tawari, E. Murphy-Chutorian, S. Y. Cheng, and M. Trivedi, "On the design and evaluation of robust head pose for visual user interfaces: Algorithms, databases, and comparisons," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2012, pp. 149–154.
- [21] X. Xie, K. G. Shin, H. Yousefi, and S. He, "Wireless csi-based head tracking in the driver seat," in *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, 2018, pp. 112–125.
- [22] M. Kotaru and S. Katti, "Position tracking for virtual reality using commodity wifi," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 68–78.
- [23] W. Mao, J. He, and L. Qiu, "Cat: high-precision acoustic motion tracking," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 69–81.
- [24] A. Wang and S. Gollakota, "Millisonic: Pushing the limits of acoustic motion tracking," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.
- [25] Y. Wang, J. Ding, I. Chatterjee, F. Salemi Parizi, Y. Zhuang, Y. Yan, S. Patel, and Y. Shi, "Faceori: Tracking head position and orientation using ultrasonic ranging on earphones," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–12.
- [26] I. sensor, <https://www.alubi.cn/>, 2022.
- [27] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 15–29.
- [28] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, Y. Liu, and X. Wang, "Swadloon: Direction finding and indoor localization using acoustic signal by shaking smartphones," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2145–2157, 2014.
- [29] R. E. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, vol. 1, no. 2, pp. 152–192, 1963.
- [30] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and i can track my user's arm," in *Proceedings of the 14th annual international conference on Mobile systems, applications, and services*, 2016, pp. 85–96.
- [31] H. Chen, F. Li, and Y. Wang, "Echotrack: Acoustic device-free hand tracking on smart phones," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.
- [32] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang, "D 3-guard: Acoustic-based drowsy driving detection using smartphones," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1225–1233.
- [33] S. A. Frohn, J. S. Matharu, and J. A. Ward, "Towards a characterisation of emotional intent during scripted scenes using in-ear movement sensors," in *Proceedings of the 2020 International Symposium on Wearable Computers*, 2020, pp. 37–39.
- [34] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic transparency and the changing soundscape of auditory mixed reality," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–16.
- [35] Y. Jin, Y. Gao, X. Guo, J. Wen, Z. Li, and Z. Jin, "Earhealth: an earphone-based acoustic otoscope for detection of multiple ear diseases in daily life," in *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 2022, pp. 397–408.
- [36] A. Ferlini, A. Montanari, C. Mascolo, and R. Harle, "Head motion tracking through in-ear wearables," in *Proceedings of the 1st International Workshop on Earable Computing*, 2019, pp. 8–13.
- [37] C. Min, A. Mathur, and F. Kawsar, "Exploring audio and kinetic sensing on earable devices," in *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*, 2018, pp. 5–10.
- [38] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 605–616.
- [39] H. Cao, H. Jiang, D. Liu, and J. Xiong, "Evidence in hand: Passive vibration response-based continuous user authentication," in *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2021, pp. 1020–1030.
- [40] W. Su, D. Liu, T. Zhang, and H. Jiang, "Towards device independent eavesdropping on telephone conversations with built-in accelerometer," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 4, pp. 1–29, 2021.
- [41] W. Chen, L. Chen, Y. Huang, X. Zhang, L. Wang, R. Ruby, and K. Wu, "Taprint: Secure text input for commodity smart wristbands," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.
- [42] X. Xu, J. Yu, Y. Chen, Q. Hua, Y. Zhu, Y.-C. Chen, and M. Li, "Touchpass: towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.
- [43] X. Xu, J. Yu, Y. Chen, Y. Zhu, and M. Li, "Leveraging acoustic signals for vehicle steering tracking with smartphones," *IEEE Transactions on Mobile Computing*, vol. 19, no. 4, pp. 865–879, 2019.
- [44] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Batmapper: Acoustic sensing based indoor floor plan construction using smartphones," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 42–55.
- [45] B. Zhou, J. Lohokare, R. Gao, and F. Ye, "Echoprint: Two-factor authentication using acoustics and vision on smartphones," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 321–336.
- [46] K. Sun and X. Zhang, "Ultras: single-channel speech enhancement using ultrasound," in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, 2021, pp. 160–173.
- [47] J. Liu, D. Li, L. Wang, and J. Xiong, "Blinklistener: "listen" to your eye blink using your smartphone," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–27, 2021.
- [48] Y. Cao, H. Chen, F. Li, and Y. Wang, "Canalscan: Tongue-jaw movement recognition via ear canal deformation sensing," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.
- [49] X. Lu, L. Xie, Y. Yin, W. Wang, Y. Bu, Q. Guo, and S. Lu, "Speedtalker: Automobile speed estimation via mobile phones," *IEEE Transactions on Mobile Computing*, 2020.
- [50] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "Earecho: Using ear canal echo for wearable authentication," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–24, 2019.



**Jingyang Hu** is currently a third-year Ph.D. student with the College of Computer Science and Electronic Engineer, Hunan University, China. From 2022 to 2023, he works as a joint Ph.D. student at the School of Computer Science and Engineering at Nanyang Technological University (NTU), Singapore. He has published papers in ACM UbiComp 2021, IEEE ICDCS 2022, IEEE ICDCS 2023, etc. His research interests include wireless sensing and deep learning.



**Hongbo Jiang** is now a full professor in the College of Computer Science and Electronic Engineering, Hunan University. He was a professor at Huazhong University of Science and Technology. He received a Ph.D. from Case Western Reserve University in 2008. He has been serving on the editorial board of IEEE/ACM ToN, IEEE TMC, ACM ToSN, IEEE TNSE, IEEE TITS, IEEE IoT-J, etc. He was also invited to serve on the TPC of IEEE INFOCOM, ACM WWW, ACM/IEEE MobiHoc, IEEE ICDCS, IEEE ICNP, etc. He is an elected Fellow of IET (The Institution of Engineering and Technology), Fellow of BCS (The British Computer Society), Senior Member of ACM, Senior Member of IEEE, and Full Member of IFIP TC6 WG6.2. Now his research focuses on computer networking, especially, wireless networks, data science in Internet of Things, and mobile computing.



**Daibo Liu** received the Ph.D. degree in computer science and engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2018. He was a Visiting Researcher with School of Software, Tsinghua University from 2014-2016, and Department of Electrical and Computer Engineering, University of Wisconsin-Madison from 2016-2017. He is currently an assistant professor with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. His research interests cover the broad areas of low power wireless networks, mobile and pervasive computing, and system security. He is a member of the IEEE and ACM.



**Zhu Xiao** (M'12-SM'19) received the M.S. and Ph.D. degrees in communication and information systems from Xidian University, China, in 2007 and 2009, respectively. From 2010 to 2012, he was a Research Fellow with the Department of Computer Science and Technology, University of Bedfordshire, U.K. He is currently a Full Professor with the College of Computer Science and Electronic Engineering, Hunan University, China. His research interests include wireless localization, Internet of Vehicles and intelligent transportation systems. He is a senior member of the IEEE and serving as an Associated Editor for IEEE Transactions on Intelligent Transportation Systems.



**Qibo Zhang** received his B.S degree in the College of Electronic and Engineering from Nanjing University of Aeronautics and Astronautics in 2018 and his M.S degree in the College of Computer Science and Electronic Engineering from Hunan University in 2021. He is currently pursuing the PhD in the College of Computer Science and Electronic Engineering, Hunan University. His research interests include electromagnetic side channels and wireless sensing.



**Jiangchuan Liu** (S'01-M'03-SM'08-F'17) is a University Professor in the School of Computing Science, Simon Fraser University, British Columbia, Canada. He is a Fellow of The Canadian Academy of Engineering, an IEEE Fellow, and an NSERC E.W.R. Steacie Memorial Fellow. He was an EMC-Endowed Visiting Chair Professor of Tsinghua University (2013-2016). In the past he worked as an Assistant Professor at The Chinese University of Hong Kong and as a research fellow at Microsoft Research Asia.

He received the BEng degree (cum laude) from Tsinghua University, Beijing, China, in 1999, and the PhD degree from The Hong Kong University of Science and Technology in 2003, both in computer science. He is a co-recipient of the inaugural Test of Time Paper Award of IEEE INFOCOM (2015), ACM SIGMM TOMCCAP Nicolas D. Georganas Best Paper Award (2013), and ACM Multimedia Best Paper Award (2012).

His research interests include multimedia systems and networks, cloud and edge computing, social networking, online gaming, and Internet of things/RFID/backscatter. He has served on the editorial boards of IEEE/ACM Transactions on Networking, IEEE Transactions on Big Data, IEEE Transactions on Multimedia, IEEE Communications Surveys and Tutorials, and IEEE Internet of Things Journal. He is a Steering Committee member of IEEE Transactions on Mobile Computing and Steering Committee Chair of IEEE/ACM IWQoS (2015-2017). He is TPC Co-Chair of IEEE INFOCOM'2021.



**Schahram Dustdar** received the Ph.D. degree in business informatics from the University of Linz, Austria, in 1992.

He is currently a Full Professor of computer science (informatics) with a focus on internet technologies heading the Distributed Systems Group, TU Wien, Wien, Austria. He has been the Chairman of the Informatics Section of the Academia Europaea, since December 2016.

Prof. Dustdar has been a member of the IEEE Conference Activities Committee (CAC), since 2016, the Section Committee of Informatics of the Academia Europaea, since 2015, and the Academia Europaea: The Academy of Europe, Informatics Section, since 2013. He was a recipient of the ACM Distinguished Scientist Award in 2009 and the IBM Faculty Award in 2012. He is an Associate Editor of the IEEE TRANSACTIONS ON SERVICES COMPUTING, ACM Transactions on the Web, and ACM Transactions on Internet Technology. He is on the Editorial Board of IEEE.