BLEAR: Practical Wireless Earphone Tracking under BLE protocol

Linfei Ge[1,2], Wentao Xie[1,2], Jin Zhang[1,*], and Qian Zhang[2,*]
[1]RITAS, CSE Dept., Southern University of Science and Technology, Shenzhen, China
[2]CSE Dept., The Hong Kong University of Science and Technology, Hong Kong, China
[*]Corresponding Author
Email: lgead@cse.ust.hk, wxieaj@cse.ust.hk, zhangj4@sustech.edu.cn, qianzh@cse.ust.hk

2024 IEEE International Conference on Pervasive Computing and Communications (PerCom) | 979-8-3503-2603-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/PERCOM59722.2024.10494424

***Abstract*—Motion tracking is an important aspect of human-computer interaction (HCI) and recent research focuses on motion tracking using earphones' embedded acoustic sensors. However, these solutions can only be deployed on wired earphones, while most of the commercial earphones are wireless ones. This limitation arises because wireless earphones utilize the Bluetooth Low Energy (BLE) protocol for handling audio data, which blocks the usage of existing acoustic sensing solutions. Firstly, the low sampling rate of BLE prevents the system from processing high-frequency ultrasounds. However, the sensing signal for earphones must be ultrasonic to prevent disturbance to the user. Secondly, BLE employs an audio compression process that is applied with different compression rates with different bandwidths. This will break the structure of wideband signals usually used for acoustic sensing. To overcome these challenges, we present BLEAR, the first earphone-tracking system compatible with the BLE audio recording protocol. To let BLE earphones receive ultrasounds, BLEAR utilizes a specially designed bandwidth conversion scheme that uses a mask signal to trigger a non-linear effect that converts high-frequency components to low-frequency ones, thereby overcoming the low audio sampling rate restriction of BLE. Additionally, by strategically designing beacon signals to align with BLE's subband compression pattern, BLEAR mitigates the influence of audio compression and achieves accurate wireless earphone tracking. We implement a wireless earphone prototype for BLEAR and conduct extensive experiments involving 8 subjects to demonstrate its feasibility. The experimental results show that BLEAR achieves a mean distance tracking error of 3.37 cm, an angle tracking error of 5.3 degrees, and an accuracy of 97.14% in recognizing 7 common user activities. This work not only introduces a BLE-compatible earphone tracking solution but also establishes a foundation for broader BLE device tracking applications.**

***Index Terms*—Human-computer interaction, earphone, motion tracking, BLE**

I. Introduction

Motion tracking plays an important role in human-computer interaction (HCI). Among these motion-tracking applications, researchers are particularly interested in earphone-based tracking because the user's attention can be inferred from the ear position, thus providing context-aware interaction for the user [1]. An example of such an application is depicted in Fig. 1 where the user is walking close to or away from a desk. To enhance the interaction experience, it is necessary to sense the activity of walking close to or away from the desk. Earphones, especially wireless earphones, are suitable for detecting these movements because they are popular accessories and are firmly attached to the user. Therefore, the user's activity can be derived by tracking the earphones' movements.

A large amount of research effort has been devoted to designing earphone tracking methods. There are three major lines of research towards this direction based on the three types of signal that an earphone can capture: Bluetooth received signal strength, inertial measurement units, and acoustic signals. First, since wireless earphones use Bluetooth Low Energy (BLE) protocol to transmit data with other devices, some works leverage BLE received signal strength (RSS) to achieve device tracking [2], [3]. Although these works demonstrate the usability of using the BLE signal to infer location information, these designs are restricted by the granularity provided by the BLE RSS signal. They can only estimate the general trending of the distance changes but not track the earphone's location quantitatively. Second, some earphone products are equipped with inertial measurement units (IMUs) [4]. These IMUs are shown to be able to perform motion tracking [5]–[7] through processing the accelerometer, gyroscope and magnetometer readings. However, pure IMU-based head tracking systems are erroneous since the on-body IMUs are intrinsically noisy because of the unconscious and inevitable motion artifacts [7]. Also, IMU-based solutions can only determine human-centered motions. Therefore, they cannot distinguish whether people are walking closer to or farther from an object. In addition, the IMU-based method cannot be widely adopted because most commercial earphones are designed without an IMU.

Recent works have shown that acoustic signals can be leveraged to track earphones or other mobile devices equipped with microphones [1], [8]–[12]. The working scenario for these systems is described as follows. An anchor device constantly transmits beacon signals such as frequency-modulated continuous wave (FMCW) and continuous wave (CW) signals. The microphone on the tracked mobile device receives the beacon signals, and the device decodes the location information from the beacons. We exclude the designs that let the mobile device transmit beacon signals and let the anchor device

This research is supported in part by RGC under Contract CERG 16204820, 16206122, AoE/E-601/22-R, Contract R8015, and 3030_006.

receive the beacons [13]. This is because the earphone speaker has extremely low power and transmits sound towards the user's ear canal rather than the air such that the beacon cannot reach the anchors. Although these works demonstrate the high accuracy of acoustic-based device tracking, there is one major limitation that prevents the wide deployment of these designs - These designs can only be applied to wired earphones where the earphones are connected to smartphones or laptops that are equipped with a rather powerful audio processing module. For wireless earphones, which are more commonly seen in recent years, these solutions are unusable. This is because wireless earphones have limited computation power and use the BLE protocol to handle audio data. It imposes two challenges that prevent using acoustic-based tracking methods, as Fig. 2 shows. First, to avoid disturbing the user while tracking the earphones, the beacon signal must be inaudible ($\geq$ 17 kHz). However, the maximum audio sampling rate under BLE is 16 kHz, which means the earphone can only handle audio signals below 8 kHz, let alone the ultrasonic beacons. Second, BLE adopts an audio compression strategy, subdividing the audio band into several subbands. Suppose a large-bandwidth beacon signal is used, and it is within the acceptable bandwidth under the BLE protocol. It will be divided into several subbands and processed with different compression rates, making its structure greatly affected and hard to perform tracking.

In this work, we resolve the above challenges and design an earphone tracking system that, for the first time, can be used with worn wireless earphones. As Fig. 3 shows, to break the low sampling rate constraint, we design a frequency conversion scheme that takes advantage of the microphone's non-linear frequency response that can help convert the high-frequency component to the low-frequency band when mixed with a properly designed mask signal. This way, the earphone can capture high-frequency signals even if they are beyond the Nyquist frequency limit. Readers can refer to Section III and Section IV for more details. To resolve the second challenge, we observe that although the entire frequency range is subdivided, the subdivision pattern is consistent as long as the BLE configuration is unchanged. Therefore, we can carefully select the narrow-band segments that remain complete after the subdivision for motion tracking.

We present BLEAR, the first wireless earphone tracking system that is compatible with BLE recording protocol with minimal hardware add-on. To utilize BLEAR, we need to set up a pair of speakers as the anchor to transmit 18.5 kHz and 19.1 kHz ultrasonic beacons separately. Meanwhile, in order to let the BLE earphone receive these ultrasounds, we propose to attach a miniature piezoelectric (PZT) transducer to trigger the non-linear effect that converts the high-frequency beacons to the low-frequency range. The rationale behind this design is that the PZT transducer emits a 20.3 kHz mask signal, which, when mixed with the two ultrasonic beacons, will produce low-frequency components because of the nonlinearity of the microphone. Notably, the two beacons in our design are used because they are necessary to achieve 2-dimension motion tracking of the earphone.

(a) (b)

Fig. 1. People walk close to/away from his/her desk.

The contributions of this work are summarized as follows.

- To the best of our knowledge, BLEAR is the first earphone tracking system that can be truly deployed on wireless earphones under the BLE audio recording protocol. This design can potentially be extended to many tracking systems that involve a BLE device.
- We propose several technical designs to overcome the challenges brought by the BLE protocol, including a frequency conversion scheme that uses a mask signal-based nonlinearity system to convert high-frequency audio into the low-frequency band to break the restriction of the low audio sampling rate of BLE, and carefully designed beacon signals to bypass the influence of BLE audio compression.
- We build a wireless earphone prototype to demonstrate the feasibility of BLEAR. We conduct extensive experiments with 8 subjects to show the performance of BLEAR. The experiment result shows that BLEAR can achieve accurate location tracking with a mean error of 3.37 cm, an angle tracking error of 5.3° and a mean accuracy of 97.14% for recognizing 7 common user activities.

II. Background

In this section, we will give some background knowledge of BLE earphones and acoustic sensing. This knowledge will help to understand the design of BLEAR.

A. Bluetooth Earphones

Bluetooth Low Energy (BLE) is a wireless communication technology designed for short-range communication between devices. It is commonly used in various applications, including wireless earphones. Usually, wireless earphones rely on the BLE protocol to exchange data with other devices, such as smartphones, tablets, or laptops.

Some studies [13] have achieved motion tracking by utilizing BLE earphones as speakers. However, this approach requires users to handle the earphone, rendering it unusable solely as an audio device. The high sampling rate of 48 kHz can be achieved when the earphone is used for audio playback only, but the sampling rate decreases significantly when recording is necessary. Wireless earphones rely on protocols like Hands-Free Profile (HFP) or Headset Profile (HSP) for audio data exchange. These protocols support bidirectional audio communication for phone calls and voice chat applications, but they only support mono channel recording at a maximum sampling rate of 16 kHz. This limited sampling rate restricts

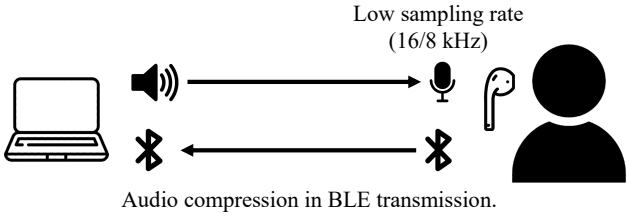


Fig. 2. The low audio sampling rate and audio compression of BLE pose limitations that prevent the acquisition of the original signal when using wireless BLE earphones for acoustic sensing.

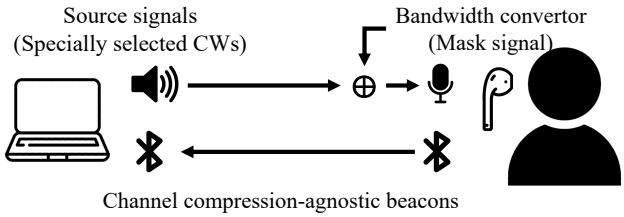


Fig. 3. Our system overcomes the limitations of BLE by utilizing a mask signal and a specially designed source audio signal.

the ability to capture ultrasound frequencies, which may be necessary for certain sensing tasks.

To best utilize the transmission bandwidth, BLE will apply compression to audio data. A typical compression algorithm includes Continuous Variable Slope Delta modulation (CVSD), Low-Complexity Subband Coding (SBC) and modified Sub-Band Coding (mSBC). Of these, we are particularly interested in mSBC as it is widely used in BLE recording protocols. The encoding process of mSBC involves dividing the audio signal into multiple subbands and applying compression to each subband. Each subband is compressed using different parameters, resulting in varying degrees of compression. As a result, mSBC significantly impacts both the audio signal quality and the system's sensing ability.

The low sampling rate and audio signal compression make BLE hard to use in practical acoustic sensing systems. However, in BLEAR, we aim to overcome these limitations by leveraging the non-linear effect and employing specially designed signals. Our goal is to achieve effective acoustic sensing within the constraints imposed by BLE.

B. Acoustic Sensing and Motion Tracking

Recent research has shown that sensing systems can be designed on commercial mobile devices that are equipped with speakers and microphones. They transform the device into a sonar system by reprogramming the speaker and microphone to sense user actions.

Various signals are designed for acoustic sensing. Continuous Wave (CW) is the most frequently used one that contains a single-frequency, continuous cosine signal. CW is usually used in motion tracking because it can track the distance change by manipulating the phase of the signal [14]. Frequency Modulated Continuous Wave (FMCW) signal is also commonly used for sensing. It is a special continuous wave signal with varying frequencies. It can be used to measure the range and velocity of objects [8].

Acoustic sensing systems typically utilize acoustic signals above 17 kHz because they are inaudible to users. To accommodate these signals, a sampling rate larger than 34 kHz is required. A common configuration for sampling rate in audio systems is 44.1/48 kHz. Additionally, acoustic sensing systems rely on original acoustic signals for modulation and demodulation processing, which is achieved through Pulse Code Modulation (PCM) for audio playing and recording in normal systems.

The CW signal is highly effective and straightforward for acoustic motion tracking. Several studies [10], [12], [14], [15] have demonstrated that CW can be utilized for phased-based distance tracking and strength-based angle tracking. Therefore, by solely employing the CW signal, it is possible to achieve accurate motion tracking.

1) Phase-based distance tracking: Phase-based distance tracking relies on utilizing the phase of a sound wave to accurately track distances. Several studies [14], [15] have successfully implemented various systems based on phase-based distance tracking, achieving impressive levels of accuracy, even at the millimeter-level.

The fundamental concept behind phase-based distance tracking is relatively simple. Suppose we have a speaker emitting a sine wave at frequency f_1 and a microphone receiving this signal. The distance between the speaker and the microphone can be determined by analyzing the phase of the sine wave. By measuring the change in phase, we can calculate the corresponding change in distance, enabling us to track the object's motion accurately.

In detail, we know the correlation between changes in distance and phase:

$$\Delta d = \frac{\Delta\phi}{2\pi} * \lambda = \frac{\phi_{d2} - \phi_{d1}}{2\pi} * \lambda. \tag{1}$$

Here, $\lambda = v_s/f_1$ represents the wavelength, where v_s is the speed of sound in air. In this context, ϕ_{d1} and ϕ_{d2} represent the phase before and after the movement, respectively. When the phase undergoes a change of 2π, it corresponds to a distance change equal to one wavelength λ.

By knowing the phase change, denoted as $\Delta\phi_d = \phi_{d2} - \phi_{d1}$, we can utilize Equation 1 to derive the change in distance and effectively track the distance.

2) Strength-based angle tracking: There is a straightforward method to achieve angle tracking by sensing the sound field generated by two speakers [12]. In our system, we utilize two speakers that emit sine waves at distinct frequencies to generate the sound field. One speaker emits a sine wave at frequency f_1, while the other emits a sine wave at frequency f_2. The combination of these two sine waves creates the sound field, which can be detected by a microphone placed within it.

When the phase difference between these sine waves is 0, indicating that they are in phase, constructive interference occurs, resulting in a high sound strength detected by the

microphone. Conversely, when the phase difference is π, indicating that they are out of phase, destructive interference occurs, leading to a low sound strength detected by the microphone. Essentially, the distribution of sound strength in the field is uneven due to the phase difference.

If $f_1 = f_2$, the phase difference is solely caused by differences in location, resulting in a static sound field. However, if $f_1 \neq f_2$, the phase difference is influenced by both location and time, creating a dynamic sound field. In practice, the sound field appears to "rotate" around the center.

Within the rotating sound field, a stationary microphone detects changes in sound strength. The frequency of these strength changes is denoted as $f_0 = |f_1 - f_2|$. If the microphone is in motion, the observed frequency of the sound field strength changes. A moving microphone will record a signal with a period larger or smaller than the standard period $T_0 = \frac{1}{f_0}$. By calculating the difference in periods ΔT, we can determine the angular speed and ultimately achieve angle tracking.

III. Beacon and Mask Signal Design

Our system, BLEAR, is specifically designed to achieve motion tracking using the BLE protocol, despite its limitations such as a limited sampling rate and audio compression. To overcome these limitations, BLEAR leverages the non-linear effect of the recording system and carefully considers the design of the beacon and mask signals. In this section, we provide a description of the non-linear effect and then explain the signal design of BLEAR in detail.

A. Nonlinearity of Recording Systems

In a recording system, the recorded signal ideally should be proportional to the input signal. However, due to imperfect microphone implementation, the recording system may exhibit nonlinearity. By leveraging this nonlinearity, we can sense higher frequency components with a restricted sampling rate. We give the rationale for this nonlinearity-based frequency conversion as follows.

Assume our system's beacon signal is a cosine wave with frequency f:

$$S(t) = cos(2\pi ft + \phi) \tag{2}$$

For an ideal recording system, the received signal is:

$$S_r(t) = AS(t) = Acos(2\pi ft + \phi) \tag{3}$$

where A is a fading factor. If we consider nonlinearity, the actual received signal is:

$$S'_r(t) = \sum_{n=0}^{\infty} A_n S(t)^n \tag{4}$$

where n is the order of the polynomial and A_n is the corresponding fading factor. To trigger the frequency conversion of a nonlinear system, we also transmit a mask signal, which is a cosine wave with frequency f_m. Therefore, the signal reaching to the microphone is

$$S(t) = cos(2\pi ft + \phi) + cos(2\pi f_m t + \phi_m) \tag{5}$$

After the nonlinear transfer function of the recording system, the received signal becomes $S'_r(t) = \sum_{n=0}^{\infty} A_n S(t)^n$. For simplicity, we only consider the first two orders of the polynomials:

$$\begin{aligned} S'_r(t) =& A_1 S(t) + A_2 S(t)^2 \\ =& A_1[cos(2\pi ft + \phi) + cos(2\pi f_m t + \phi_m)] \\ &+ A_2[cos(2\pi ft + \phi) + cos(2\pi f_m t + \phi_m)]^2 \\ =& A_1[cos(2\pi ft + \phi) + cos(2\pi f_m t + \phi_m)] \\ &+ A_2[\frac{cos(4\pi ft + 2\phi)}{2} + \frac{cos(4\pi f_m t + 2\phi_m)}{2} \\ &+ cos(2\pi(f + f_m)t + (\phi + \phi_m)) \\ &+ cos(2\pi(f - f_m)t + (\phi - \phi_m))] \end{aligned} \tag{6}$$

In the recording system, the signal will be then sent to an Anti-Aliasing Filter (AAF). Typical, AAF is a low-pass filter at half of the sampling rate f_s. Thus, all components with higher frequency than $\frac{f_s}{2}$ will be removed. If $|f - f_m| < \frac{f_s}{2} < f, f_m$, the remaining signal is:

$$S'_{r,AAF}(t) = A_2 cos(2\pi(f - f_m)t + (\phi - \phi_m)) \tag{7}$$

This way, the original transmitted signal with frequency f will be converted to $|f - f_m|$. Therefore, by leveraging the nonlinearity of the recording system and a mask signal, we can make the acoustic system sense a signal at frequency $|f - f_m|$ and its phase $\phi - \phi_m$. Note that we have $|f - f_m| < \frac{f_s}{2} < f, f_m$. It means that we can sense the phase change even if the signal frequency is beyond the Nyquist frequency $\frac{f_s}{2}$.

B. BLEAR Signal Design

As discussed in Section II-A, there are two main restrictions of BLE: low sampling rate and audio signal compression. We need to overcome these two restrictions through beacon and mask signal designs.

To address the low sampling rate limitation, we leverage the nonlinearity of the recording system. As explained Section III-A, even with a low sampling frequency, if there are two signals present - a target signal at frequency f and a mask signal at frequency f_m - such that $|f - f_m| < \frac{f_s}{2} < f, f_m$, we can still detect the phase change $\phi - \phi_m$ and perform distance tracking.

To overcome the signal compression imposed by BLE, we avoid using complex signals such as FMCW. This is because, as mentioned in Section II-B, the BLE audio compression algorithm, known as subband coding, divides the audio into several frequency bands and applies different compression ratios to each band. If a wideband signal like FMCW were employed, spanning across multiple frequency bands, it would be significantly affected by the compression algorithm. Consequently, CW signals are employed for motion tracking, as they are less susceptible to the impact of the compression algorithm. CW signals operate at a single frequency, making it easier to avoid the frequency range between two subbands.

In addition, BLE typically divides audio data into 4/8 subbands during mSBC compression. We carefully select the frequencies of the CW signals to stay clear of the cutoff frequencies of these subbands. In order to ensure that the

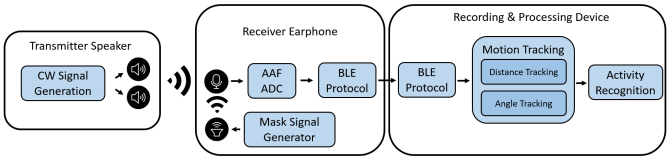


Fig. 4. System Overview.

signals remain inaudible to humans, it is recommended to use frequencies above 17 kHz. Also, the available bandwidth of a commercial audio system is up to 24 kHz because of the typically used 48 kHz sampling rate, the frequency range of the CW signals should be within 17 to 24 kHz. However, it is important to note that frequencies near the upper limit (24 kHz) should be avoided, as conventional speakers typically have a poor frequency response in that range.

IV. System Design

In this section, we will begin by providing an overview of the system. Following that, we will introduce the details of BLEAR. We will then delve into the process of how we handle the received audio data and extract motion tracking information. Lastly, we will introduce the activity recognition component, which is built upon the motion tracking results.

A. System Overview

Fig. 4 presents an overview of the BLEAR system, which consists of three main parts. The first part, "Transmitter Speaker," includes a pair of speakers that play CW signals at distinct frequencies f_1 and f_2. The second part, "Receiver Earphone," consists of BLE earphones that receive acoustic signals and transmit the data to a processing device using the BLE protocol. The last part, "Recording and Processing Device," is connected to the earphones and is responsible for recording and processing the received audio data. This device further enables motion tracking and activity recognition.

1) Transmitter Speakers: In our system, we utilize a transmitter that plays two sine waves at distinct frequencies. This allows us to achieve both distance and angle tracking. By utilizing stereo mode, we can easily control the left and right channels to transmit CW at 18.5 and 19.1 kHz independently.

2) Receiver Earphone: We need to generate a mask signal at the receiver end in order to account for the nonlinearity introduced in Section III-A, as Fig. 5(a) shows. The phase we sense is given by $\phi - \phi_m$, where ϕ is the phase of the target signal and ϕ_m is the phase of the mask signal. Both the target signal and the mask signal contribute to the resulting phase. In order to achieve distance tracking based on the target signal, we need to ensure that ϕ_m remains constant. To achieve this, we have chosen to use a PZT transducer to generate the mask signal near the earphone microphone. By keeping the PZT transducer and the microphone relatively static, we can ensure that ϕ_m remains constant, allowing us to derive ϕ from $\phi - \phi_m$ and further track the distance based on the target signal.

The PZT transducer can be driven by a voltage signal, and we utilize a channel of a 3.5mm audio playback jack to drive the PZT transducer. We have found that the PZT transducer performs well in the frequency range of 15-21 kHz, with a central frequency of 18 kHz.

Specifically, in the BLEAR system, the beacon signals are configured as CW signals with frequencies of $f_1 = 18.5$ kHz and $f_2 = 19.1$ kHz, resulting in a frequency difference of $f_0 = |f_1 - f_2| = 0.6$ kHz. The mask signal, on the other hand, is set as a CW signal with a frequency of $f_m = 20.3$. As a result, the signals used to derive the phase are at frequencies of 1.2 kHz and 1.8 kHz. It is important to note that these frequencies are not unique. We can choose other frequency combinations as long as they meet the requirement of the frequency range mentioned above and Section III-B.

B. Motion Tracking

The recording and processing device used in our system is typically a laptop, which is connected to the receiver BLE earphone. As mentioned in Section III, we have opted to use CW acoustic signals as the sensing signal in our motion tracking system due to the limitations of BLE. Consequently, we have implemented two tracking methods: phase-based distance tracking and strength-based angle tracking.

One major challenge in implementing these methods is the requirement for calibration between the transmitter and receiver. This is necessary due to slight variations in clock frequencies among different devices, which can lead to frequency shifts. To overcome this challenge, a clock calibration system is utilized. Initially, the system searches for a motionless data segment, which is easily identifiable as frequency-induced motion is typically minimal and consistent. This segment is then used to compute the frequency difference between the known playback frequency and the recorded frequency.

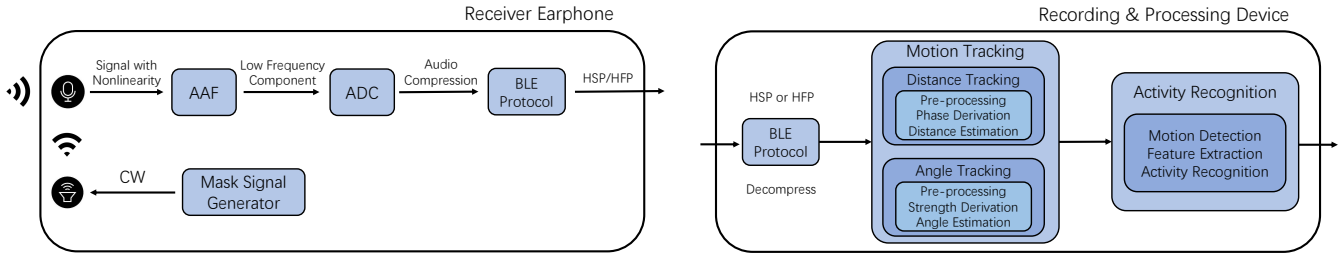
The obtained frequency difference from calibration is subsequently utilized in distance and angle estimation to improve the accuracy of the results. By incorporating the frequency shift, the system can compensate for any discrepancies and achieve more precise tracking. It's important to note that all calibration processes are conducted on the recording and processing device, as illustrated in Fig. 5(b).

The first step in both distance and angle tracking is pre-processing, which involves filtering the received signal using a bandpass filter to remove irrelevant noise. For instance, if we want to derive the distance from f_1 using a mask signal f_m, the central frequency of the bandpass filter would be $|f_1 - f_m|$.

After the pre-processing stage, the BLEAR system utilizes phase-based distance estimation and strength-based angle estimation separately, as mentioned in Section II-B to derive accurate motion tracking results. While it is possible to derive distance and angle at the same rate as the audio sampling rate (typically 8/16 kHz), it is not necessary to maintain such a high derivation rate for our system. Instead, we can average the values and reduce the derivation rate to 50 Hz, which enhances stability and accuracy.

C. Activity Recognition

After estimating motion tracking results, we obtain the results for distance and angle tracking. These measurements



(a) Receiver earphone. (b) Recording and processing device.

Fig. 5. System details.

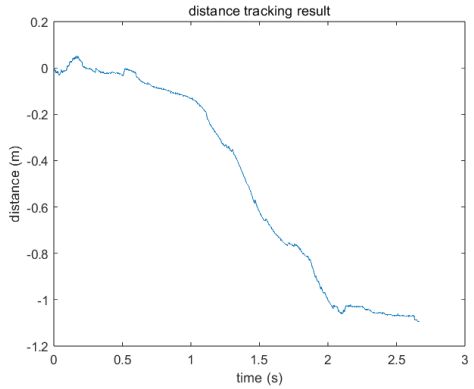


Fig. 6. An example of a person coming back to his computer.

can be utilized to analyze user activity. An example application scenario is the automatic lock and resume feature on a computer, as depicted in Fig. 1. Our objective is for the computer to perceive our actions, such as automatically locking when we leave and resuming when we return.

In addition to leaving and returning, there may be instances where individuals simply stand up without walking away. In such cases, the computer should not lock itself. Similarly, if a user passes by without sitting down, the computer should not resume. To account for these practical scenarios, we have designed seven activities: standing up, sitting down, walking away, walking close, standing up and walking away, walking close and sitting down, and passing by.

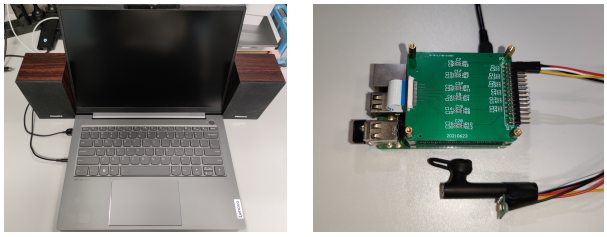
Furthermore, we have identified several significant features that can effectively differentiate between different activities. For instance, in Fig. 6 we observe the distance change when a user wearing a BLE earphone returns to their computer. This scenario can be divided into two stages: walking close and sitting down. The speed at which these stages occur differs significantly, which aids in distinguishing between activities.

After careful consideration, we have opted to employ the K-Nearest Neighbors Algorithm (KNN) as our classification algorithm. It has proven to be both efficient and accurate for this particular scenario. We have selected 12 features to serve as inputs for the KNN algorithm. These features and their explanations are summarized in Table I. By employing the KNN classifier, we can achieve precise activity recognition, thereby facilitating the implementation of practical BLE earphone sensing applications. The performance of our system will be discussed in detail in Section VI.

TABLE I
FEATURES USED IN THE KNN CLASSIFIER.

Feature	Explanation
Direction	Overall motion direction (closer/further).
Length	Distance travelled.
Angle	Angle travelled.
Time	Time consumed.
Speed @ 1/4	Speed at 1/4 of the whole journey.
Speed @ 3/4	Speed at 1/4 of the whole journey.
Acceleration @ 1/5	Acceleration at 1/5 of the whole journey.
Acceleration @ 2/5	Acceleration at 2/5 of the whole journey.
Acceleration @ 3/5	Acceleration at 3/5 of the whole journey.
Acceleration @ 4/5	Acceleration at 4/5 of the whole journey.
Jerk*@ 1/4	Jerk at 1/4 of the whole journey.
Jerk @ 3/4	Jerk at 3/4 of the whole journey.

* The derivative of acceleration.



(a) Transmitter. (b) Receiver.

Fig. 7. Transceivers. (a) Transmitter: laptop & speaker. (b) Receiver: Raspberry Pi & Reapeaker Kit.

V. IMPLEMENTATION

Our system implementation consists of two parts, transmitter and receiver. The transmitter utilizes a basic stereo speaker, while the receiver is implemented using the Respeaker Kit [16]. Further details regarding the implementation of the system will be discussed in the following sections.

A. Transmitter

Our system uses stereo mode to control a pair of speakers, allowing simultaneous playback of two sine waves. One speaker plays a sine wave at frequency f_1 while the other plays a sine wave at frequency f_2. For convenience, we use a pre-generated file in the form of a .wav file. We use a laptop as the speaker controller, primarily for playing back the pre-generated two-channel audio file.

We use the "Philips SPA20" speaker (Fig. 7(a)) in our system. To set up the speaker system, we position the speakers

on either side of a laptop, approximately 40 cm apart. This arrangement ensures a stereo effect, particularly useful when the laptop has downward-facing speakers.

B. Receiver

At the receiver end, we need a PZT to generate mask signal and BLE earphones to receive signal.

A PZT transducer can be directly powered by a voltage signal. We use one channel of a 3.5 mm headphone jack to drive a thin PZT transducer. The central frequency of the PZT that we use is 18 kHz. Thus it has good performance at ultrasound frequency band. The PZT transducer is placed near the microphone to ensure relatively strong signal strength.

For BLE earphones, initially we implemented our system on commercial BLE earphones such as the Apple AirPods2 [17]. We found that we could achieve motion tracking only some of the time. The signal strength varies extensively, even when the earphones are stationary. We surmise that this is caused by denoising algorithms, which sometimes reduce the received signal. Unfortunately, we are not able to disable these algorithms as they are embedded in the firmware and cannot be accessed. Thus, we have implemented a separate BLE earphone without additional audio processing such as denoising using a Raspberry Pi and Respeaker. We then connect a computer to these simulated "BLE earphones" and record the audio data.

As shown in Fig. 7(b), the PZT transducer is attached to the back of a microphone. The microphone, in turn, is attached to a commercial BLE earphone casing. It is important to note that the BLE earphone is non-functional; its purpose is solely to provide a wearable form factor for the microphone.

VI. Evaluation

This section reports our extensive experiments that evaluate BLEAR's feasibility. We first give an overview of the core evaluation results of this work. Then we report the detailed experiments and findings.

A. Overview

We conduct two series of experiments to extensively evaluate the performance of BLEAR in a real-world setting. In the first evaluation, we assess BLEAR's performance in motion tracking regarding angle and distance estimation. The result shows that BLEAR can achieve errors of 3.37 cm and 5.3 degrees in distance and angle estimation, respectively. In the second evaluation, we design a classifier to recognize the earphone wearer's daily activities, and the result shows that BLEAR can recognize seven common activities with an accuracy of 97.14%.

B. Motion Tracking Performance

In this section, we evaluate BLEAR's performance in estimating two separate motion metrics, namely, distance estimation and angle estimation. For these tests, the speakers and earphone are placed on the ground, ensuring they are at the same height and get the best performance.

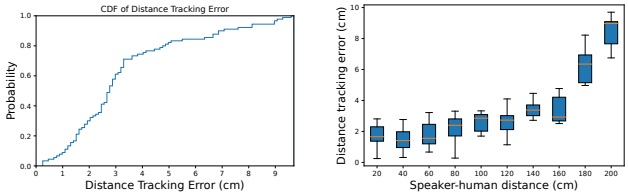
1) Distance Tracking Error: To accurately measure the distance tracking error respective to the anchor speakers, we use a linear actuator with a stepper motor to control the movement of the earphone. This linear actuator can be programmed to follow a predefined movement pattern, offering a high level of accuracy at the millimeter scale. Consequently, it serves as the ground truth for changes in distance. The errors are calculated as the absolute differences between the measured distance changes and the corresponding estimated values. While conducting the experiments, we align the linear actuator with the perpendicular bisector of the two anchor speakers to ensure consistency. To ensure reliability, distance measurements are repeated 9 times at each distance. Fig. 8 shows the experimental results. Specifically, Fig. 8(b) shows the distance estimation errors at different distances. We find the result is quite encouraging that the mean error is within 3.37 cm and, notably, when the earphone is within 1.5 m with respect to the anchors, the mean error is as low as 2.24 cm.

While we do not offer a quantitative comparison, it is worth noting that the similarly designed system in [12], which however does not incorporate BLE, also achieves cm-level distance tracking error. Instead, the CAT system [8], which also does not incorporate BLE, by incorporating synchronization between transceivers, can achieve accuracy up to 5 mm, surpassing the performance of BLEAR. Thus, even though BLEAR incorporates BLE, it can achieve comparable accuracy compared with normal acoustic motion tracking systems.

2) Angle Tracking Error: To assess the angle tracking performance, we let the earphone move in an arc-shaped motion within a fan-shaped area in front of the speaker, with the speaker as the center. The error is measured as the difference between the estimated angle and the established ground truth. To establish the ground truth angle measurements, designated points are marked on the ground. The earphone, held by a human, then transits between these points, allowing measurement of angle changes as the ground truth. To ensure reliability, these measurements are also repeated 9 times. It is worth noting that the ground truth angles are taken at different distances in order to show how the distance between the earphone and the anchors affects the angular accuracy. Fig. 9 shows the angle estimation error statistics. Specifically, Fig. 9(b) shows the relationship between angle estimation error and distance. It is observed that within a particular range, our system achieves an extremely low mean estimation error of 1.9 degrees. Nevertheless, this error increases up to 11.9 degrees as the distance continues to increase. The overall estimation error in the range of 2 m is 5.3 degrees.

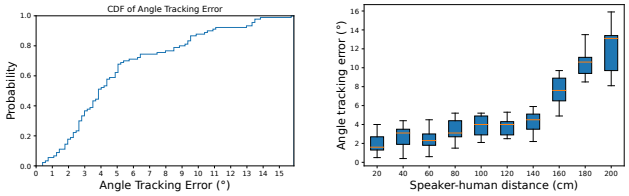
C. Activity Recognition Performance

In addition to separately tracking the raw earphone's angle and distance, we design a classifier that can predict the earphone-wearer's activity based on the sensed distance and angle changing patterns as described in Section IV. In this section, we evaluate the performance of BLEAR's activity recognition accuracy. The speakers are placed beside a laptop on the desk and the earphone is worn by the subjects. We



(a) Error CDF. (b) Error vs. distance

Fig. 8. Distance estimation error.



(a) Error CDF. (b) Error vs. distance

Fig. 9. Angle estimation error.

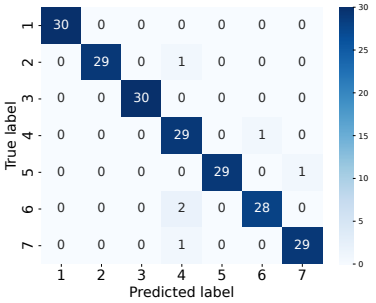
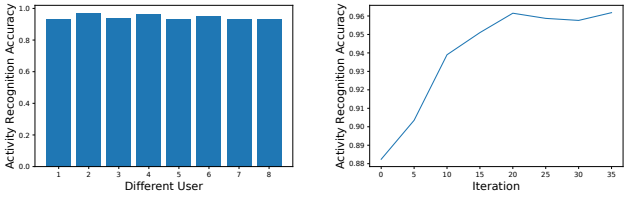
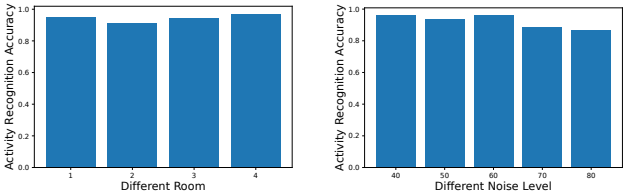


Fig. 10. Activity recognition confusion matrix.



(a) Impact of users. (b) Impact of scenarios.

Fig. 11. Activity recognition accuracy with different users and iterations.



(a) Impact of rooms. (b) Impact of noise.

Fig. 12. Activity recognition accuracy with different factors.

have recruited 8 subjects to participate in this experiment. For each subject, we ask him/her to perform the seven target activities (labeled from 1 to 7), namely, standing up, sitting down, walking away, walking close, standing up and walking away, walking close and sitting down, and passing by, and each subject is asked to repeat the above activities for at least five rounds. In total, we have collected 210 samples. As discussed in Section IV, the prediction of these activities is accomplished using a KNN classifier. We use leave-one-subject-out (LOSO) cross-validation to evaluate the classification performance. Fig. 10 shows the confusion matrix of the classification result. The precision, recall and F1-score are 97.30%, 97.14% and 97.18%. In general, all the activities can be distinguished from the others with high accuracy. Note that 'sit down' could be misclassified into 'walk close' in some cases, and similarly, some 'pass by' samples are misclassified into "walk close". This is because these activities are similar in nature, and it can be hard to distinguish them.

Additionally, we conducted an evaluation of several impact factors that could potentially affect the accuracy of activity recognition. The results, which illustrate the activity recognition accuracy for different users, are presented in Fig. 11(a). The accuracy ranges from 88.2% to 97.1% with the original data. It is worth noting that each user has unique walking and sitting patterns, leading to variations in accuracy. To address this, newly recognized data is incorporated into the training dataset, enabling the optimization of recognition parameters of each user, just as Fig. 11(b) shows. As a result, an enhanced accuracy of over 93% is achieved for all users.

Fig. 12(a) illustrates the activity recognition accuracy in different rooms. These rooms are: 1. a 6m x 10m meeting room with desks and chairs; 2. a 5m x 5m semi-open space; 3. a 2m x 6m room with many of objects inside; 4. a 10m x 20m office space. These rooms have different sizes and item placements, resulting in different acoustic interference. The results indicate consistently good accuracy across all rooms, with accuracy consistently above 90%. In Fig. 12(b), we present the accuracy of recognition under the influence of noise at different levels. The noise is played by a speaker and comprises multiple types of sounds, such as music, speech, and machine-generated noise. The final accuracy is an average of multiple tests. However, due to the similarity in frequency bands employed by BLEAR and ambient noise, simply applying a filter is insufficient to solve the noise problem. The accuracy decreases to 40% with noise at 80 dB. To address this, we integrated noise-affected data into the training dataset, resulting in an accuracy of 88% at 80 dB of noise. Although slightly lower than the accuracy without noise, this level of accuracy is still acceptable considering the target scenario of a normal office environment.

VII. Discussion and Limitations

a) Limitations: Although this is the first work to achieve earphone tracking under the BLE audio recording protocol, there are limitations in our current design and implementation. **Tracking distance**. We use the microphone's nonlinearity to convert the high-frequency beacon signal to the low-frequency band. However, the signal power after this conversion is largely reduced because of the nature of the

nonlinear effect. Therefore, the effective sensing range is limited to 150 cm. Future designs may use speakers with larger power to extend the sensing range. **Beacon design**. As discussed in Section II, the BLE employs an audio compression strategy to reduce data traffic which will, in return, damage the continuity of the bandwidth. To bypass the influence of audio compression, we carefully select 3 frequencies and use single-frequency continuous wave (CW) signals as the beacon signal. However, previous works have shown that wide-band signals, such as FMCW, have better tracking ability. Therefore, future work should also explore how to design wide-band signals that are still functional after BLE's audio compression. **Noise**. BLEAR makes use of the acoustic signal within the frequency range of 0.5 - 2 kHz (after non-linear effects). Unfortunately, this range overlaps with human voice and ambient noise commonly encountered in daily life. Loud noises can cause tracking failure and result in low recognition accuracy. While we have taken speech noise into consideration during the evaluation, we are unable to address continuous loud noises. BLEAR has the potential to utilize frequencies up to 7 kHz if speakers and microphones with a good frequency response near 24 kHz are available.

b) Compatibility: Although we use a Raspberry Pi to build our prototype, it is only used to simulate a BLE earphone and does not apply any processing to the audio signal. Thus, BLEAR can be easily deployed on a normal BLE earphone, provided that the earphone does not process the audio. Since our system is based on the BLE audio protocol, it is compatible with audio applications such as phone calls. However, the user's voice during a phone call can interfere with the motion tracking signals in our system, potentially decreasing performance.

VIII. Related Works

In this section, we give an overview of the related work of this study. We group the related works into two categories: motion sensing systems and acoustic sensing systems.

A. Motion Sensing

Research in motion sensing is most related to this study. We summarize the related works according to their sensing modality. **Computer vision**. Computer vision (CV) methods are most prevalent for human motion capture. Some works [18]–[20] design methods to extract 3D human motion from videos of RGB or depth camera. **Bluetooth power**. Recent commercial products and research employ Bluetooth Low Energy (BLE) Received Signal Strength (RSS) measurements to estimate the approximate proximity between two devices [2], [3]. **IMU**. Inertial measurement units, including accelerometer, gyroscope and magnetometer, can be used to track motions. In recent years, there has been research focused on pedestrian dead-reckoning (PDR) [21], [22]. Alternatively, some works [5], [23] use IMUs attached to headgear for head tracking and detecting user attention. A number of eye-gaze tracking systems [24], [25] have also adopted IMU-based compensatory measures for head movement to enhance tracking precision.

B. Acoustic Sensing

The research community is particularly interested in leveraging acoustic signals to realize various kinds of sensing tasks due to the wide accessibility of this signal. Here, we summarize these works by their applications. **Motion trakcing**. The solutions proposed in some works [8], [9], [12], [15], [26] are device tracking systems that employ multiple speakers to track a device that is equipped with a microphone. Apart from device tracking, researchers are also interested in device-free tracking designs. Recent studies [14], [27]–[29] proposed motion-tracking solutions that can track hand, finger or human body movements. These designs derive the motion of the target object by analyzing the acoustic signals that are reflected by the target object. **Gesture recognition**. AudioGest [30] and RobuCIR [31] utilize a speaker-microphone pair to detect alterations in Doppler shift or channel impulse response (CIR) caused by hand movements to recognize different hand gestures. **Sensing on earables**. Faceori [1] uses an acoustic ranging method that involves the microphones and the headphone and speakers on anchor devices to achieve user orientation detection. EarphoneTrack [13] proposes an earphone tracking system that overcomes practical challenges inherited from the earphone's form factor. EHTrack [10] presents an earphone tracking system that can track the wearer's location and head orientation simultaneously. **Nonlinearity**. The feasibility of using common off-the-shelf (COTS) microphones to record over high-frequency components using nonlinearity is first reported in Chen et al. [32]. They demonstrate that using a 48 kHz sampling rate microphone can sense the power spectrum density (PSD) of signals at 60 kHz, which is over twice the conventional sensible threshold.

IX. Conclusion

In conclusion, existing methods utilizing acoustic signals for earphone tracking are infeasible for wireless earphones, primarily due to their reliance on the Bluetooth Low Energy (BLE) protocol for audio data transmission. BLE uses a low audio sampling rate and implements audio compression, which makes the existing solutions impossible to deploy. This study introduces BLEAR to overcome these challenges, presenting the first BLE-compatible earphone tracking solution. Through innovative approaches involving a nonlinearity-triggering piezoelectric transducer and strategically designed beacon signals, BLEAR enables wireless earphone tracking while adhering to BLE protocol restrictions. A wireless earphone prototype is implemented, and extensive experiments involving 8 subjects are conducted to showcase BLEAR's practicality. The experimental results demonstrate that, in the range of 2 m, BLEAR can achieve mean angle error of 5.3 degrees and distance tracking error of 3.37 cm. Moreover, an accuracy of 97.14% in recognizing seven common user activities can be achieved.

REFERENCES

[1] 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 *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–12.

[2] "Google pixel watch," 2023, accessed Aug 7, 2023. [Online]. Available: https://store.google.com/us/product/google_pixel_watch

[3] O. Hashem, K. Alkiek, M. Youssef, and K. A. Harras, "Leveraging earables for natural calibration-free multi-device identification in smart environments," in *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, 2021, pp. 92–98.

[4] "Airpods pro (2nd generation) - apple," 2023, accessed Aug 7, 2023. [Online]. Available: https://www.apple.com/airpods-pro/

[5] Z. Yang, Y.-L. Wei, S. Shen, and R. R. Choudhury, "Ear-ar: indoor acoustic augmented reality on earphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.

[6] A. Ahuja, A. Ferlini, and C. Mascolo, "Pilotear: Enabling in-ear inertial navigation," in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 139–145.

[7] J. Gong, X. Zhang, Y. Huang, J. Ren, and Y. Zhang, "Robust inertial motion tracking through deep sensor fusion across smart earbuds and smartphone," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–26, 2021.

[8] 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.

[9] 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.

[10] L. Ge, Q. Zhang, J. Zhang, and H. Chen, "Ehtrack: Earphone-based head tracking via only acoustic signals," *IEEE Internet of Things Journal*, 2023.

[11] Y. Zhang, J. Wang, W. Wang, Z. Wang, and Y. Liu, "Vernier: Accurate and fast acoustic motion tracking using mobile devices," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1709–1717.

[12] L. Ge, Q. Zhang, J. Zhang, and Q. Huang, "Acoustic strength-based motion tracking," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–19, 2020.

[13] 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.

[14] 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.

[15] Y. Liu, W. Zhang, Y. Yang, W. Fang, F. Qin, and X. Dai, "Pamt: Phase-based acoustic motion tracking in multipath fading environments," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2386–2394.

[16] "Respeaker 4-mic linear array kit — seeed studio wiki," https://wiki.seeedstudio.com/ReSpeaker_4-Mic_Linear_Array_Kit_for_Raspberry_Pi/, 2023, accessed Aug 7, 2023.

[17] "Airpods (2nd generation) - apple," 2023, accessed Aug 7, 2023. [Online]. Available: https://www.apple.com/airpods-2nd-generation/

[18] K. Wang, J. Xie, G. Zhang, L. Liu, and J. Yang, "Sequential 3d human pose and shape estimation from point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7275–7284.

[19] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, "Vitpose: Simple vision transformer baselines for human pose estimation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 38 571–38 584, 2022.

[20] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7753–7762.

[21] W. Kang and Y. Han, "Smartpdr: Smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors journal*, vol. 15, no. 5, pp. 2906–2916, 2014.

[22] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu," in *2009 IEEE International Symposium on Intelligent Signal Processing*. IEEE, 2009, pp. 37–42.

[23] J. Windau and L. Itti, "Walking compass with head-mounted imu sensor," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 5542–5547.

[24] C.-H. Fang and C.-P. Fan, "Effective marker and imu based calibration for head movement compensation of wearable gaze tracking," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2019, pp. 1–2.

[25] T.-L. Liu and C.-P. Fan, "Visible-light wearable eye gaze tracking by gradients-based eye center location and head movement compensation with imu," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2018, pp. 1–2.

[26] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, "Indoor follow me drone," in *Proceedings of the 15th annual international conference on mobile systems, applications, and services*, 2017, pp. 345–358.

[27] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 1515–1525.

[28] 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.

[29] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, "Strata: Fine-grained acoustic-based device-free tracking," in *Proceedings of the 15th annual international conference on mobile systems, applications, and services*, 2017, pp. 15–28.

[30] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, "Audiogest: enabling fine-grained hand gesture detection by decoding echo signal," in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, 2016, pp. 474–485.

[31] Y. Wang, J. Shen, and Y. Zheng, "Push the limit of acoustic gesture recognition," *IEEE Transactions on Mobile Computing*, vol. 21, no. 5, pp. 1798–1811, 2020.

[32] Y. Chen, W. Gong, J. Liu, and Y. Cui, "I can hear more: Pushing the limit of ultrasound sensing on off-the-shelf mobile devices," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2015–2023.