# HeadTrack: Real-Time Human–Computer Interaction via Wireless Earphones

Jingyang Hu, *Graduate Student Member, IEEE*, Hongbo Jiang, *Senior Member, IEEE*, Zhu Xiao, *Senior Member, IEEE*, Siyu Chen, *Student Member, IEEE*, Schahram Dustdar, *Fellow, IEEE*, and Jiangchuan Liu, *Fellow, IEEE*

***Abstract*— Accurate head movement tracking is crucial for virtual reality and Metaverse in ubiquitous human-computer interaction (HCI) applications. Existing works for head tracking with wearable VR kits and wireless signals require expensive devices and heavy algorithmic processing. To resolve this problem, we propose HeadTrack, a low-cost, high-precision head motion tracking system that uses commercially available wireless earphones to capture the user's head motion in real-time. HeadTrack uses smartphones as 'sound anchors' and emits inaudible chirps picked up by the user's wireless earphones. By measuring the time-of-flight of these signals from the smartphone to each microphone on the earphone, we can deduce the user's face orientation and distance relative to the smartphone, enabling us to accurately track the user's head movement. To realize HeadTrack, we use the cross-correlation method to optimize the Frequency Modulated Continuous Wave (FMCW) based acoustic ranging method, which solves the problem of insufficient wireless earphone bandwidth. Moreover, we solve the problems of asynchronous startup time between devices and the existence of sampling frequency offset. We conduct excessive experiments in real scenarios, and the results prove that HeadTrack can continuously track the direction of the user's head, with an average error under $6.3^\circ$ in pitch and $4.9^\circ$ in yaw.**

***Index Terms*— Human–computer interaction, acoustic sensing, acoustic ranging, head motion tracking.**

Manuscript received 15 March 2023; revised 1 August 2023; accepted 31 August 2023. Date of publication 25 December 2023; date of current version 19 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62272152, Grant 62372161, and Grant U20A20181; in part by the National Key Research and Development Program of China under Grant 2022YFE0137700; in part by the Key Research and Development Project of Hunan Province of China under Grant 2022GK2020 and Grant 2021WK2001; in part by the Hunan Natural Science Foundation of China under Grant 2022JJ30171; in part by the Open Research Fund from the Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant GML-KF-22-22 and Grant GML-KF-22-23; in part by the Shenzhen Science and Technology Program under Grant CYJ20220530160408019; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011915. *(Corresponding author: Hongbo Jiang.)*

Jingyang Hu, Hongbo Jiang, Zhu Xiao, and Siyu Chen are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China, and also with the Shenzhen Research Institute, Hunan University, Shenzhen 518055, China (e-mail: fbhhjy@hnu.edu.cn; hongbojiang@hnu.edu.cn; zhxiao@hnu.edu.cn; csy990406@hnu.edu.cn).

Schahram Dustdar is with the Research Division of Distributed Systems, TU Wien, 1040 Vienna, Austria (e-mail: dustdar@dsg.tuwien.ac.at).

Jiangchuan Liu is with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada, and also with the Jiangxing Intelligent Research and Development Department Inc., Nanjing 210000, China (e-mail: jcliu@sfu.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2023.3345381.

Digital Object Identifier 10.1109/JSAC.2023.3345381

## I. Introduction

In RECENT years, human-computer interaction (HCI) has drawn wide interest as it offers a key enabler for emerging applications such as Metaverse [1] and virtual reality (VR) [2]. Various HCI technologies can provide the Metaverse with a better user experience and more efficient interaction methods. For example, gesture recognition technology can realize the operation of virtual objects by detecting the user's gestures [3] in the virtual space. Voice recognition [4] technology enables users to use voice commands to interact with the Metaverse. In particular, virtual reality head-mounted kits, which implement head tracking for HCI, offer users a more vivid way to experience objects and scenes in virtual environments.

Existing efforts have explored a variety of approaches for tracking head motion, which is mainly divided into three categories. *i*) Wearable-based approach, including IMU [5] and VR glasses [6]. Users wear additional devices performing head motion tracking. *ii*) Vision-based approaches based on cameras for head tracking, such as Animoji [7]. Recent researches use depth cameras [8] and RGB [9] for head pose tracking. *iii*) Wireless signal-based approach. For instance, DriverSonar [10] utilizes acoustic signal for driver head motion detection, Xie et al. [11] uses on-board Wi-Fi for head steering estimation. Unfortunately, those methods have the following drawbacks. The wearable VR headsets [12] and VR glasses [13] require the deployment of additional camera arrays (an infrastructure that is not always available) to achieve cm-level tracking accuracy, which may limit their applicability to smart environments. In addition, current devices such as VR glasses are equipped with a large number of sensors, which not only increases the cost but also reduces the user experience. Vision-based methods can achieve high accuracy for head tracking, but severe privacy concerns hinder their widespread adoption. Methods based on wireless signals are often designed and tested from ideal experimental environments, easily affected by individual differences and background noise.

In this paper, to overcome the abovementioned issues, we strive to develop a low-cost head tracking system, namely HeadTrack, to achieve high-precision and real-time head motion tracking via wireless earphones. Specifically, the main advantages of the proposed HeadTrack are two-fold. First, it uses ubiquitous commercial off-the-shelf (COTS) devices, i.e., wireless earphones, rather than customized or expensive

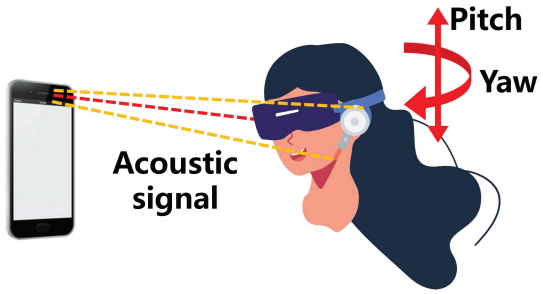



Fig. 1. Vision of HeadTrack. Tracking user's head motion using wireless earphones.

VR kits. Second, HeadTrack can adapt to different environments. To specify, we capture the wireless earphone's position change, which is highly related to head movement and steering. To that end, we propose leveraging the microphones' unique placement on each earphone to enable head motion tracking via acoustic ranging. Turning the wireless earphone into a spatial input device can unlock a broader range of interactions closely related to the user's physical environment. This approach can significantly enhance user experience in the HCI and other intelligent environments.

Inspired by the above observations, in the proposed HeadTrack as shown in Fig. 1, we utilize built-in microphones in wireless earphones to determine the user's spatial position and head orientation relative to various devices. Specifically, we use Frequency Modulated Continuous Wave (FMCW) emanating from the computing devices (e.g., smartphone in Fig. 1) and calculate the time of arrival to estimate the distance from each microphone to the device. Based on these measurements, HeadTrack can measure the distance from the device to the face and the orientation (e.g., see Pitch and Yaw in Fig. 1) of the face relative to the device. We have conducted a lot of experiments in different environments, and the experimental results show that the one-dimensional ranging error within 0.5 m is not more than 0.4 cm.

Although promising by linking wireless earphones to head-tracking technology in HCI, it is challenging to meet the demand for high-precision and real-time head tracking for HCI scenarios. In particular, we are faced with the following technical challenges:

*i*) The FMCW signal cannot provide a large enough signal bandwidth due to the inherent hardware limitations in wireless earphones and smartphones. Accordingly, it is challenging to achieve high-precision acoustic distance and head orientation angle measurements with traditional methods.

*ii*) It is challenging to ensure synchronization for the start-up times of wireless earphones and smartphones. Indeed, the start-up time difference is different whenever the FMCW signal is started to be transmitted. Such out-of-sync will inevitably cause large ranging errors, which significantly degrade the accuracy of head tracking.

*iii*) The presence of sampling frequency offsets between wireless earphones and smartphone can significantly affect acoustic-based cross-correlation, leading to severe ranging errors.

To address the first challenge, We optimize the traditional FMCW ranging scheme using the cross-correlation method to improve ranging accuracy to the millimeter level. To address the second challenge, we set an acoustic reference point on the smartphone's microphone to account for inconsistencies in device startup times. To deal with the third challenge, we design a novel sampling frequency shift algorithm to solve the sampling frequency offset problem between the signal transmission side and the receiving end. On top of that, we design multiple functions to promote HeadTrack to realize real-world HCI applications, including intelligent closing screens and device-switching functions.

The main contributions of this paper are outlined as follows:

- We propose a head motion tracking system named HeadTrack, which uses COTS wireless earphones and FMCW-based acoustic modules on smart devices to track fine-grained head motion in real-time.
- We utilize the cross-correlation to solve the problem that the wireless earphone cannot provide a large bandwidth of the FMCW signal due to its hardware limitations. Besides, we design a frequency calibration algorithm to eliminate the sampling frequency offset between devices.
- We conduct extensive experiments in real-world settings and show that the average error in tracking head rotation angles is $4.9^\circ$ in yaw and $6.3^\circ$ in pitch.
- We finally explore the potential application scenarios of HeadTrack and demonstrate the applicability.

The remainder of this paper is organized as follows. Section II introduces the related work. Section III presents the feasibility analysis of applying wireless earphone for head tracking. Section IV introduces the overview of HeadTrack. Section V present the whole system. Section VI present the implementation of HeadTrack. Section VII presents the system evaluation. Finally, Section VIII concludes the paper.

## II. Related Work

In this section, we first discuss approaches employing earable device sensing [14], [15], [16], [17], [18], [19], [20], [21], [22]. Subsequently, we present related research work in IMU-based [23], [24], [25], [26], [27], [28], [29], [30] and acoustics-based tracking methods [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41].

### A. Sensing Based on Earable Devices

As an important interface of human-computer interaction equipment, wireless earphones are expected to fundamentally promote human wireless sensing applications' development. There are many excellent ear-worn devices in the field of ubiquitous computing. EarphoneTrack [15] designed an acoustic headphone motion tracking system. Frohn et al. [16] used an accelerometer in an in-ear headset to sense the user's facial expression. McGill et al. [17] discussed the impact of acoustic transparency and compared the directional tracking and acoustic noise reduction capabilities of different indoor and outdoor headphones through experiments. Besides, they suggested how to improve the application of headphones in mixed reality.

EarHealth [18] used commercial smart earphones to monitor the health of the user's ear canal, a low-cost, non-invasive and efficient way to monitor ear health. Ferlini et al. [19] used eSense [20] to study the value changes of gyroscopes and acceleration sensors when people's head is moving. Still, they did not carry out more in-depth trajectory restoration.

### B. IMU-Based Motion Tracking

An IMU consists of an accelerometer and a gyroscope that can measure an object's acceleration and angular velocity. By analyzing and integrating these data, the motion pose of the head can be accurately estimated. Li et al. [23] integrated IMU and EEG electrodes into the helmet to alert the operator when the calculated risk level (fatigue, high stress, or error) reaches a threshold. Leelasawassuk et al. [24] assessed participants' temporal and spatial visual attention using a head-mounted inertial measurement unit (IMU). SmoothMoves [25] enables on-screen interaction by tracking objects and calculating the correlation between the objects and the user's head movements. Hwang et al. [26] perform gait symmetry analysis using head trajectory maps obtained from a single head-mounted IMU data. Fang and Fan [27] design an IMU-based calibration technique for head motion compensation for wearable gaze trackers. Head-AR [28] used weighted ensemble learning for human activity recognition through a head-mounted IMU and achieved the highest performance in the competition with 11 other algorithms.

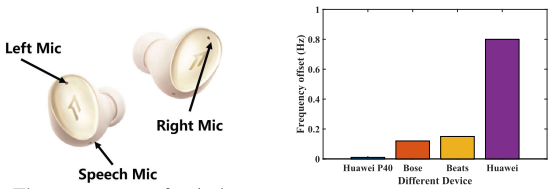
### C. Acoustic-Based Motion Tracking

Acoustic-based sensing has been widely 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. [31] used an acoustic sensor on a smartphone to capture changes in the angle of the driver's hand on the steering wheel. Wang et al. [32] designed and implemented an audio-based high-precision human respiration monitoring system using a commercial acoustic platform. EchoPrint [33] provided a user authentication scheme that combines visual and acoustic features. UltraSE [34] used ultrasound for single-channel speech enhancement in commercial equipment. In DriverSonar [10], the authors used commercial smart devices to detect dangerous driving in a moving vehicle. BlinkListener [35] found out the acoustic response characteristics corresponding to the blink pattern, and used commercial smart devices to perform blink detection for the first time. CanalScan [36] uses existing smartphones to detect lingual and jaw movements by capturing sound signals from the ear canal. Via using the phone's built-in microphone and camera, SpeedTalker [37] attempted to estimate the speed of the car through a combination of sound and image signals. FaceOri [38] use ultrasound for head Orientation tracking.

## III. Feasibility Analysis

### A. Wireless Earphones Prototype

Wireless earphones are commonly equipped with noise-canceling microphone technology to enhance call and music quality. Such technology employs various techniques, including noise-canceling algorithms and signal processing, to minimize interference from ambient noise. In doing so, the microphone can capture external sound and reduce ambient noise distractions, leading to a clearer audio experience. Moreover, some wireless earphones feature dual-microphone noise cancellation, as shown in Fig. 2(a), wherein one pair of microphones [42] records ambient noise, and another captures the user's voice, resulting in even better noise cancellation. The integration of these technologies can significantly enhance call and music quality, making wireless earphones a potent portable audio device.




(a) The prototype of wireless earphones. (b) Frequency offset of different devices.

Fig. 2. Prototype and benchmark experiments of earphone acoustic components.

HeadTrack uses a pair of wireless earphones to track the head's distance and direction, providing a better HCI experience. Overall, the workflow of HeadTrack is divided into two steps. First, we use the FMCW signal to measure a set of acoustic distance measurements from the speaker of the smart earphone to the wireless earphone. Then, we build a model to calculate the face turning (both yaw and pitch) relative to the smartphone by calculating the data these microphones pass.

### B. Frequency Offset in Wireless Earphones

The hardware of wireless earphones is a scaled-down version of that in smartphones, resulting in smaller device sizes. As such, the oscillator in wireless earphones is inferior to smartphones, leading to a higher frequency offset between the transmitting and receiving signals. While this frequency shift may not have an observable impact on calls and music playback, it could result in significant errors for fine-grained acoustic ranging and tracking. For instance, a minor frequency offset of 1 Hz can generate a ranging error of 2.125 cm [43] in one second for traditional FMCW-based ranging.

We conduct comparative experiments to compare the frequency offset of smartphones and wireless earphones. We detect the 16 kHz acoustic signal frequency offset using three wireless earphones and one smartphone. As shown in Fig. 2(b), we found that the frequency offset of the Bose QuietComfort2 and Beats Fit Pro ranges between 0.1 Hz-0.2 Hz, while the frequency offset of Huawei Freebuds Pro is approximately 0.8 Hz. Such a difference may come from the relatively weak high-frequency response of Huawei Freebuds Pro. The result proves difficult to perform high-precision acoustic tracking on the wireless earphone platform using a traditional FMCW ranging solution.

## IV. System Overview

HeadTrack is an innovative wireless earphone-based head motion tracking system that utilizes commercial wireless

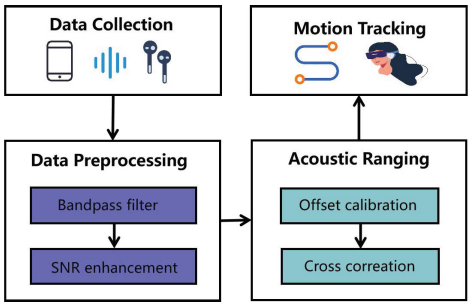



Fig. 3. System overview.

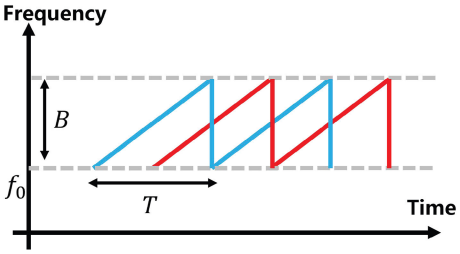



Fig. 4. The FMCW signal.

earphones for motion tracking without any hardware modification. In order to ensure accurate motion tracking, we designed a series of algorithms to filter out background noise. Since the device hardware bandwidth limits the traditional FMCW ranging scheme, we use a ranging scheme based on cross-correlation. It is noteworthy that wireless earphones and smartphones have a major problem with hardware asynchrony. To resolve this problem, we purposely designed a calibration algorithm enabling HeadTrack to track fine-grained head movement.

As shown in Fig. 3, the overall architecture of the HeadTrack system consists of four modules: *i*) data collection to obtain continuous FMCW acoustic signals; *ii*) signal preprocessing to eliminate the environment noise, improving the signal-to-noise ratio and obtaining a clean received signal; *iii*) performs frequency offset calibration followed by acoustic-based distance measurement using cross-correlation to accurately measure the distance between the wireless headset and the smartphone; and *iv*) Facial rotation measurements which measure head movements at different angles. With its innovative design, HeadTrack enables accurate and reliable COTS wireless earphone-based head motion tracking.

## V. System Design

### A. FMCW Signal Design

The FMCW signal is a commonly used method for calculating distance. It is usually estimated to be delayed according to the frequency movement spacing of the FMCW signal. As shown in Fig. 4, the blue line indicates the transmission signal and the frequency increases linearly over time, and the relationship between the transmission signal and the frequency can be expressed as:

$$f(t) = f_0 + \frac{Bt}{T}, \tag{1}$$

where $f_0$ denotes the starting frequency, $B$ and $T$ are the bandwidth and period, respectively. The phase $u(t)$ of the transmitted signal can be calculated as:

$$u(t) = \int_0^t f(t')\,dt' = 2\pi\left(f_0 t + B\frac{t^2}{2T}\right), \tag{2}$$

We then represent the transmission signal as:

$$v_{tx} = cos(u(t)), \tag{3}$$

Without loss of generality, we ignore the attenuation of the amplitude. Then the signal can be expressed as:

$$v_{rx} = cos(u(t-\tau)), \tag{4}$$

Assuming $R$ represents the distance from the receiving end to the transmitted end, $v$ represents the speed of the target. Then we define delay $\tau$:

$$\tau = 2(R + vt)/C, \tag{5}$$

where $C$ denotes the speed of acoustic signal. The receiving will multiply the signal and obtain $v_m = v_{tx} \times v_{rx}$. $v_m$ can be expressed as:

$$v_m = \cos\left(2\pi\left(f_0\tau - \frac{B\left(\tau^2 - 2t\tau\right)}{2T}\right)\right), \tag{6}$$

We define the frequency of $V_m$ to $f_b$:

$$f_b = \frac{1}{2\pi}\frac{\delta\,\text{phase}\,(v_m)}{\delta t} = \frac{2f_0 v}{C} + \frac{2BR}{CT} + \frac{4Bvt}{CT} - \frac{4Bv^2t + 4BRv}{C^2T}, \tag{7}$$

where $1/C^2$ and $v$ is closed to 0 (for slowly moving target). We can simplify the relationship between the formula to get the relationship between $r$ and $f_b$:

$$R = \frac{CT}{2B}f_b, \tag{8}$$

Based on Eq. (8), we can infer the resolution of the FMCW signal as follow:

$$\delta R = \frac{CT}{2B}\delta f_b, \tag{9}$$

According to Eq.(9), $\delta R$ relies on $\delta f_b$, and $\delta f_b$ is limited by the frequency $1/T$. Hence the distance resolution $\delta R$ of FMCW is calculated as follows:

$$\delta R \geq \frac{CT}{2B}\cdot\frac{1}{T} = \frac{C}{2B}, \tag{10}$$

The sampling rate is set to 48 kHz for widely-used wireless earphones and smartphones. According to the Nyquist sampling theorem [44], it is more appropriate to set the highest frequency of acoustic equipment below 24 kHz so as not to cause serious distortion. In addition, note that human hearing is usually below 18 kHz. We choose a signal exceeding 18 kHz so that it will not cause interference to people. According to Eq. (10), the highest resolution of 18 kHz-22 kHz FMCW signal can be expressed as:

$$\delta R \geq \frac{C}{2B} = \frac{343}{2\times 4000} = 0.0428m, \tag{11}$$

The measurement error caused by the 4.28 cm ranging resolution is catastrophic for fine-grained head tracking. To resolve this issue, we propose a more reliable ranging method in Section. V-B.

### B. Cross-Correlation Based Acoustic Ranging

The traditional FMCW-based acoustic ranging can only achieve a ranging resolution of 4.28 cm due to hardware limitations on wireless earphones. This is unsatisfactory for fine-grained head tracking. To increase the ranging resolution, we propose to use a cross-correlation-based approach. For the FMCW signal whose modulation period is $T$, we apply the cross-correlation into the transmitted signal and the peak value of the received signal to perform high-precision ranging. Motivated by this, the cross-correlation function is defined as follows:

$$R(n) = \begin{cases} \dfrac{1}{N-n} \displaystyle\sum_{m=0}^{N-\mathrm{n}-1} v_{tx}(m) \cdot v_{rx}(m+n), & n \geq 0 \\ \dfrac{1}{N-|n|} \displaystyle\sum_{m=0}^{N-|n|-1} v_{rx}(m) \cdot v_{tx}(m+n), & n < 0 \end{cases} \tag{12}$$

where $N$ is the number of signal samples obtained by the system within a complete period $T$, suppose $v_{rx}$ gets $v_{tx}$ through displacement $d$, then $R(n)$ records the displacement of the signal in discrete time, and $R(d)$ is the displacement $d$ signal peak. Therefore, the time delay $\tau$ in discrete time can be calculated as follows:

$$\tau = \frac{d}{F_s}, \tag{13}$$

where $F_s$ is the sampling frequency of the acoustic signal. We combine Eq. (5) and Eq. (13) to obtain:

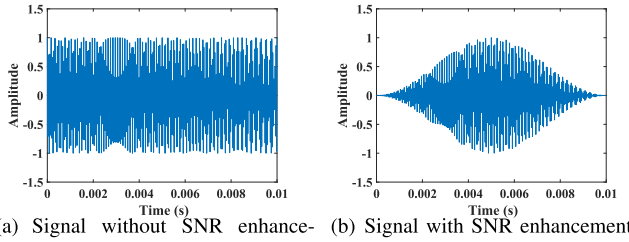
$$R = \frac{C \times d}{2F_s} - vt, \tag{14}$$

Since we calculate the resolution of $R$, we assume $v = 0$ to simplify the calculation. According to Eq. (14), the distance resolution of the FMCW signal based on cross-correlation is expressed by:

$$\delta R = \frac{C \cdot \delta d}{2F_S} = \frac{C}{2F_S}, \tag{15}$$

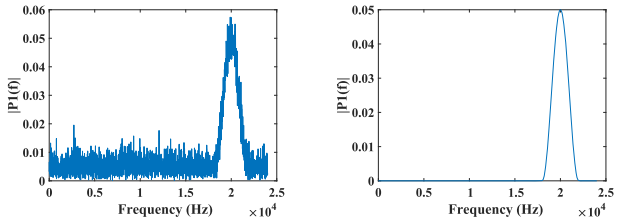
Based on Eq. (12), the resolution of cross-correlation corresponds to the cross-correlation computed for each audio sample. So the theoretical resolution of the variable $d$ is $\delta d = 1$. where $C$ is the speed of sound and $F_s$ is 48 kHz for wireless earphones and smartphones. Then we finally obtain a distance resolution $\delta R$ of 0.357 cm. To achieve a higher ranging resolution, we use interpolation technology to increase the sampling rate to 96 kHz. In this way, we can control the ranging resolution within 2mm.

Given that the signal is periodically sent from the smartphone, at the receiving end, it is required to identify the starting point of each segment of the FMCW signal in the received signal. However, the speaker and microphone are not perfectly in sync despite being embedded in the same smartphone. This results in a time difference between the transmitted and received signal within each cycle, known as Symbol Time Offset. To resolve the time offset, a cyclic prefix is added to the beginning of each transmitted signal block. This prefix is a copy of the last 16 samples of the 48-sample signal symbol.




(a) Signal without SNR enhancement. (b) Signal with SNR enhancement.

Fig. 5. Apply a Hanning window to reshape the envelope of the signal to improve the SNR.




(a) The raw FMCW signal (b) The FMCW signal with noise removal

Fig. 6. Noise removal of the FMCW signal.

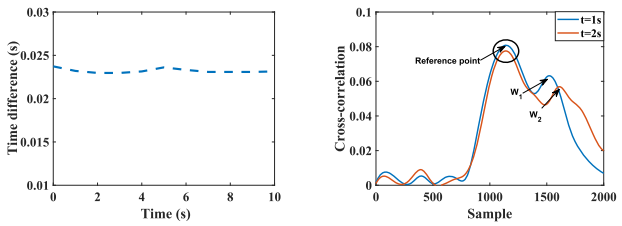
Accordingly, the time offset can be eliminated by using a cyclic prefix and corresponding data section. A pair of sliding windows $W$ and $W'$ are adopted to specify. The length of each window, denoted by $L_0$, equals 16 samples, and the distance between two windows is 32. By aligning the beginning of the window W with the beginning of the cyclic prefix in the transmitted signal block and aligning the window $W'$ with the end of the block, the similarity between $W$ and $W'$ is maximized based on the transmitted signal. In doing so, the time offset can be estimated by minimizing the difference between these two sliding windows. We obtain:

$$\delta = \arg\min_k \left\{ \sum_{i=k}^{k+L_0} \left( |W_i| - |W_i'| \right)^2 \right\}. \tag{16}$$

The sliding window, denoted by $k$, is moved across the received signal in steps of 1 sample to estimate time offset, denoted by $d$. Once the value of time offset is determined, the start point of each signal in the received signal can be adjusted accordingly.

### C. FMCW Signal Preprocessing

We designed an FMCW acoustic signal ($f_0$ = 18 kHz, $B$ = 4 kHz, $T = 0.01s$, $F_s$ = 48 kHz), the speaker continuously transmits the FMCW signal, and the signal over 18 kHz exceeds the human audible range. The wireless earphone microphone accepts the signal at a sampling rate of 48 kHz. We first need to preprocess the signal before cross-correlation to remove noise caused by hardware and the environment.




(a) Starting time difference between earphones and smartphone. (b) The peak after cross-correlation as t=1s and t=2s.

Fig. 7. Using cross-correlation to solve the start-up desynchronization of wireless earphones and smartphones.

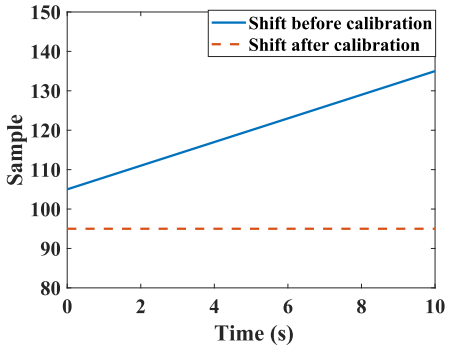



Fig. 8. Using calibration algorithm to solve sampling frequency offset.

There are two main steps in signal preprocessing, Signal to Noise Ratio (SNR) enhancement and noise removal.

*1) SNR Enhancement:* We apply a Hanning window [45] on the pulse to reshape their envelops to increase their peak to side lobe ratio, thus higher SNR. The hanning window is described by:

$$H[n] = 0.5 * \left(1 - \cos\left(\frac{2 * \pi * n}{N - 1}\right)\right), \tag{17}$$

where $N$ denote the number of samples within the window, and $n$ represent the index of a sample. Given that the pulse duration is 0.01 s and the sample rate is 48 kHz, the window spans 480 samples. To shape the discretized pulse, each element of the discretized pulse vector is multiplied element-wise with the corresponding element of the discretized Hanning window. The signal in the Fig. 5 is windowed using a Hanning window.

*2) Noise Removal:* To eliminate background noise and enhance the quality of received signals, a Butterworth bandpass [46] filter is applied with a passband range of [18 kHz, 22 kHz]. The purpose of this filtering process is to remove any unwanted noise that may obscure weak reflections. In environments with high levels of noise as shown in Fig. 6(a), this step is crucial for ensuring accurate data collection. As shown in Fig. 6(b), low frequency noise is removed from the filtered signal.

### D. Starting Time Offset Cancelling

For ideal cases, smartphones and wireless Earphones run simultaneously, then we can directly calculate the absolute distance from wireless earphones to smartphones. While under actual situations, wireless earphones and smartphones cannot reach complete synchronization, as such devices have an inherent delay. To validate this, we have tested the delay between smartphones and wireless earphone, the results are shown in Fig. 7(a). In particular, the signal received by wireless earphone takes time to return to the phone through Bluetooth, which will cause additional delay.

In the design of HeadTrack, the microphone of the smart device (e.g., smartphone) itself receives part of the speaker transmission signal without reflection, called self-interference. This part of the receiving signal generates a unique peak in the interconnected function of sending and receiving signals, representing the reference point (shown in Fig. 7(b)). Because the part of the receiving signal is not affected by the target, the amplitude and position of the reference point remain relatively stable, thereby promoting its recognition from other peaks.

Using the difference between the reference point and peak from the receiving signal, we can cancel the start time offset between the transmitter and the receiver. Subsequently, through the application FMCW distance estimation equation, we can accurately estimate the distance. This method can eliminate the interference from the starting time difference between the wireless earphone and the smartphone. The supporting reason for this method stems from the fact that the lagging shift caused by the cause of the time difference is the same for the reference point and the peak generated by the receiving signal.

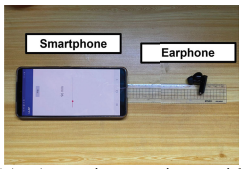
### E. Sampling Frequency Offset Calibration

The smartphone and the wireless earphone have sampling frequency offset when performing acoustic ranging. Such frequency offset will bring serious cumulative errors over time. As shown in Fig. 8, the offset of 30 sampling points is generated in ten seconds. This will produce a ranging error of around 10 cm based on Eq. (8). The impact of such errors in the fine-grained head tracking system is catastrophic. To reliably track the direction of head movement, we need to eliminate the frequency offset between the smartphone and the wireless headset. To that end, we design a frequency calibration algorithm to mitigate the sampling frequency offset between devices.

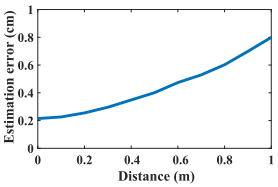
**Algorithm 1** Frequency Calibration Algorithm

**Input :** Duration $T$, and offset threshold $T_d$
**Output:** Calibration rate $r$ and calibration unit $\Delta s$
$r_{\mathrm{L}} \leftarrow 0$; $r_{\mathrm{U}} \leftarrow 1$; Find the smallest $\Delta s$ that makes $\mathrm{getShift}(r_{\mathrm{L}}, \Delta s, T)$ and $\mathrm{getShift}(r_{\mathrm{U}}, \Delta s, T)$ produce different symbols;
**while** $|\mathrm{shift}| > T_d$ **do**
  $\mathrm{shift} \leftarrow \mathrm{getShift}(r \leftarrow (r_{\mathrm{L}} + r_{\mathrm{U}})/2, \Delta s, T)$;
  **if** $\mathrm{shift} < 0$ **then**
    $r_{\mathrm{U}} \leftarrow r$;
  **else**
    $r_{\mathrm{L}} \leftarrow r$;
**return** $r$ and $\Delta s$;

For the design of the frequency calibration algorithm, we choose to supplement a displacement in the transmission




(a) Acoustic ranging with smartphone and wireless earphone.




(b) Estimation error.

Fig. 9. Estimation error of acoustic ranging.

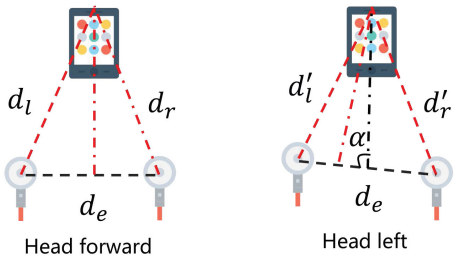



Fig. 10. HeadTrack estimates head orientation towards the acoustic signal.

signal to calibrate the sampling rate between the smartphone and the wireless earphone. Let $T_d$ and $T$ denote the threshold of the offset sampling point and the duration of time, respectively. We define parameters $r$ and $\Delta s$ to represent the permitted and calibrated acoustic samples. The procedure of the frequency calibration algorithm is presented in Algorithm 1. First, the algorithm determines the shift unit, denoted by $\Delta s$, which results in different symbols of system shift when the calibration rate $r$ equals 0 and 1. Then, a binary search method is utilized to identify an appropriate calibration rate $r$ that ensures the system shift during period $T$ to be smaller than the shift threshold $T_d$. Note that during the execution of the algorithm, the target object must remain static to guarantee calibration accuracy. Target movement may introduce shift and affect the calibration performance. A static object is only needed to obtain the calibration parameters once during system calibration before detection. The computational time incurred by the algorithm depends on the input parameters $T$ and $T_d$. Larger $T$ and smaller $T_d$ lead to longer running time but more accurate calibration rate $r$ and shift unit $\Delta s$. As shown in Fig. 8, the frequency offset is almost eliminated after our calibration algorithm, which validates the proposed calibration algorithm's effectiveness.

For acoustic ranging using wireless earphones and smartphones, we use a HUAWEI P30 and Sony wireless earphones for benchmarking, as shown in Fig. 9(a). We set our smartphones to continuously transmit FMCW signals ($F_s = 48$ kHz, $f_0 = 18$ kHz, $B = 4$ kHz) beyond human hearing, and at the same time, wireless earphones receive these signals. To establish a reference point, we place a ruler on a table (see in Fig. 9(a)), and the position that reads 0 is considered the reference position.

To evaluate the ranging resolution of HeadTrack, we move the target to 10 different positions between 0m and 1 m. The distance estimation results are shown in Fig. 9(b). Within 0.5 m, the estimated distance error is less than 0.4 cm. Furthermore, when the distance between the acoustic device is within 1 m, the maximum odometry error is around 0.8 cm, which demonstrates that fine-grained head motions can be detected using HeadTrack. However, as the distance increases, the error also increases due to the attenuation of the reflected acoustic signal.

### F. Face Orientation Angle Measurement

We utilize the acoustic-based FMCW ranging method to estimate the distance between the smartphone speaker and the wireless earphone's three microphones. The description of three microphones can be referred to Section III-B. Both wireless earphones are equipped with noise-canceling microphones. On the right side of the wireless earphone, there is a voice microphone for calls. We can calculate the yaw rate by comparing the distance between the left and right microphones. The pitch can be calculated by comparing the distance between the noise canceling microphone and the voice microphone of the wireless earphone on the right.

The determination of both yaw and pitch angles of the face towards the speaker involves measuring the angles between the face orientation vector and the vector from the speaker's location to the center of the head. These angles can be visualized as triangles in the transverse and sagittal planes of the head as shown in Fig. 10. In each triangle, the altitude corresponds to the face orientation vector, while the median corresponds to the vector between the head center and the speaker. The left microphone is designated as $d_l$, the right microphone as $d_r$, and the right speech microphone as $d_s$. The distance between the left and right microphones is $d_e$, which can be either manually measured or set to an average value across users. Additionally, the distance between the right speech microphone and the right microphone is a known quantity denoted as $d_b$. To calculate the yaw angle, our method utilizes the values of $d_l$, $d_r$, $d_e$, $d_s$, and $d_b$. We can obtain the length of the median line $d_m$ by using the formula:

$$d_m = \frac{\sqrt{2d_l^2 + 2d_r^2 - d_e^2}}{2}, \tag{18}$$

The yaw angle $\phi$ is defined by the angle $\alpha$ between the median line and the base line of the triangle in the following:

$$\alpha = \arccos\left(\frac{d_m^2 + \frac{1}{2}{d_e}^2 - {d_r}^2}{d_m d_e}\right), \tag{19}$$

The relationship between the steering angle $\phi$ and $\alpha$ is $\varphi = \alpha - 90°$. In a similar way, we can calculate the angle $\theta$ of the head pitch. Overall, through this mathematical conversion, we can track the turning orientation of the head.

### G. Cumulative Error Elimination

Although HeadTrack is able to track the rotation angle of head motion continuously, it is non-trivial to avoid the accumulation of estimation errors because the estimation process is time-dependent. To address this problem, we propose a cumulative error cancellation method.

According to people's head movement, specifically, when a person's head looks forward, we set the angle of the head

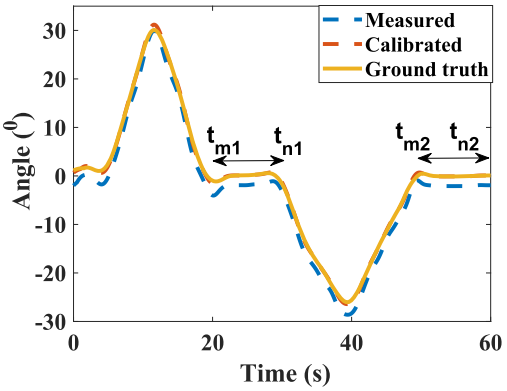



Fig. 11. Illustration of calibrating accumulative error.

to be 0 degree. We use this angle for calibration. The head is initially oriented with an angle of 0 degree and passes through two 0 degree segments starting from $t_{m1}$ to $t_{n1}$, and $t_{m2}$ to $t_{n2}$. We perform calibration when the head is at 0 degrees. To specify, HeadTrack first calculates the differences between the measured steering wheel angle and the corresponding ground truth at times $t_{n1}$ and $t_{m2}$, which are denoted by $\Delta A\left(t_{b1}\right)$ and $\Delta A\left(t_{a2}\right)$, respectively. Then the estimated angle at time $t$ can be calibrated by:

$$A'(t) = A(t) - \Delta A\left(t_{b1}\right) - (t - t_{b1}) \times l, \tag{20}$$

The variable $A(t)$ represents the head angle measurements obtained before calibration. $l$ represents the linear trend of the estimated error observed between two linear segments. This trend can be expressed as:

$$l = \frac{\Delta A\left(t_{a2}\right) - \Delta A\left(t_{b1}\right)}{t_{a2} - t_{b1}}, \tag{21}$$

We assume that the estimation error between two consecutive linear segments accumulates in a linear fashion. This observation comes from a series of experiments involving different smartphones, volunteers and wireless earphones. As shown in Fig. 11, the calibration process effectively removes the accumulated estimation error in subsequent estimates.

## VI. Implementation

In this section, we introduce the prototype implementation and data collection of HeadTrack.

### A. Hardware Prototype

we use six different wireless earphones namely Apple AirPods [47], Bose [48], Beats [49], Honor [50], Audio-Technica [51] and Samsung Galaxy Buds+ [52] for experimental evaluation. Each wireless earphone has a similar microphone configuration, with a noise-canceling microphone on the top of the earcup to pick up ambient noise and a voice mic on the lower part. We use a HUAWEI P40 smartphone as the acoustic signal transmitter. All data is transmitted via Bluetooth to an Acer TravelMate laptop (CPU: i7-1165G7, Quad-core, 2.8 GHz, RAM: 16 GB, storage: 512 GB) for real-time audio signal processing algorithms, with the sample rate and bit depth set to 48 kHz and 16 bits, respectively.

### B. Ground Truth

We conducted an experiment with 32 participants (18 females, 14 males) having an average age of 25.5 years. All participants had prior experience with headphones and smartphones. The experiments took place in a conference room. To assess the accuracy of HeadTrack, we needed to compare its performance against the ground truth. For this purpose, we utilized the Xsens MTI-3 IMU module, a compact device measuring about 1 square centimeter. The IMU module features a 9-axis accelerometer, 9-axis gyroscope, and 9-axis magnetometer. During the experiments, we placed the IMU module on top of each volunteer to capture their movements. The data was then transmitted in real-time back to a laptop, serving as the ground truth reference. Prior to each experiment, we performed calibration on the IMU module to ensure accurate measurements. Finally, we compared the data obtained from the HeadTrack system with the measurements from the IMU module (ground truth) to complete the evaluation process.

### C. Data Collection

Prior to each experiment, we conducted a comprehensive explanation of the experiment's purpose and procedures to each participant. With the assistance of the experimenter, we provided the participants with earphones for the experiment. Each participant took part in three experiments, and each experiment consisted of six subtasks: (1) participant gazed at the neutral state of the smartphone speaker for 5 seconds; (2) head moved back and forth three times; (3) head turns 3 times in yaw direction to maximum range, then returns to neutral; (4) head rotates 3 times in the pitch direction to the maximum range, then returns to the neutral position; (5) head draw a zigzag from upper left to lower right with two creases; (6) move the head randomly for five seconds. Before starting each experiment, we conducted a calibration of the HeadTrack system to ensure accurate tracking results. As a token of appreciation, each participant received a 50 dollar shopping card after completing the experiment.

## VII. Evaluation

In this section, we conduct extensive experiments to evaluate the performance of HeadTrack for head motion tracking under different conditions.

### A. Overall Performance

We first test HeadTrack's average ranging, Yaw and Pitch errors at different distances. In the experiments, we place the smartphone horizontally on the level with the participant's wireless earphones. Then constantly adjusting the distance between the smartphone and the wireless earphone, we move the distance between the smartphone and the wireless earphone from 10 cm to 205 cm, with a step length of 15 cm. As shown in Fig. 12, the average Yaw error was $4.9°$, and the average Pitch error was $6.3°$. It is worth noting that as the distance continues to increase, the measurement errors of Yaw and Pitch increase. However, at 1 m, the measurement errors of

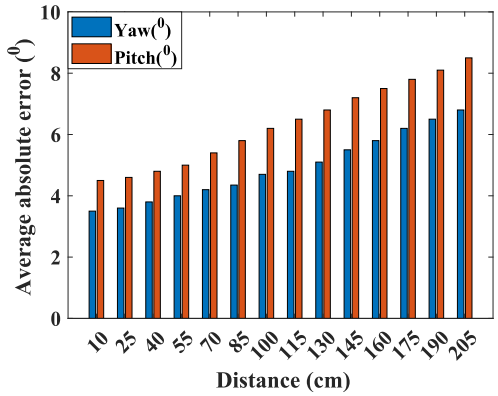



Fig. 12. Overall performance of HeadTrack.

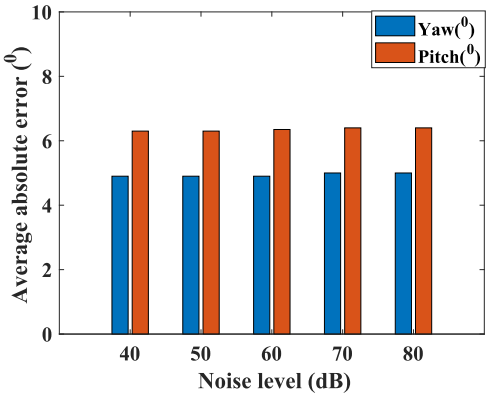



Fig. 14. The impact of different noise level.

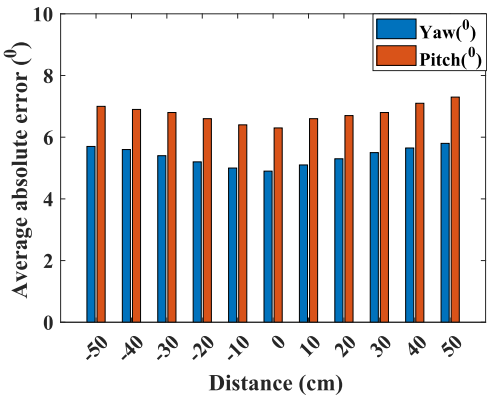



Fig. 13. Different height of smartphone.

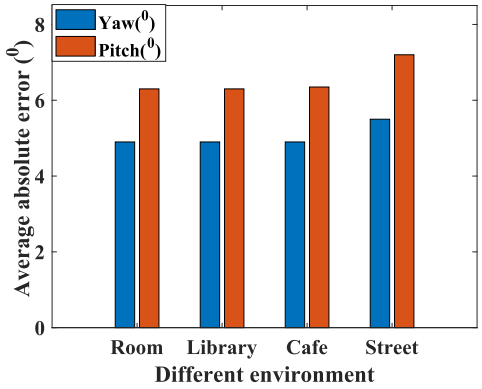



Fig. 15. The impact of different environment.

Yaw do not exceed $4.3°$, and the measurement errors of Pitch are below $6.2°$. Those results indicate that within the normal interaction range, HeadTrack can maintain a stable head turning measurement.

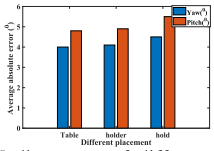
*1) Impact of Level Difference:* We investigate the effect of different smartphone positions on HeadTrack by placing the smartphone at different horizontal heights relative to the wireless earphone. We first placed the smartphone at a horizontal position of 0 degrees and kept a distance of 1m from the subject, then moved it up and down to 50 cm and $-50$ cm in steps of 10 cm, respectively. As shown in Fig. 13, the HeadTrack system performs best when the smartphone is at 0 cm. When the smartphone moves up or down, the error of Yaw and Pitch starts to increase, but within 50 cm, the maximum error of Yaw is less than $5.9°$, and the maximum error of Pitch is within $7.2°$. Those results demonstrate that the HeadTrack system can perform satisfactorily for head motion tracking.

*2) Impact of Noise Strength:* Acoustic-based systems face the problem of environmental noise interference. The potential application scenarios of HeadTrack are AR/VR and driving and other scenarios with severe noise interference. To verify the anti-interference ability of HeadTrack under different noise levels, We place speakers next to the HeadTrack system and play noises with different sound pressure levels. The experimental results are shown in Fig. 14. It can be seen that HeadTrack can maintain stable operation under different noise levels. This is because HeadTrack uses a constant 18-22 kHz signal for acoustic perception, and such frequency bands can be effectively distinguished from acoustic signals in the "audible" range.

*3) Impact of Usage Environment:* We test HeadTrack in four environments: conference room, library, cafe, and street. In indoor scenarios, the smartphone is placed on a fixed stand. On the street, the participants are required to hold their smartphones. The results are shown in Fig. 15. Due to the fixed position of the smartphone indoors, we found that HeadTrack's performance remains stable. When the smartphone is held in the street, the angle error of Yaw and Pitch increases significantly. The results show that in the scenario of holding a smartphone, hand motion impacts the results. We recommend that users place their smartphones in a fixed position to use HeadTrack to achieve a better user experience.

*4) Impact of Different Smartphone Placement:* We place the smartphone in different positions (holder, on a table, and in hand, as shown in Fig. 16(a)) to study the effect of different positions on HeadTrack. We control the distance between the earphone and the smartphone at 50 cm in these three positions. The final results are shown in Fig. 16(b). The average error of Yaw is less than 4 degrees, and the average error of Pitch is within 5 degrees under the first and second positions. When holding a mobile phone, the average error is higher than the first two positions. Because when holding the smartphone in hand, the hand produces involuntary movements, thus interfering with tracking accuracy.


(a) Different smartphone placements.




(b) Median error of different placements.

Fig. 16. The impact of different smartphone placements.

TABLE I
LATENCY OF HEADTRACK FOR DIFFERENT OPERATION

| Operation | Latency (ms) |
|---|---|
| Data preprocessing | 2.3 |
| Acoustic ranging | 25.5 |
| Motion tracking | 5.2 |

TABLE II
POWER CONSUMPTION OF HEADTRACK FOR DIFFERENT SMARTPHONE

| Smartphone | Power (mW/min) |
|---|---|
| Huawei | 2100 |
| Galaxy | 2120 |
| MI 10 | 2243 |

*5) Latency and Power Consumption:* HeadTrack uses earphones and a smartphone to perform head tracking. There is no additional sensor module for head tracking on the headset. Most earphones today stream data directly to your phone or over the web. We evaluate the power consumption and latency of HeadTrack on smartphones. We implement algorithms (including data preprocessing, acoustic ranging and motion tracking) on the smartphone. When we execute the algorithm on the mobile phone, the latency of different operations is shown in the table I. The signal processing time is usually very short, so the actual energy consumption will be much lower, as shown in the table II.

*6) Comparison With Other Approaches:* We compare HeadTrack with two existing solutions, i.e., the IMU-based tracking [53] and vision-based solution [54]. As shown in Fig. 17, we found that the steering tracking accuracy of HeadTrack is almost equivalent to that with IMU and Camera. However, vision-based and IMU-based solutions require the deployment of additional equipment. In addition, vision-based solutions may lead to potential privacy leaks, and the use of HeadTrack is broader than the deployment of cameras. Compared with the IMU solution, HeadTrack has a precise ranging function. Therefore, HeadTrack can be a potential commercial head tracking solution.

### *B. Applications in Human-Computer Interaction*

To achieve continuous head tracking, HeadTrack enforces a calibration procedure. In particular, we propose an alternative approach to attention classification that uses binarized detection to determine whether a user is looking at a device without any calibration. This approach is valuable in a variety of settings, such as attentive user interfaces. To specify, we apply bandpass filters (frequency range 18 kHz to 22 kHz) to the audio signals received from the three microphones. We then divide the frequency range evenly into 40 frequency bands

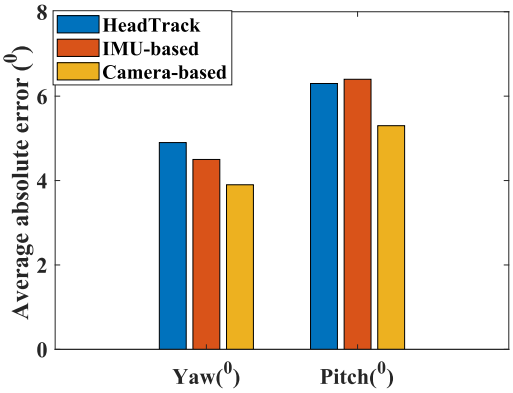
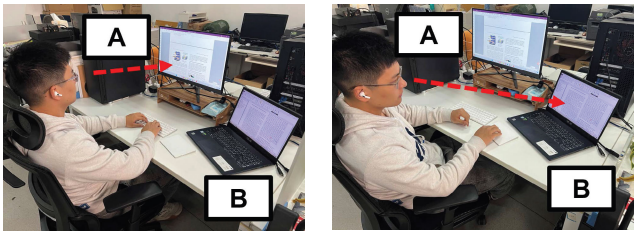



Fig. 17. Comparison with other approaches.




(a) Bluetooth keyboard and mouse switch to device A.

(b) Bluetooth keyboard and mouse switch to device B.

Fig. 18. Detecting attention across multiple devices.

and calculate the level difference (LD), which corresponds to the amplitude-level ratio of the audio signals from the two microphones in each frequency band. In total, we obtained 45 LD features (i.e., 15 LD features per microphone). Further, we include three time-difference features between the microphones, each of which represents the frequency gap (fpd) between the peaks in the frequency domain of the mixed signal from the two microphones. Using those 45 features, HeadTrack can detect whether the user is looking at the device by training a binary classifier [38].

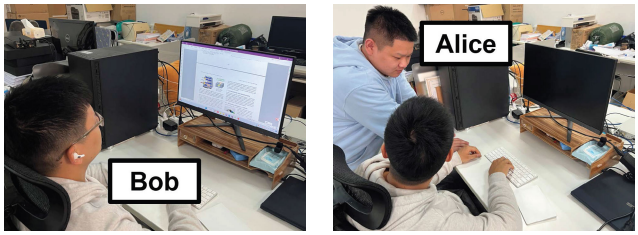
*1) Detecting Attention Across Multiple Devices:* Head tracking is regarded as an important application scenario in human-computer interaction, as it can make devices smarter and improve user experience. For example, when the user uses multiple smart devices, when the user stares at a certain device, the Bluetooth keyboard or mouse can be switched freely. We set the user's head position as the origin when the user's head is between the two devices. As shown in Fig. 18(a), when the user's head turns to device A, the Bluetooth mouse and keyboard switch to device A. When the user's head turns to device B as shown Fig. 18(b), the Bluetooth mouse and keyboard switch to device B.

*2) Attention Detection:* Employing attention detection has the potential to facilitate transfer between different tasks or points of interest. An example of this is that when the user is inconvenient to use the phone to read important information immediately, such as washing hands, the phone can automatically turn off the screen and resume after the user's attention returns to the screen. As shown in Fig. 19, The user needs to check the information on the phone screen while washing his hands, but his hands are wet and he cannot click on the phone.

(a) User is washing their hands or is inconvenient to reach the smartphone.

(b) HeadTrack detects that the user is facing the mobile phone and automatically turns on.

Fig. 19. Practical attention detection scenarios.




(a) Bob is working alone, the screen remains at normal brightness.

(b) When Bob communicates with Alice, the screen turns off.

Fig. 20. HeadTrack's attention detection feature prevents attackers from peeking at the screen.

At this time, HeadTrack detects that the user's face is facing the phone, and can automatically turn on the phone screen.

*3) Anti-Peeping Mechanism:* Another potential scenario is privacy defense at work. When Bob is concentrating on his work, he keeps eyes on the computer screen most of the time as shown in Fig. 20(a), his body will maintain a fixed posture, and his eyes will move within a small range. When Alice comes to talk to Bob, Bob's earphone will detect that Bob's eyes are away from the computer, and then turn off the screen as shown in Fig. 20(b), preventing Bob's work privacy from being leaked by Alice.

## VIII. Conclusion

In this paper, we have developed HeadTrack, an innovative solution for achieving low-cost and high-precision head tracking for HCI applications. Our proposed approach is universal, user-friendly, and can be worn on the ears. With COTS wireless earphones, HeadTrack is able to track the user's head direction continuously, with an average error of no more than $6.3^\circ$ in pitch and $4.9^\circ$ in yaw, respectively. To validate the efficacy of HeadTrack, we conduct extensive experiments and obtain promising results. Moreover, we design the practical attention detection function, which extends HeadTrack's usability to real-life applications. Our solution can be integrated with virtual reality head-mounted displays, enabling users to interact with the virtual environment more realistically.

Summarizing, HeadTrack can provide an exciting opportunity for users to explore the rapidly emerging world of the Metaverse. As the concept of the Metaverse continues to evolve, HeadTrack can help users to immerse themselves in this virtual world, where they can experience objects and scenes with greater realism. By tracking the user's head movements, HeadTrack can also enable users to engage in a variety of interactive activities, such as socializing with other avatars, attending virtual events, or playing games.

## References

[1] Z. Meng, C. She, G. Zhao, and D. De Martini, "Sampling, communication, and prediction co-design for synchronizing the real-world device and digital model in metaverse," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 288–300, Jan. 2023.

[2] O. Postolache, D. J. Hemanth, R. Alexandre, D. Gupta, O. Geman, and A. Khanna, "Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 562–573, Feb. 2021.

[3] R. Zhang, C. Jiang, S. Wu, Q. Zhou, X. Jing, and J. Mu, "Wi-Fi sensing for joint gesture recognition and human identification from few samples in human–computer interaction," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2193–2205, Jul. 2022.

[4] K. Khotimah et al., "Validation of voice recognition in various Google voice languages using voice recognition module v3 based on microcontroller," in *Proc. 3rd Int. Conf. Vocational Educ. Electr. Eng. (ICVEE)*, Oct. 2020, pp. 1–6.

[5] J. D. Hincapié-Ramos, K. Ozacar, P. P. Irani, and Y. Kitamura, "GyroWand: IMU-based raycasting for augmented reality head-mounted displays," in *Proc. 3rd ACM Symp. Spatial User Interact.*, Aug. 2015, pp. 89–98.

[6] (2022). *G. A. Headset*. [Online]. Available: https://www.theverge.com/2022/1/20/22892152/google-project-iris-ar-headset-2024

[7] (2018). *Apple.ARKit*. [Online]. Available: https://support.apple.com/en-sg/HT208986

[8] G. Borghi, M. Fabbri, R. Vezzani, S. Calderara, and R. Cucchiara, "Face-from-depth for head pose estimation on depth images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 596–609, Mar. 2020.

[9] A. F. Abate, P. Barra, C. Bisogni, M. Nappi, and S. Ricciardi, "Near real-time three axis head pose estimation without training," *IEEE Access*, vol. 7, pp. 64256–64265, 2019.

[10] H. Jiang, J. Hu, D. Liu, J. Xiong, and M. Cai, "DriverSonar: Fine-grained dangerous driving detection using active sonar," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 3, pp. 1–22, 2021.

[11] X. Xie, K. G. Shin, H. Yousefi, and S. He, "Wireless CSI-based head tracking in the driver seat," in *Proc. 14th Int. Conf. Emerg. Netw. Exp. Technol.*, Dec. 2018, pp. 112–125.

[12] X. Hu, R. Su, and L. He, "The design and implementation of the 3D educational game based on VR headsets," in *Proc. Int. Symp. Educ. Technol. (ISET)*, Jul. 2016, pp. 53–56.

[13] J. T. Panachakel, A. G. Ramakrishnan, and K. P. Manjunath, "VR glasses based measurement of responses to dichoptic stimuli: A potential tool for quantifying amblyopia?" in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 5106–5110.

[14] Y. Cao, C. Cai, A. Yu, F. Li, and J. Luo, "EarAcE: Empowering versatile acoustic sensing via earable active noise cancellation platform," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 2, pp. 1–23, Jun. 2023.

[15] G. Cao et al., "EarphoneTrack: Involving earphones into the ecosystem of acoustic motion tracking," in *Proc. 18th Conf. Embedded Networked Sensor Syst.*, Nov. 2020, pp. 95–108.

[16] S. A. L. Frohn, J. S. Matharu, and J. A. Ward, "Towards a characterisation of emotional intent during scripted scenes using in-ear movement sensors," in *Proc. Int. Symp. Wearable Comput.*, Sep. 2020, pp. 37–39.

[17] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic transparency and the changing soundscape of auditory mixed reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–16.

[18] 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 *Proc. 20th Annu. Int. Conf. Mobile Syst., Appl. Services*, Jun. 2022, pp. 397–408.

[19] A. Ferlini, A. Montanari, C. Mascolo, and R. Harle, "Head motion tracking through in-ear wearables," in *Proc. 1st Int. Workshop Earable Comput.*, Sep. 2019, pp. 8–13.

[20] C. Min, A. Mathur, and F. Kawsar, "Exploring audio and kinetic sensing on earable devices," in *Proc. 4th ACM Workshop Wearable Syst. Appl.*, Jun. 2018, pp. 5–10.

[21] X. Fan et al., "HeadFi: Bringing intelligence to all headphones," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2021, pp. 147–159.

[22] L. Ge, Q. Zhang, J. Zhang, and H. Chen, "EHTrack: Earphone-based head tracking via only acoustic signals," *IEEE Internet Things J.*, early access, p. 1, 2023, doi: 10.1109/JIOT.2023.3298412.
[23] P. Li, R. Meziane, M. J.-D. Otis, H. Ezzaidi, and P. Cardou, "A smart safety helmet using IMU and EEG sensors for worker fatigue detection," in *IEEE Int. Symp. Robot. Sensors Environ. (ROSE)*, Oct. 2014, pp. 55–60.
[24] T. Leelasawassuk, D. Damen, and W. W. Mayol-Cuevas, "Estimating visual attention from a head mounted IMU," in *Proc. ACM Int. Symp. Wearable Comput.*, 2015, pp. 147–150.
[25] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley, "SmoothMoves: Smooth pursuits head movements for augmented reality," in *Proc. 30th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2017, pp. 167–178.
[26] T.-H. Hwang, J. Reh, A. O. Effenberg, and H. Blume, "Validation of real time gait analysis using a single head-worn IMU," in *Proc. EKC*. Cham, Switzerland: Springer, 2021, pp. 87–97.
[27] C.-H. Fang and C.-P. Fan, "Effective marker and IMU based calibration for head movement compensation of wearable gaze tracking," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–2.
[28] H. Gjoreski et al., "Head-AR: Human activity recognition with head-mounted IMU using weighted ensemble learning," in *Activity and Behavior Computing*. Singapore: Springer, 2021, pp. 153–167.
[29] S. Shen, H. Wang, and R. Roy Choudhury, "I am a smartwatch and I can track my user's arm," in *Proc. 14th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2016, pp. 85–96.
[30] Y. Cao, A. Dhekne, and M. Ammar, "ITrackU: Tracking a pen-like instrument via UWB-IMU fusion," in *Proc. 19th Annu. Int. Conf. Mobile Syst., Appl., Services*, Jun. 2021, pp. 453–466.
[31] X. Xu, J. Yu, Y. Chen, Y. Zhu, and M. Li, "Leveraging acoustic signals for vehicle steering tracking with smartphones," *IEEE Trans. Mobile Comput.*, vol. 19, no. 4, pp. 865–879, Apr. 2020.
[32] 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.
[33] B. Zhou, J. Lohokare, R. Gao, and F. Ye, "EchoPrint: Two-factor authentication using acoustics and vision on smartphones," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2018, pp. 321–336.
[34] K. Sun and X. Zhang, "UltraSE: Single-channel speech enhancement using ultrasound," in *Proc. 27th Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2021, pp. 160–173.
[35] J. Liu, D. Li, L. Wang, and J. Xiong, "BlinkListener: 'Listen' to your eye blink using your smartphone," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 1–27, Jun. 2021.
[36] Y. Cao, H. Chen, F. Li, and Y. Wang, "CanalScan: Tongue-jaw movement recognition via ear canal deformation sensing," in *Proc. IEEE Conf. Comput. Commun.*, May 2021, pp. 1–10.
[37] X. Lu et al., "SpeedTalker: Automobile speed estimation via mobile phones," *IEEE Trans. Mobile Comput.*, vol. 21, no. 6, pp. 2210–2227, Jun. 2022.
[38] Y. Wang et al., "FaceOri: Tracking head position and orientation using ultrasonic ranging on earphones," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–12.
[39] Y. Cao, C. Cai, F. Li, Z. Chen, and J. Luo, "HeartPrint: Passive heart sounds authentication exploiting in-ear microphones," in *Proc. IEEE Conf. Comput. Commun.*, May 2023, pp. 1–10.
[40] P. Wang, R. Jiang, and C. Liu, "Amaging: Acoustic hand imaging for self-adaptive gesture recognition," in *Proc. IEEE Conf. Comput. Commun.*, May 2022, pp. 80–89.
[41] C. Liu, P. Wang, R. Jiang, and Y. Zhu, "AMT: Acoustic multi-target tracking with smartphone MIMO system," in *Proc. IEEE Conf. Comput. Commun.*, May 2021, pp. 1–10.
[42] (2022). *1More-Comfobuds Z.* [Online]. Available: https://global.1more.com/
[43] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "EarEcho: Using ear canal echo for wearable authentication," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 3, pp. 1–24, Sep. 2019.
[44] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, 1967.
[45] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*. New York, NY, USA: McGraw-Hill, 2001.
[46] S. F. Hussin, G. Birasamy, and Z. Hamid, "Design of Butterworth band-pass filter," *Politeknik & Kolej Komuniti J. Eng. Technol.*, vol. 1, no. 1, pp. 1–12, 2016.
[47] (2023). *AirPods Pro*. [Online]. Available: https://www.apple.com.cn/airpods-pro/
[48] (2023). *Bose*. [Online]. Available: https://www.bose.cn/zh_cn/index.html
[49] (2023). *Beats*. [Online]. Available: https://www.beatsbydre.com.cn/
[50] (2023). *Honor*. [Online]. Available: https://consumer.huawei.com/cn/audio/
[51] (2023). *Audio-Technica*. [Online]. Available: https://www.audio-technica.com.cn/
[52] (2023). *Galaxy Buds*. [Online]. Available: https://www.samsung.com/audio-sound/galaxy-buds/
[53] C. Xu, J. He, Y. Li, X. Zhang, X. Zhou, and S. Duan, "Optimal estimation and fundamental limits for target localization using IMU/TOA fusion method," *IEEE Access*, vol. 7, pp. 28124–28136, 2019.
[54] 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*, Jun. 2017, pp. 345–358.

**Jingyang Hu** (Graduate Student Member, IEEE) is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University, China. From 2022 to 2023, he was a joint Ph.D. Student with the School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore. He has published papers in ACM Ubicomp, ACM CCS, IEEE ICDCS, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and IEEE INTERNET OF THINGS JOURNAL. His research interests include wireless sensing and deep learning.

**Hongbo Jiang** (Senior Member, IEEE) received the Ph.D. degree from Case Western Reserve University in 2008. He was a Professor with the Huazhong University of Science and Technology. He is currently a Full Professor with the College of Computer Science and Electronic Engineering, Hunan University. His research interests include computer networking, especially, wireless networks, data science in the Internet of Things, and mobile computing. He is elected as a fellow of The Institution of Engineering and Technology (IET), a fellow of The British Computer Society (BCS), a Senior Member of ACM, and a Full Member of IFIP TC6 WG6.2. He has been serving on the editorial board of IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON MOBILE COMPUTING, *ACM Transactions on Sensor Networks*, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and IEEE INTERNET OF THINGS JOURNAL. He was also invited to serve on the TPC of IEEE INFOCOM, ACM WWW, ACM/IEEE MobiHoc, IEEE ICDCS, and IEEE ICNP.

**Zhu Xiao** (Senior Member, IEEE) 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, the Internet of Vehicles, and intelligent transportation systems. He is currently an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

**Siyu Chen** (Student Member, IEEE) received the B.S. degree in communication engineering from Hunan University, Changsha, China, in 2021, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering. He has published articles in IEEE TRANSACTIONS ON MOBILE COMPUTING and IEEE INTERNET OF THINGS JOURNAL. His research interests include WiFi sensing.

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

He is currently a Full Professor in computer science (informatics) with a focus on internet technologies heading the Distributed Systems Group, TU Wien, Wein, Austria. He has been the Chairperson 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 IEEE TRANSACTIONS ON SERVICES COMPUTING, *ACM Transactions on the Web*, and *ACM Transactions on Internet Technology*. He is on the editorial board of IEEE.

**Jiangchuan Liu** (Fellow, IEEE) received the B.Eng. degree (cum laude) in computer science from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree in computer science from The Hong Kong University of Science and Technology in 2003.

He was an Assistant Professor with The Chinese University of Hong Kong and a Research Fellow at Microsoft Research Asia. He was also the EMC-Endowed Visiting Chair Professor of Tsinghua University from 2013 to 2016. He is currently a University Professor with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. His research interests include multimedia systems and networks, cloud and edge computing, social networking, online gaming, and the Internet of Things/RFID/backscatter.

Dr. Liu is a fellow of The Canadian Academy of Engineering and an NSERC E.W.R. Steacie Memorial Fellow. He was a co-recipient of the Inaugural Test of Time Paper Award of IEEE INFOCOM in 2015, the ACM SIGMM TOMCCAP Nicolas D. Georganas Best Paper Award in 2013, and the ACM Multimedia Best Paper Award in 2012. He was a Steering Committee Member of IEEE TRANSACTIONS ON MOBILE COMPUTING and the Steering Committee Chair of IEEE/ACM IWQoS from 2015 to 2017. He was the TPC Co-Chair of the IEEE INFOCOM in 2021. He has served on the editorial boards for 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.