



MagSound: Magnetic Field Assisted Wireless Earphone Tracking

LIHAO WANG, Nanjing University, China

WEI WANG, Nanjing University, China

HAIPENG DAI*, Nanjing University, China

SHIZHE LIU, Nanjing University, China

Wireless earphones are pervasive acoustic sensing platforms that can be used for many applications such as motion tracking and handwriting input. However, wireless earphones suffer clock offset between the connected smart devices, which would accumulate error rapidly over time. Moreover, compared with smartphone and voice assistants, the acoustic signal transmitted by wireless earphone is much weaker due to the poor frequency response. In this paper, we propose MagSound, which uses the built-in magnets to improve the tracking and acoustic sensing performance of Commercial-Off-The-Shelf (COTS) earphones. Leveraging magnetic field strength, MagSound can predict the position of wireless earphones free from clock offset, which can be used to re-calibrate the acoustic tracking. Further, the fusion of the two modalities mitigates the accumulated clock offset and multipath effect. Besides, to increase the robustness to noise, MagSound employs finely designed Orthogonal Frequency-Division Multiplexing (OFDM) ranging signals. We implement a prototype of MagSound on COTS and perform experiments for tracking and handwriting input. Results demonstrate that MagSound maintains millimeter-level error in 2D tracking, and improves the handwriting recognition accuracy by 49.81%. We believe that MagSound can contribute to practical applications of wireless earphones-based sensing.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Acoustic sensing, Distributed ranging, Earable computing

ACM Reference Format:

Lihao Wang, Wei Wang, Haipeng Dai, and Shizhe Liu. 2023. MagSound: Magnetic Field Assisted Wireless Earphone Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 33 (March 2023), 32 pages. <https://doi.org/10.1145/3580889>

1 INTRODUCTION

Acoustic ranging has been a prosperous technique in human-computer interaction on mobile devices. Compared with radio frequency signals, acoustic signals have a much smaller signal propagation speed and thus are competent for fine-grained tracking tasks [46, 62, 65, 72]. Most of the existing works implement acoustic ranging systems on smartphones due to their pervasive usage. Recently, many researchers observe that wireless earphones have been widely used in daily life [12]. Connected via Bluetooth, the movement of an earphone can be tracked by smart devices, such as smartphones, and hence this gadget can be a novel interaction tool. Considering its compact size, holding a wireless earphone is more user-friendly than holding a smartphone for tracking task. Besides, for smart devices without or with a small screen, e.g., smart speakers or smartwatches, earphones can

*Haipeng Dai is the corresponding author.

Authors' addresses: Lihao Wang, Lihao_Wang@hotmail.com, Nanjing University, Nanjing, Jiangsu, China, 210023; Wei Wang, ww@nju.edu.cn, Nanjing University, Nanjing, Jiangsu, China, 210023; Haipeng Dai, haipengdai@nju.edu.cn, Nanjing University, Nanjing, Jiangsu, China, 210023; Shizhe Liu, shizhelu@smail.nju.edu.cn, Nanjing University, Nanjing, Jiangsu, China, 210023.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2474-9567/2023/3-ART33 \$15.00

<https://doi.org/10.1145/3580889>

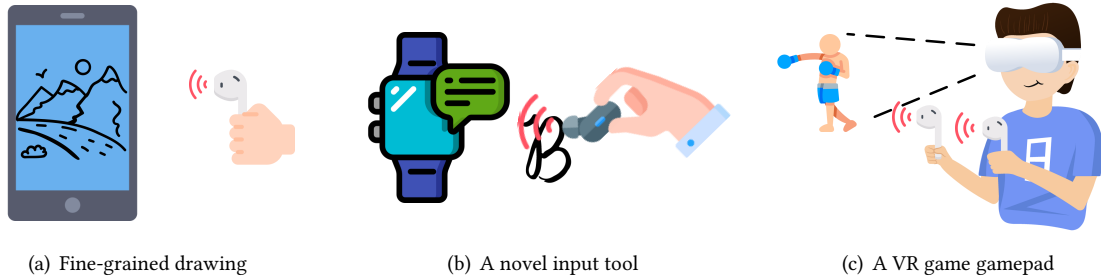


Fig. 1. Typical applications of wireless earphone tracking

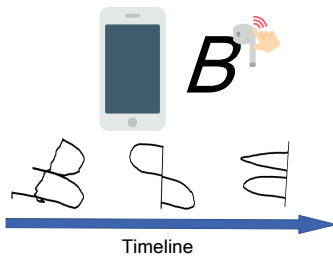


Fig. 2. One-shot calibration example

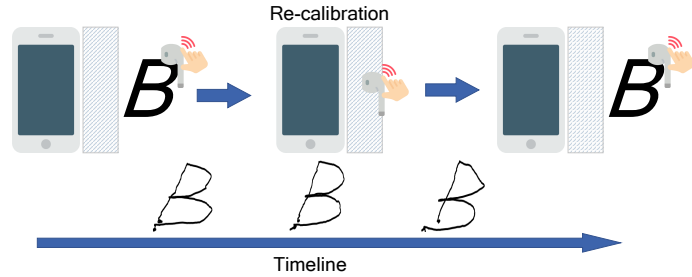


Fig. 3. MagSound example

be an input tool that does not rely on the touch screen, which is more convenient than other interaction tools such as stylus. Figure 1 shows three typical applications of wireless earphone tracking. In Figure 1(a), wireless earphones can be used to do fine-grained drawing as an alternative to styluses. In Figure 1(b), wireless earphone tracking empowers the earphone to be a novel input tool for a small screen without contact, which avoids the fat finger problem [56]. In Figure 1(c), wireless earphones become gamepads for Virtual Reality (VR) games. We believe that the wireless earphone has the potential to be a ubiquitous acoustic sensing platform.

Albeit promising, tracking the wireless earphones is non-trivial. We find that there are two main challenges faced by wireless earphone tracking. The first challenge is that since the smart device and the wireless earphone have their own clocks, the asynchronous clock would accumulate error rapidly for acoustic ranging, which is formidable for fine-grained tracking. Previous works [62, 66] for acoustic ranging systems view the offset of each signal frame as a constant, so they perform one-shot calibration linearly, *i.e.*, holding the device for seconds before tracking movement to count the clock offset and then cancel it using the average of the clock offset hereafter. Unfortunately, for wireless earphones, we observe that the clock offset per frame varies over time, and hence the one-shot calibration is not applicable to the considered scenario. According to our experiment, using the one-shot calibration, the distance error caused by the clock offset can be up to 1.43 m in 240 seconds, given that the sound speed is 343 m/s. The second challenge is that the signal transmitted by the earphone is much weaker than the built-in speaker in the smartphone, and some earphones even fail to transmit acoustic signal at ultrasound band [7]. While it brings less discomfort to users even in the audible band, a weaker acoustic signal is more susceptible to noise and the tracking accuracy suffers as a result.

To deal with the challenges, we propose MagSound, which can calibrate the clock offset multiple times nonlinearly in one run. The key insight is that the built-in magnetometer in the smart device shares the clock with built-in microphones, which means that the collected magnetic field strength data is not affected by the asynchronous clock. Besides, we find that the earphones use embedded magnets to produce sound, so we can exploit the strength of the magnetic field generated by the magnets to predict the position of the earphone free from clock offset, which enables us to estimate the offset caused by acoustic ranging and then cancel it. The advantage of this method is that the predicted position is neither affected by clock offset nor multipath, and the disadvantage is that the effective range is limited due to the property of the magnetic field. To drive our point, we draw two examples of one-shot calibration and MagSound in Figure 2 and Figure 3, respectively. In the examples, the user is writing the letter ‘B’ by tracking the movement of the wireless earphone. Using one-shot calibration, the writing trace of ‘B’ gets distorted over time due to the accumulated clock offset. In contrast, using MagSound, the user can re-calibrate the acoustic ranging system by moving the earphone into the shaded area where we predict the position using magnetic field strength. In this area, users do not need to keep the earphone still, and experimental results show that the duration can be small enough to be negligible. Meanwhile, MagSound collects the clock offset data to fit it using an explicit function. Once the earphone moves out of the area, MagSound predicts the clock offset using the fitted function and cancels the clock offset to correct the trace. As a result, we can see that trace ‘B’ keeps stable over time. Besides, to achieve accurate sensing under weaker signals, we design a robust signal model to transmit the Zadoff-Chu (ZC) sequence in an Orthogonal Frequency-Division Multiplexing (OFDM) manner. We mitigate the noise interference through signal correlation, and the nice properties of the ZC sequence allow us to separate signals propagating through Line-of-Sight (LOS) paths from multipath.

In a nutshell, we summarize our contribution into four points:

- (1) We perform integrated magnetic and acoustic sensing. Based on the pros and cons of magnetic and acoustic sensing, we carefully design a fusion algorithm to integrate the two modalities.
- (2) To the best of our knowledge, we are the first to make on-the-fly clock offset corrections for wireless earphones. Consequently, the performance of MagSound is stable over time, which pushes the earphone tracking system into real-life applications.
- (3) We design a robust signal model for earphone sensing. We use ZC sequence as the acoustic signal for the earphone tracking system to mitigate problems, such as weaker transmitted signals, brought by the unique characteristics of earphones.
- (4) We implement a prototype of MagSound to evaluate the performance for potential applications on Commercial Off-The-Shelf (COTS) devices, such as motion tracking and handwriting input.
- (5) Experimental results show that MagSound maintains millimeter-level accuracy for 2D tracking under clock drift and achieves a recognition accuracy of 95.43% for handwriting, which improves the baseline by 49.81%. User evaluation shows that 93.75% of participants prefer MagSound over the touch screen based input method on smartphones.

2 RELATED WORK

Existing works related to earphone tracking can be divided into three categories: sound-based sensing, magnet-based tracking, and Inertial Measurement Unit (IMU) based tracking.

2.1 Sound-based Sensing

Techniques on sound-based sensing have developed rapidly in recent years, which can be further divided into two categories, *i.e.*, device-free sensing and device-based sensing, depending on whether a device is needed on the object being sensed.

For device-free sensing [11, 41, 43, 47, 58, 63, 65, 70], although circumvent the complicated synchronization, they are more vulnerable to noise and multipath effect than device-based system. Recently, VSkin [58] and StruGesture [63] track the finger gesture on the back of a smartphone by distinguishing the structure-borne sounds. They transmit a novel signal called Zadoff-Chu (ZC) sequence to improve robustness to noise.

For device-based sensing [22, 46, 61, 62, 72], CAT [46] tracks the movement of a smartphone in millimeter-level tracking accuracy using a distributed Frequency Modulated Continuous Wave (FMCW) ranging system, where the transmitter (speaker) and the receiver (smartphone) are separate and asynchronous. CAT uses the smartphone built-in IMU to detect whether the smartphone is stationary, and thereby makes re-calibration on the run. SoundTrack [72] devises a ring with an embedded miniature speaker to transmit an acoustic signal, employs an array of microphones to receive the signal, and localizes the finger's 3D position afterwards. MilliSonic [62] achieves sub-millimeter tracking accuracy based on the phase of FMCW signal. ASAT [22] performs motion tracking by sensing the sound strength and achieves centimeter-level accuracy. VECTOR [61] monitors temperature-field based on the sound speed with distributed acoustic devices.

Moreover, earphones have become a novel platform for device-based sensing with the development of earable computing [12]. EarphoneTrack [7] applies wired earphones to 2D tracking and device-free scenario, achieving millimeter-level accuracy. Between wireless earphone and smartphone, EarphoneTrack only presents the result of 1D tracking. Liu *et al.*, [45] proposed an application called acoustic ruler, where the smartphone receives FMCW acoustic signal transmitted by the earphone. Limited by Bluetooth microphones, the acoustic ruler employs the audible spectrum (< 10 kHz) for FMCW signal to measure the 1D distance between the smartphone and the earphone, and achieves a 1.7 cm error within 5 meters. FaceOri [66] modifies the earphone to get the raw acoustic stream of built-in microphones and tracks the face movement using the headset as the receiver and a smartphone as the transmitter. EarIO [40] tracks facial expressions in an AI-powered way. It transmits acoustic signals on a wired earable towards the face, and inputs the received signal into a customized deep learning pipeline, which predicts facial expressions. EarHealth [34] uses wired earphones to monitor hearing health by analyzing the recorded echoes evoked by a chirp sound stimulus in ear.

To conclude, for earphone-based tracking systems, the tracking accuracy of the state of the arts is millimeter-level. Besides, earphones have shown the potential to be a ubiquitous sensing platform. For device-based systems, the difficulty lies in the synchronization between the transmitter and the receiver. Most related works assume that the offset caused by the asynchronous clock is constant, and stall the device for a few seconds to perform linear calibration before tracking. For specified hardware or smartphones the linear calibration may work, yet for COTS wireless earphones, whose oscillators are smaller, the one-shot calibration would fail. We will further discuss the clock offset problem in Section 3.2.2.

2.2 Magnet-based Tracking

Magnet-based tracking uses either permanent magnet or electromagnet. For permanent magnets, MagPen [30] augments a pen with magnets, and recognizes the relative gesture based on the magnetic field strength directly. TMotion [69] fuses the data from both magnetometer and gyroscope to track the movement of a pen, and achieves millimeter-level error in 3D. MagX [9] utilizes an array of magnetometers to track the wearable permanent magnet with robustness and high-efficiency. MAGIC [64] proposes a system framework for automatic calibration of soft-iron and hard-iron disturbances of MEMS magnetometer arrays. For electromagnet, Finexus [10] equips fingertips with electromagnets and performs fine-grained finger tracking with an average accuracy of 1.33 mm in 3D. AuraRing [49] designs an untethered ring that carries an electromagnet, and requires fewer sensors compared with Finexus. Huang and Wu [29] pushed the error of electromagnet-based tracking into sub-millimeter level.

However, all the related works above require either extra permanent magnets or electromagnets. Permanent magnet-based tracking requires multiple magnetometers to track the magnet. Besides, the distance is limited

between the magnet and the magnetometer [9]. On one hand, to satisfy the assumption of dipole, the distance should not be too small. On the other hand, due to the rapid decrease of the magnetic field strength, the ranging distance should not be too large. As we make no modifications to COTS smartphone and earphone, their methods are not applicable to the considered scenario.

2.3 IMU-based Tracking

In the field of motion tracking, IMU is another important sensor to sense, which includes gyroscope, accelerometer, and compass. In this subsection, we mainly focus on accelerometer and gyroscope.

WriteAS [73] inputs the IMU data from a smartwatch to train a multimodal convolutional neural network for handwriting recognition task at word-level. VibWriter [21] inputs the vibration signals collected by an accelerometer in the smartphone to train a convolutional neural network for the letter recognition task. SoM [75] tracks the wrist using the built-in IMU of a smartwatch. It requires a fixed smartphone that transmits acoustic tones at intervals to get the radial displacement of the smartwatch as an auxiliary to IMU-base tracking. ITrackU [8] fuses data from ultra-wideband radios (UWB) and IMU to track a pen-like instrument. Handwriting-Assistant [6] designs a pen cap with IMU to track the movement of a pen at millimeter-level accuracy.

Nevertheless, most COTS earphones do not support IMU sensor, or have an IMU but hard to acquire data by user interface. Therefore, existing works to sense on the earphone platform usually have to add specified IMU sensors to the earphone [44, 68]. Nokia Bell lab once launched an experimental product that carries IMU called eSense [36], but the product has reached its end of life according to the official announcement [37]. Besides, distributed IMU sensing also suffers from synchronization [75].

3 OPPORTUNITY AND LIMITATIONS OF WIRELESS EARPHONE SENSING

Compared with other commercial sound emitting and recording devices, such as mobile phones and voice assistants, the earphones have unique characteristics, which bring new opportunities as well as challenges when they are used for sensing. In this section, we first discuss the opportunities of magnetic sensing and then the limitations of sound-based sensing for earphones.

3.1 Opportunities of Magnetic Sensing on Earphones

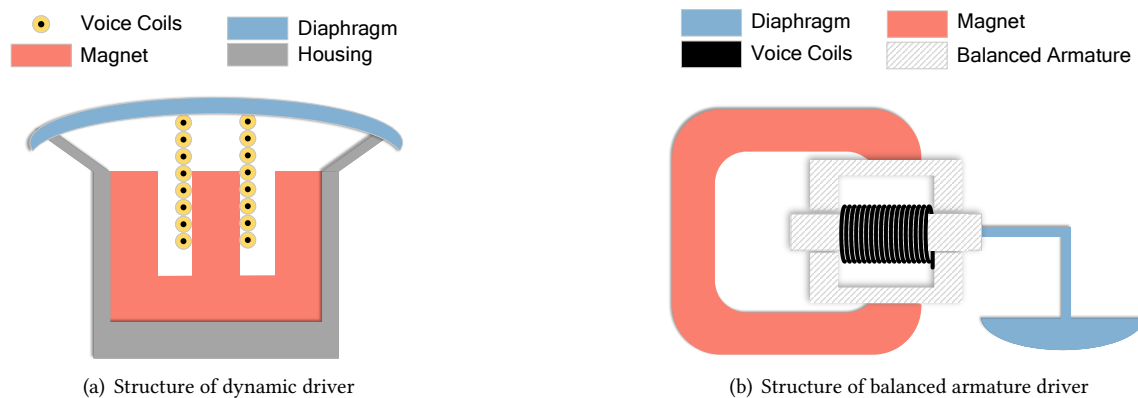


Fig. 4. Two major earphone structures

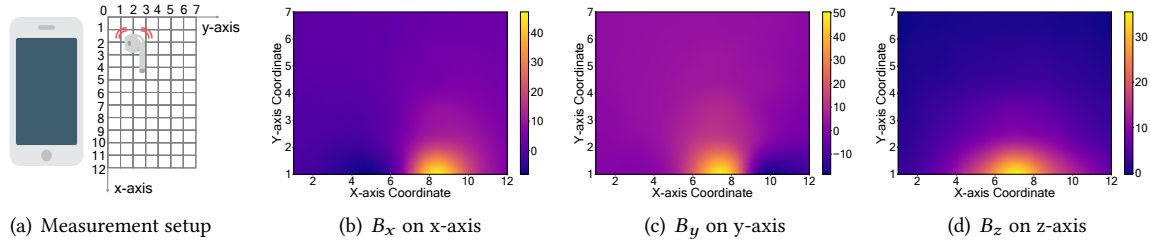


Fig. 5. Magnetic field strength B (μT) measured at the smartphone under different earphone positions

Most earphones have integrated magnet components which produce a magnetic field. To understand the properties of the magnetic field of the earphone, we first provide a brief introduction to the structure of the earphone. In general, earphones use headphone drivers to vibrate and *drive* air into the ear canal. There are two major types of earphone drivers: dynamic driver and balanced armature driver [4], as illustrated in Figure 4. A dynamic driver consists of three core components: diaphragm, voice coils, and magnet. To produce sound, the corresponding current signal runs through voice coils, which then become electromagnetic and repel or attract alternately with magnets. The diaphragm would vibrate with the movement of voice coils since they are connected, and thus moves the air to make the sound. The balanced armature driver also uses the principle of electromagnetic induction. However, in balanced armature drivers, voice coils warp a balanced armature which would become an electromagnet when the current signal runs through voice coils. The electromagnetic balanced armature would move periodically and stimulate the vibration of the diaphragm. Most earphones use one of the two structures, or a mix of both structures. Generally, both types of earphone drivers use strong magnets to convert electrical energy into acoustic energy. To improve performance, an earphone often uses multiple drivers, which means there are multiple magnets in a single earphone. The magnetic field strength around the magnet used in earphones can reach 1.45 T [67], which is at the same level as magnetic resonance imaging. However, strong magnets may cause health problems for long-term users [39]. Therefore, most earphones use careful designed structure, such as electromagnetic shielding, to reduce the magnetic field strength to conform with the WHO standard [48]. For COTS earphones, the strength of static magnetic field is lower than 1 mT outside the earphone [35]. Meanwhile, the time-varying magnetic field is reduced to below $0.1 \mu T$ [42], which is negligible compared with the static magnetic field due to the accuracy of the built-in magnetometer in smart devices [2].

To study the feasibility of using magnetic field to assist earphone tracking, we measure the magnetic field strength B of the earphone using the smartphone's magnetometer. The coordinate system used in the measurement is shown in Figure 5(a), where we measure the reading of the magnetometer when the earphone is placed at different grid points separated by one centimeter. The measurement device is a Google Pixel 4 smartphone and the earphone is a Samsung Galaxy Buds Pro. The measurement of different components of the magnetic field are presented in Figures 5(b), 5(c), and 5(d). We observe that the distributions of B_x , B_y and B_z are different, with a maximum magnetic field strength less than $50 \mu T$. The highest magnetic readings for all magnetic field components appear at around the point (8, 1), where the magnetometer of the Google Pixel 4 is located. Moreover, as the static magnetic field decays with the distance d with a rate proportional to d^{-3} , the field is barely detectable when the distance between the earphone and the magnetometer is more than 10 cm.

In summary, while the magnetic field of the earphone is weak, a commercial smartphone can still reliably detect the earphone when it is close. In addition, the distribution of the three components of the magnetic field is closely related to the position of the earphone so that they can serve as fingerprints to localize the earphone.

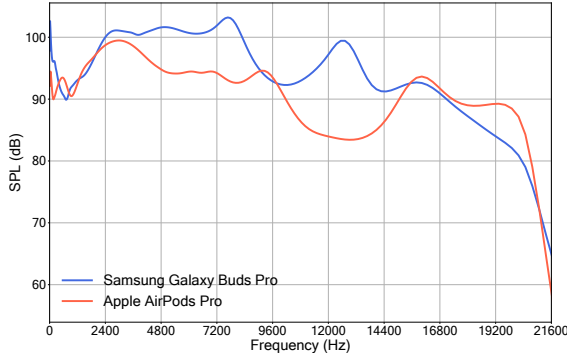


Fig. 6. Frequency response of two types of earphones

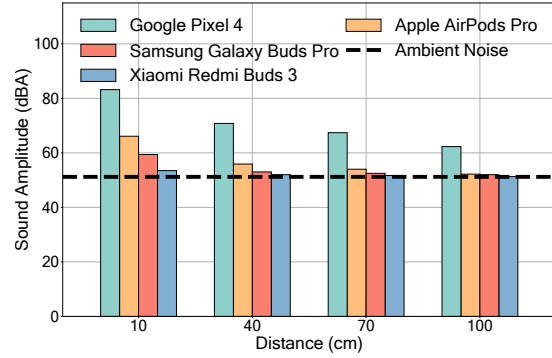


Fig. 7. Signal strength of different speakers

3.2 Limitations of Sound-base Sensing on Earphones

There are two key limitations of the sound signal produced by the earphone, when compared with the speakers on smartphones or voice assistants.

3.2.1 Poor Frequency Response. The first limitation is the poor frequency response of the earphone. Frequency response is a quantitative metric to evaluate the magnitude of the sound output as a function of input frequency [57]. Most earphones support a frequency band from 20 Hz to 20 kHz, which is also generally accepted as the audible frequency range. However, limited by their small shape, earphones focus on boosting the frequency response below 10 kHz, while ignoring the high frequency bands around 20kHz that are widely used for sensing. Figure 6 plots the frequency response curve of Samsung Galaxy Buds Pro and Apple AirPods Pro, using the data from RTINGS [32, 53]. The data from RTINGS are measured when the earphones are worn by a head simulator. The unit used for frequency response is Sound Pressure Level (SPL), which is defined as

$$SPL = 20 \log_{10} \left(\frac{\mathcal{P}}{\mathcal{P}_0} \right), \quad (1)$$

where \mathcal{P} is the stimulated sound pressure, \mathcal{P}_0 is the constant reference pressure. In brief, SPL is proportional to the amplitude of the measured signal in decibels. We observe that the frequency response fluctuates beyond 10kHz. Moreover, the frequency response drops significantly when the frequency is over 16 kHz.

To investigate the impact of distance on the signal strength, we place the speaker at different distances to a sound level meter. We set the volume to the maximum value. Then on the speaker, we play the ZC sequence, which is an acoustic signal with a center frequency at 14 kHz, and profile the signal strength in unit of dBA. We perform the experiments on a smartphone Google Pixel 4 [15] and three different earphones, including Apple AirPods Pro [3], Samsung Galaxy Buds Pro [18], and Xiaomi Redmi Buds 3 [19]. As a reference, we also profile the ambient noise level during the measurement. The results are drawn in Figure 7. We can see that the signal strength of all earphones are significantly weaker than smartphones. Worse still, when the distance is more than 10 cm, the strength of the signal transmitted by earphones is close to the ambient noise, which means a much lower signal-to-noise ratio (SNR) than the smartphone.

As the frequency response determines the amplitude of emitted signals, poor frequency response may reduce the SNR of the captured sound. Therefore, we should carefully design the sensing signal to handle such poor frequency response on most earphones.

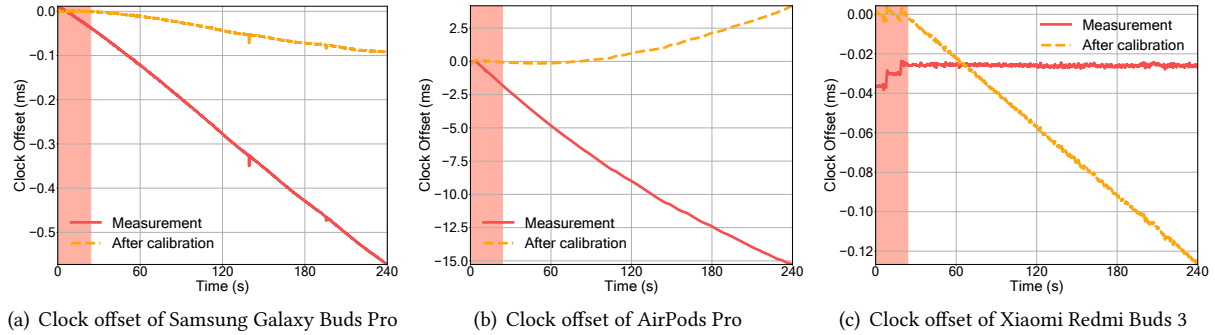


Fig. 8. Measured clock offset between smartphone and different earphones

3.2.2 Synchronization Problem in Wireless Earphones. Clock offset is a common problem for distributed ranging systems. Wireless earphones and smartphones use local oscillators separately, which brings the clock offset problem. What's worse, limited by its small size, earphones have to use oscillators that are less accurate than that in smartphones [7]. According to [14], the frequency accuracy for local oscillators on consumer applications can be ± 50 ppm (high accuracy mode) or ± 1000 ppm (normal accuracy mode). Using different clocks, the clock offset would accumulate over time.

To deal with this problem, existing methods perform calibration every time when the system begins to play back, or record [7, 62, 66, 72]. At the beginning of each run, they fix the positions of the two devices individually, take several seconds to record the signal, and compute the average offset between two frames. Then during the tracking, they calibrate the clock offset linearly by compensating the average offset for each frame.

However, it is difficult to calibrate the clock offset between the smartphone and wireless earphones. We perform a feasibility study to calibrate the clock offset linearly. We measure the clock offset between a smartphone and three different types of wireless earphones when their positions are fixed, and present the results in Figure 8. We compute the clock offset from the phase offset of ZC sequence at a frequency of 14 kHz, and the sampling rate is 48 kHz. We collect the data for 240 seconds, and use the first 24 seconds to linearly fit the offset, which is shaded in red. Although we take longer time to calibrate offset than existing methods, after 240 seconds, the accumulated clock offsets with calibration are 0.09 ms, 4.18ms, and -0.126 ms for Samsung Galaxy Buds Pro, Apple AirPods Pro, and Xiaomi Redmi Buds 3, respectively. Even for the earphone with the smallest clock offset, there would be a ranging error of 3.10 cm given that the sound speed is 343 m/s, which is beyond the millimeter-level error in phase based ranging. Besides, we also observe that curves jitter occasionally over time. From the figures, we can see that the clock offset between smartphone and earphones is nonlinear over time. Besides, for different types of device, the patterns of clock offset over time are also different.

There are two possible reasons for this phenomenon. For one thing, the performance of the oscillator is affected by temperature, and the relationship between the temperature and the frequency offset is polynomial [55], influenced by hardware-specific parameters. So, once the system starts, the temperature rises inevitably, and the clock offset changes nonlinearly. For another thing, from Figure 8, we observe that the clock offset pattern in the first tens of seconds is different from that after. Since Phase-Locked Loops (PLL) have been adopted in audio systems [28], it would take some time to initialize PLL, which interferes with the estimation on the offset. Therefore, even if we use a nonlinear function for one-shot calibration, the error due to the clock offset is usually larger than if we use a linear function. Additionally, the duration of the PLL is usually unknown for COTS wireless earphones, so it is difficult to avoid this problem by skipping a short period of time at the beginning.

To conclude, calibrating clock offset in a fashion of one-shot fitting would accumulate errors over time when tracking the wireless earphones, which limits the application of acoustic tracking system.

4 SYSTEM DESIGN

4.1 System Overview

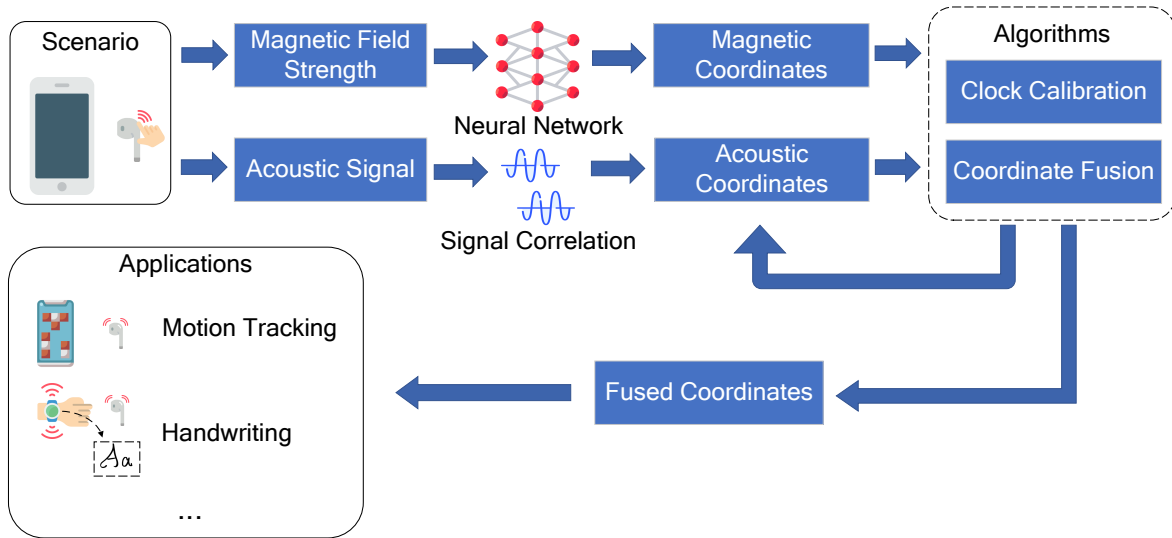


Fig. 9. Processing pipeline of MagSound

In this paper, we build an integrated wireless earphone tracking system, MagSound, which uses the built-in magnet to improve the tracking robustness against clock skew and multipath. Figure 9 illustrates the architecture of the proposed system. In our considered scenario, the user moves the wireless earphone, which emits the acoustic signal. We collect both magnetic field strength and acoustic signals from the microphones and the magnetometer of the smartphone, respectively. For magnetic field strength, we feed it into a pretrained neural network, and get the predicted coordinates, which we call *magnetic coordinates*. For acoustic signals, we first demodulate it and then compute the distance from the earphone to microphones based on the impulse response after cross-correlation. With the distance to the two microphones of the smartphone, we compute the coordinates, which we call *acoustic coordinates*, and feed it together with the magnetic coordinates into the fusion and calibration module, where we perform clock calibration and coordinate fusion. The result of clock calibration would improve the accuracy when we compute acoustic coordinates. We fuse the magnetic coordinates and the acoustic coordinates, and the fused coordinates are more accurate than the acoustic coordinates over time. We can apply it to applications such as motion tracking and handwriting.

4.2 From Magnetic Field Strength to Coordinates

In this section, we first discuss the characteristics of the magnetic field around earphones. Then we perform magnetic field strength based tracking using machine learning models, which take the magnetic field strength as the input and the coordinate as the output. Based on the comparison between different machine learning models, we finally choose neural network model for MagSound.

4.2.1 Principles of Magnetic Field Tracking. Since the earphone is magnetic, the built-in magnetometer of the smartphone can sense the movement of the earphones. If far enough from the phone, the earphone can be modeled as a point of magnetic dipole. According to [13], the relationship between the earphone's relative position and the measured magnetic field strength can be formulated as

$$\mathbf{B}(\mathbf{d}) = \mathbf{G} + \frac{\mu_0}{4\pi} \left(\frac{3\mathbf{d}(\mathbf{m} \cdot \mathbf{d})}{\|\mathbf{d}\|^5} - \frac{\mathbf{m}}{\|\mathbf{d}\|^3} \right), \quad (2)$$

where \mathbf{d} is the vector from the earphone to the built-in magnetometer, \mathbf{m} is the vector of magnetic dipole moment, and μ_0 is the vacuum permeability constant. Background magnetic field, including the Earth's geomagnetic field, also influences the measurement, and we denote it as \mathbf{G} in the equation.

However, deriving the distance $\|\mathbf{d}\|$ using Equation 2 is difficult in the considered scenario due to two reasons:

- (1) The equation is underdetermined. In total, there are 9 parameters to determine from \mathbf{d} , \mathbf{m} , and \mathbf{G} . However, we only have one magnetometer, which brings three equations corresponding to three axes.
- (2) The ideal model may not hold in practice. As we have illustrated in Section 3.1, the changes in magnetic field strength with distance are discriminable only in the range of a few centimeters. Considering the size of earphones, the assumption of magnetic dipole may fail in this small range.

In fact, it is impractical to propose an explicit model of the magnetic field between the earphone and the smartphone, because different types of devices have different magnetic field distributions. Fortunately, there is a deterministic correlation between \mathbf{d} and \mathbf{B} , as \mathbf{B} decreases from the earphone to all around according to the preliminary study in Section 3.1.

4.2.2 Learning-based Tracking using Magnetic Field Strength. Denote the 2D coordinate of the earphone as \mathbf{P} , and the measured magnetic field strength vector as

$$\mathbf{B} = \mathbf{G} + \mathbf{B}^*, \quad (3)$$

where \mathbf{B}^* is the magnetic field strength vector from the earphone. As aforementioned, there is a stable relationship between \mathbf{P} and \mathbf{B}^* . Therefore, we can use machine learning models to fit the implicit mapping from \mathbf{B}^* to \mathbf{P} .

To collect the ground truth, we first draw a grid, and then place the earphone on each grid point to collect the readings of the built-in magnetometer. The grid can be placed close to the phone's built-in magnetometer. The size of the grid is 12 cm \times 7 cm, with a granularity of 1 cm. For each grid point, we collect the readings for 2 seconds and set the sampling rate to 100 Hz, which means we can collect more than 20,000 samples within 5 minutes. In total, we collect five groups of data for each pair of earphone and smartphone.

However, we have to remove the background magnetic field before we start training, because the strength of the background magnetic field is close to that of the magnetic field around the earphone. According to [38], geomagnetic field strength ranges from 20 μT to 68 μT , and will change due to the location of the device or the change of time. We find that the total intensity of the earth's magnetic field varies by only a few tens of nT with a fundamental period of 24 hours [38]. Besides, we observe that the background magnetic field strength is almost the same at each grid point in one run. Therefore, we can eliminate the background magnetic field strength by choosing one grid point as a reference point. Formally, we denote \mathbf{B}_k as the measured magnetic field strength vector of the selected point, and \mathbf{B}_i , $i \neq k$, as the measured magnetic field strength vector of the i -th point. Then we perform

$$\begin{aligned} \mathbf{B}_i - \mathbf{B}_k &= (\mathbf{G}_i - \mathbf{G}_k) + (\mathbf{B}_i^* - \mathbf{B}_k^*) \\ &\approx \mathbf{B}_i^* - \mathbf{B}_k^*. \end{aligned} \quad (4)$$

Next, we train machine learning models using $\mathbf{B}_i - \mathbf{B}_k$. Meanwhile, when tracking the earphone, we hold the earphone and collect the readings of the built-in magnetometer in the first few seconds, and perform the same operation to feed the data into the model for inference.

Table 1. Settings of each machine learning model

Name	Setting	Error (cm)	Size (Kb)	Inference Time per Sample (ms)
RF	100 trees, at least 2 samples to split	0.76	43,581	0.011
SVR	RBF kernel, regulation parameter $C=100$	0.94	683	5.790
NN	3 hidden layers, each with 100 neurons	0.82	495	0.014

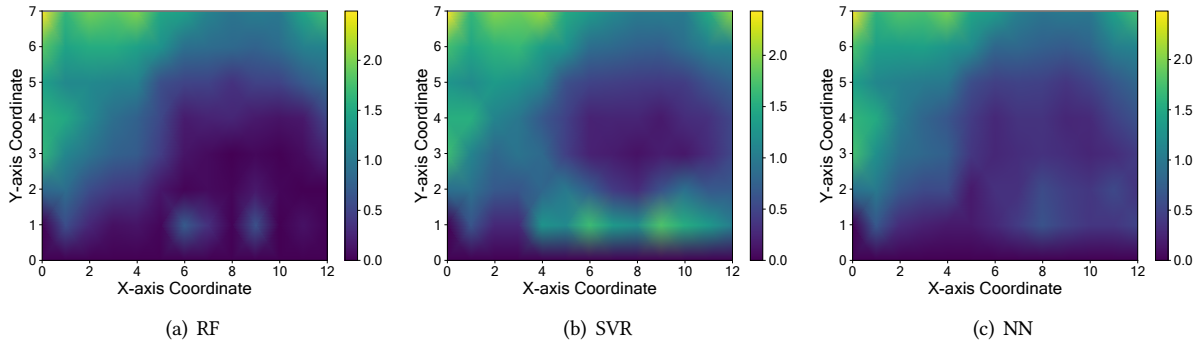


Fig. 10. Localization error distribution (cm)

We use 4 groups of data for training and 1 group of data for testing, and compare the performance between 3 light-weighted machine learning models: random forest (RF), support vector Regression (SVR), and neural network (NN). The setting, the average localization error, and the size of each model are illustrated in Table 1. We use radial basis function (RBF) as the kernel function of SVR. We evaluate the inference time per sample of different models on a desktop with Intel i7-8700U CPU. We can see that all the models achieve millimeter-level error, and the overhead of NN is more balanced compared with RF and SVR. As is shown in Figure 10, the error of NN distributes more evenly than the others. So, we choose NN as the machine learning model to track the earphone using magnetic field strength. Note that a machine learning model trained on data collected from one model of earphone - smart device pair can be directly used on other earphone - smart device pairs of the same model, since the magnetic field strength distribution of the same model of device is expected to be similar. However, when the model of the earphone or smart device is changed, it is necessary to re-collect the data and train a new neural network model.

Finally, we call the area where we have collected data as *magnetic trackable area*, and the coordinate from magnetic field strength as *magnetic coordinate* P^m . Based on localization error distribution, we define an area Ω called *effective magnetic area* from the magnetic trackable area, such that the localization error of P^m in Ω is smaller than a threshold ε . Here, a feasible algorithm to obtain Ω in a depth first search (DFS) manner is presented in Algorithm 1, where the input P_0 can be the coordinate with the strongest magnetic field strength, and ε is usually set to 1 cm. As is shown in Figure 11, depending on the shape, users can hold the earphone in different ways. When the earphone moves into the effective magnetic area (shaded area in Figure 11), the posture of holding the earphone should be the same as when the magnetic fingerprint was collected.

Algorithm 1: Effective Magnetic Area Division Algorithm

Input: Error distribution of the magnetic trackable area, a starting point $P_0 = (a, b)$, threshold ε .

Output: Boundaries of effective magnetic area Ω .

- 1 Mesh the magnetic trackable area to get a graph G of $m \times n$, where $G[i, j] \in \{0, 1\}$ denotes whether the localization error at point (i, j) is smaller than ε or not;
- 2 $stack.push(P_0)$;
- 3 $Z = \emptyset$;
- 4 $visited = zeros(i, j)$;
- 5 $Z = DFS(G, stack, visited)$;
- 6 Output Z .



Fig. 11. Holding different earphones

4.3 From Acoustic Signal to Coordinates

In this section, we build a robust signal model for MagSound. We first explain our design on transmission signals and the corresponding benefits. Then, we introduce how we perform signal modulation and demodulation. Finally, we calculate the phase shift caused by the movement of the earphone, and consequently the acoustic coordinates by optimization.

4.3.1 Transmission Signal Design. Using the earphone, we transmit Zadoff-Chu (ZC) sequence [71] in an OFDM manner for acoustic ranging. Formally, a ZC sequence with a length of N can be described with

$$ZC[n] = \exp\left(-j\frac{\pi nr(n + c_f + 2q)}{N}\right), j = \sqrt{-1}, \quad (5)$$

where $0 \leq n < N$, $q \in Z$, $c_f = N \bmod 2$, $0 < r < N$, and $\gcd(N, r) = 1$. $ZC(n)$ represents the n -th sample point. Parameter r is also called the root of the ZC sequence. ZC sequence owns the property of constant amplitude and zero auto-correlation at non-zero lags [50], which brings the following advantages in the acoustic sensing system:

- (1) The power of ZC sequence is stable, and thereby reduces requirements for earphone quality. Signals with inconstant power would cause the instantaneous power of the speaker to fluctuate significantly at times. Hence, the speaker needs more circuit components to operate without distortions and being damaged.
- (2) ZC sequence has the ideal periodic auto-correlation function. The auto-correlation result of a ZC sequence is non-zero only at $n = 0$. This would help us determine the phase shift at the receiver.

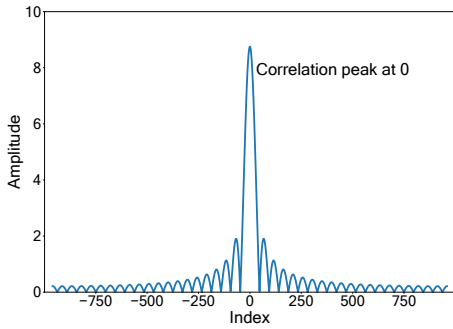


Fig. 12. Auto-correlation of ZC sequence without noise

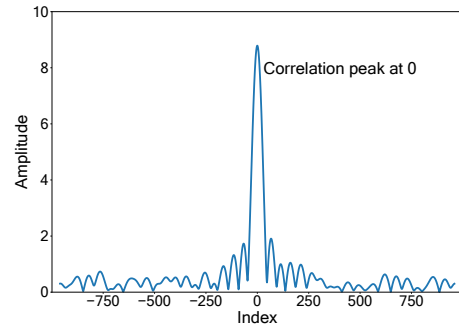


Fig. 13. Auto-correlation of ZC sequence with noise

- (3) ZC sequence has a good cross-correlation function. For two ZC sequences whose roots are co-prime, their correlation peak is almost zero. This means the ZC sequence is robust to noise and can be distinguished from other signals.
- (4) The result of the Fourier transform of a ZC sequence is still a ZC sequence. Hence, either Fast Fourier Transform (FFT) or Inverse Fast Fourier transform (IFFT) would not change the property of the ZC sequence.

In fact, ZC sequence has been widely used in recent related works [58, 63]. In our study, we set $q = 0$, $N = 41$, and $r = 5$ to generate a ZC sequence.

4.3.2 Signal Modulation and Demodulation. We modulate the ZC sequence in an OFDM manner. A widely accepted frequency threshold for ultrasound is 20 kHz [59]. Unfortunately, as mentioned in Section 3.2.1, most COTF wireless earphones have poor frequency response when the frequency is above 15 kHz. Therefore, to minimize the disturbance to the user while ensuring compatibility with most earphones, we set the center frequency f_c to 14 kHz and the bandwidth to 1025 Hz. After interpolation, we get a sequence of length $L = 1920$. We set the sampling rate to 48 kHz, so the period of a received frame is $\frac{1920}{48000} = 40$ ms, which is small enough to sense movement.

Meanwhile, the smartphone records the signals using built-in microphones, with the sampling rate $f_s = 48$ kHz. Denote the transmitted signal as ZC^T , the received signal as ZC^R , and $\text{round}(a)$ as the integer part of a . Due to the multipath effect, the received signal is a superposition of multiple copies of transmitted signals with different delays, which can be described with

$$ZC^R[n] = \sum_{i=1}^M A_i \exp(-j\phi_i) ZC^T[n - \text{round}(\tau_i f_s)], \quad (6)$$

where M is the number of propagation paths, A_i is the attenuation coefficient for the i -th path, $\phi_i \in (-\pi, \pi)$ represents the phase shift caused by propagation, and τ_i is the time of flight (ToF) for the i -th path.

To get the ToF for the direct path, *i.e.*, LOS path, we first perform FFT on the received signal. Next, we estimate the Channel Impulse Response (CIR) by conjugating the received signal ZC^R and cross-correlating it with the transmitted signal ZC^T by multiplication in the frequency domain. We upsample the signal in the frequency domain by zero-padding to a length of L and then perform IFFT back to the time domain. Since the OFDM signal has a rectangular frequency gate function and a limited band, the corresponding time-domain CIR is the convolution of the sinc function with the impulse response. Finally, since the ZC sequence has the ideal

periodic auto-correlation function, we would observe several peaks which represent different delays of different propagation paths.

To test the robustness of the generated ZC sequence to noise, we first simulate the auto-correlation of the ZC sequence without noise in Figure 12. Then, we set SNR to 0 dB, and simulate the auto-correlation with noise in Figure 13. We can see that the peak is clear and distinguishable in both figures.

4.3.3 Phase-based Tracking. In practice, we care about the direct path from the earphone to the smartphone, *i.e.*, the LOS path. In other words, we calculate the ToF τ_1 of the first propagation path. From Equation 6, the ZC sequence transmitted through LOS path is

$$ZC_1^R[n] = A_1 \exp(-j\phi_1) ZC^T [n - \text{round}(\tau_1 f_s)]. \quad (7)$$

To calculate the ToF τ_1 of the first propagation path, we take the index of the first peak in CIR, which is actually the integer part of $\tau_1 f_s$, *i.e.*, $\text{ind}_1 = \text{round}(\tau_1 f_s)$. To find ind_1 , we first find the maximum value A_{\max} in CIR, and then iterate the CIR to find the first value over $A_{\max} \cdot \epsilon$ and the corresponding index, where ϵ is a threshold to control the robustness.

The corresponding CIR of the LOS path is a sinc function as

$$\text{CIR}[\text{ind}_1] = A_1 \exp(-j\phi_1) \text{sinc} [n - \text{round}(\tau_1 f_s)]. \quad (8)$$

We compute the phase shift ϕ_1 of this complex value. As $\phi_1 \in (-\pi, \pi)$, we accumulate the phase change by adding or subtracting 2π to phase to get the phase shift Φ_1 caused by ToF through the LOS path. Then the ToF through the LOS path is derived as

$$\tau_1 = \frac{\Phi_1}{2\pi f_c}. \quad (9)$$

Since there are two built-in microphones, we can achieve 1D or 2D tracking with COTS smartphones. Without loss of generality, we introduce our method for 2D tracking as an example. Following related works in [7, 66], we first place the earphone at a known position P_0 for calibration of a few seconds. Since the devices are fixed at this time, we can collect the clock offset δ of each frame, and then take the average $\bar{\delta}$ as the constant clock offset after the calibration. Besides, we compute the initial distance from the earphone to the microphones and the initial phase.

Traditionally, the 2D coordinate of the earphone can be computed by solving a system of geometric equations:

$$\begin{cases} \text{dist} (P, P^1) = d_1 \\ \text{dist} (P, P^2) = d_2 \end{cases}, \quad (10)$$

where $\text{dist}(A, B)$ computes the distance between point A and point B . However, different from existing works, we compute the coordinates while tracking the earphone by solving an optimization problem. This is because the clock offset problem in wireless earphones is significant, and it is common that the system of equations has no solution. In contrast, solving an optimization problem can always return a solution. Specifically, for the i -th microphone, we cancel out the clock offset, compute the accumulated distance change of each channel and update the distances between the earphone and the microphone as d_1 and d_2 . We denote the coordinate of the earphone as P , the i -th microphone as P^i , and optimize the following objective function:

$$\min_P \left\{ (\text{dist} (P, P^1) - d_1)^2 + (\text{dist} (P, P^2) - d_2)^2 \right\}. \quad (11)$$

To optimize the objective function, we use COBYLA solver [52] that can solve a constrained problem with low delay by linear approximation. Finally, we call the coordinate from acoustic signal as *acoustic coordinate* P^a , which uses one-shot linear calibration.

Algorithm 2: Clock Offset Calibration Algorithm

Input: Magnetic coordinate P_i^m in the i -th frame, acoustic coordinate P_i^a in the i -th frame, effective magnetic area Ω , objective function L .

Output: Estimated parameter vector \mathbf{x} of $h(i)$.

```

1  $flag = 0$ ;
2  $F = []$ ;
3 foreach  $frame_i$  do
4   if  $P_i^m \in \Omega$  then
5      $flag = 1$ ;
6      $offset = getOffset(P_i^m, P_i^a)$ ;
7      $F.add(offset)$ ;
8   else
9     if  $flag == 1$  then
10       $\mathbf{x} = \underset{\mathbf{x}}{\operatorname{argmin}} L(F; \mathbf{x})$ ;
11    end
12     $flag = 0$ ;
13  end
14 end

```

4.4 Clock Calibration

Using magnetic coordinate P_m , the tracking of the earphone would not be affected by the clock offset when the earphone moves inside Ω . However, the clock offset between two adjacent frames would change over time. When the earphone keeps moving outside Ω , the offset still accumulates. Thus, in this subsection, we try to fit the clock offset multiple times using P_m in a nonlinear way to calibrate it.

Without loss of generality, for one channel, the clock offset function can be expressed explicitly by $h(t; \mathbf{x})$, where \mathbf{x} denotes the vector of parameters to estimate. For example, if we assume that $h(t; \mathbf{x})$ is a quadratic function of t , then the expression is

$$h(t; \mathbf{x}) = at^2 + bt + c \quad (a \neq 0), \quad (12)$$

where $\mathbf{x} = (a, b, c)$.

The calibration algorithm is presented in Algorithm 2. When $P_i^m \in \Omega$, we take P_i^m as the ground truth, and compute the clock offset in P_i^a . Then we collect the offsets in the array F , and for the first frame where the earphone moves out of Ω , we estimate \mathbf{x} by minimizing the objective function $L(F; \mathbf{x})$. When $h(t)$ is a quadratic function of t , an example of the objective function is

$$L(F; a, b, c) = \sum_{i=1}^n (ai^2 + bi + c - F[i]) + \left| \frac{a - a_0}{a_0} \right| + \left| \frac{b - b_0}{b_0} \right|, \quad (13)$$

where a_0 and b_0 are the initial guess of a and b , respectively, and n is the size of F . Note that we regularize a and b in case that the parameters deviate too much. To minimize the objective function, we use the Powell solver [51] that avoids calculating derivatives and converges fast.

In the beginning, when the size of F is small, it is difficult to fit $h(t)$, yet the performance of Algorithm 2 improves as the size of F increases. To verify the idea, we perform a simulation test based on the data collected in Section 3.2.2. We assume that the earphone moves into the effective magnetic area Ω every 60 seconds, and moves inside Ω for 6 seconds. We calibrate the clock offset linearly in the first two rounds, and then model $h(t)$

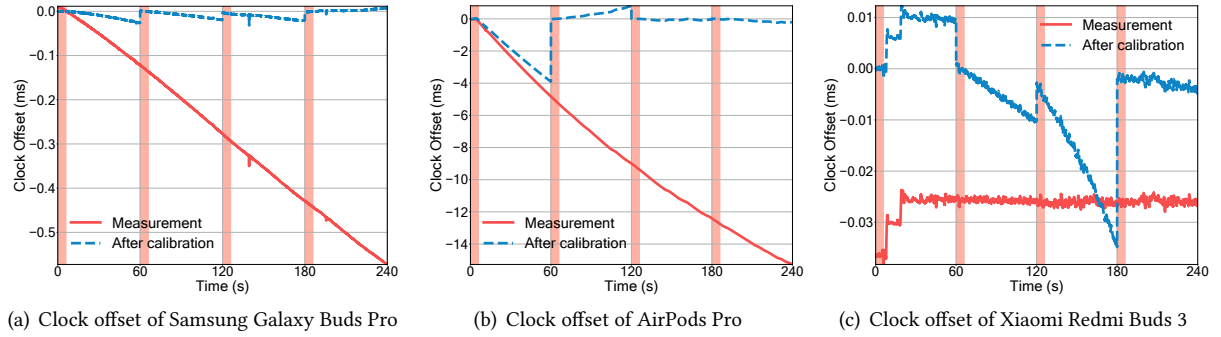


Fig. 14. Simulated clock offset calibration on different earphones

Table 2. Comparison between magnetic coordinate and acoustic coordinate

Name	Clock Offset	Multipath	Effective Distance	Localization Error Level
Magnetic coordinate	No	No	5 cm to 10 cm	Millimeter
Acoustic coordinate	Big	Yes	Within 1 meter	Millimeter

as a quadratic function. The calibration result is shown in Figure 14, where the calibration stage is represented by red shaded areas. Finally, the clock offsets after calibration are 0.008 ms, -0.22 ms, and -0.004 ms for Samsung Galaxy Buds Pro, Apple AirPods Pro, and Xiaomi Redmi Buds 3, respectively. Although we only used a quarter of the calibration time of one-shot calibration, the nonlinear multi-shot calibration reduced the offset to 3.17% of the one-shot calibration for Xiaomi Redmi Buds 3. As expected, after the first calibration, the calibrated offset is even larger than one-shot calibration since we take much less calibration time. However, the calibrated offset decreases quickly since the second calibration, and converges to 0 in the end. In this way, MagSound maintains millimeter-level error when tracking the earphone over time. We also tried higher order expressions of $h(t)$, but the fitting error is bigger than the quadratic function at the beginning. In the end, we can get $h(t)$ for each channel, *i.e.*, there would be two clock offset functions in 2D tracking. We hereafter denote the clock offset of all channels as a vector $\mathbf{h}(t)$.

4.5 Coordinate Fusion

Now we have a magnetic coordinate \mathbf{P}^m and an acoustic coordinate \mathbf{P}^a , and the features of the two modalities are compared in Table 2. Before we introduce the coordinate fusion algorithm, we first analyze the clock offset in \mathbf{P}^a . In the i -th frame, denote the clock offset as δ_i , the accumulated propagation delay by Equation 9 as τ_i , and the accumulated propagation delay without clock offset as τ_i^* . For the n -th frame of tracking, we have

$$\tau_n = \tau_n^* + \sum_{i=1}^n \delta_i. \quad (14)$$

Hence for two adjacent frames, the difference between their propagation delay is

$$\tau_{n+1} - \tau_n = (\tau_{n+1}^* - \tau_n^*) + \delta_{n+1}. \quad (15)$$

Algorithm 3: Coordinate Fusion Algorithm

Input: Magnetic coordinate P_i^m in the i -th frame, acoustic coordinate P_i^a in the i -th frame, effective magnetic area Ω , clock offset function $h(t)$.

Output: Fused Coordinate P_i^f in the i -th frame.

```

1 foreach  $frame_i$  do
2    $V = \text{calibrate}(P_i^a, h(i)) - \text{calibrate}(P_{i-1}^a, h(i))$ ;
3   if  $P_i^m \in \Omega$  then
4      $P_i^f = P_i^m$ ;
5   else
6      $P_i^f = P_{i-1}^f + V$ ;
7   end
8 end

```

The key insight is that in one frame, the clock offset δ_i is small, and thereby the difference between acoustic coordinates from two adjacent frames is small enough to ignore. For example, in Figure 8(a), the distance error caused by clock offset between two adjacent frames is less than 0.005 mm. What's even better, this small error would be further mitigated when we use the clock offset function $h(t)$ to calibrate per frame. So, we can improve the accuracy by limiting the offset accumulation using magnetic coordinate P^m .

Accordingly, we formulate the coordinate fusion algorithm in Algorithm 3. Suppose that we have obtained the clock offset function $h(t)$ corresponding to each frame by Algorithm 2. For the i -th frame, we first calibrate P_i^a , and then calculate the displacement V between calibrated acoustic coordinates in two adjacent frames. Next, if $P_i^m \in \Omega$, we take $P_i^f = P_i^m$. Otherwise, we take $P_i^f = P_{i-1}^f + V$. It is worth noting that when the earphone moves outside the preset grid, the predicted coordinates by the model would be around the edges of the grid, because the measured magnetic field strength is indistinguishable when the earphone moves far enough away.

There are mainly two advantages of the Algorithm 3:

- (1) The clock offset accumulation is limited. When the earphone moves into Ω , we straightforwardly take $P_i^f = P_i^m$. Denote the last fused coordinate when the earphone is in Ω as P_j^f . When the earphone moves out of Ω , the following fused coordinate would be

$$P_{j+k}^f = P_j^f + \sum_{i=j+1}^{j+k} V_i. \quad (16)$$

Thus, the clock offset would accumulate for at most k frames, and once the earphone moves into Ω , the accumulated clock offset is removed.

- (2) The fused coordinate is robust to multipath. For one thing, when the earphone moves near the smartphone, there would be a serious multipath effect by the reflection of the smartphone's body. Fortunately, this is just the time when the earphone moves into Ω , and we take the magnetic coordinate P_i^m as the fusion result, which is not affected by multipath. For another thing, similarly to clock offset, the error caused by multipath when the earphone is outside Ω would not accumulate and would be removed once the earphone moves into Ω .

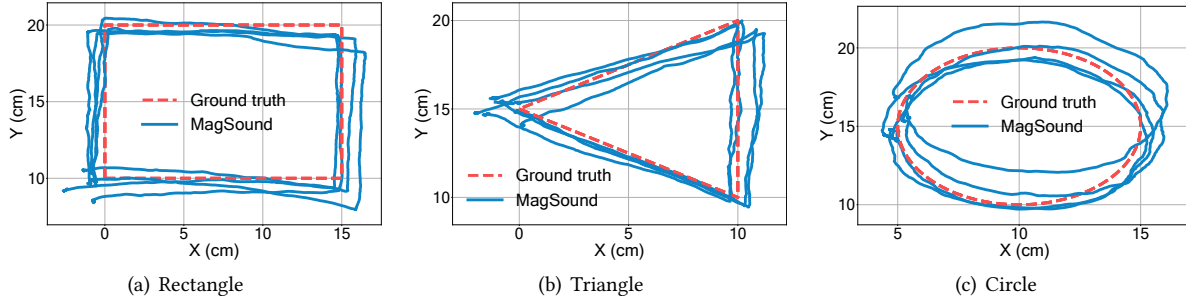


Fig. 15. Trajectories of MagSound

5 EVALUATIONS

5.1 Implementation

We implement a prototype of MagSound to verify its performance on 2D tracking and handwriting. To facilitate the use of optimizers and machine learning models, we implement MagSound using a client-server architecture. We develop an app on the Android platform, which transmits the acoustic signal and collects both the audio and magnetometer readings. The transmitted ZC signal has a center frequency of 14 kHz and a frame length of 40 ms. The sampling rate of the microphone is 48 kHz. We set the sampling rate of the magnetometer to 100 Hz, which means there would be 4 readings in a frame of 40 ms. We take the average of the 4 readings and pack it together with the audio frame into a binary packet. Besides, a stand-alone software is developed on PC platform to receive, process and visualize data in real time. The default devices are Google Pixel 4 and Samsung Galaxy Buds Pro. At the beginning of each run, we hold the earphone for 4 seconds for one-shot calibration. During the experiments, users may constantly move the earphone inside the effective magnetic area for 5-10 seconds for calibration. Once we detect the earphone entering the effective magnetic area, we collect the clock offset, and then perform the re-calibration when the earphone moves out of the effective magnetic area. In practical applications, users do not need to remember the boundaries of the effective magnetic area in detail. Still, they only need to move the earphone close to the smartphone after writing a trace each time. We call the tracking method that only uses acoustic signals and calibrates the clock offset once at the beginning as *One-shot* in the experiment.

5.2 2D Tracking

5.2.1 Setup. For 2D tracking, we first draw the trace on a grid of paper as the ground truth, and then move the earphone along the trace. Unless stated, we draw a rectangle of 15 cm \times 10 cm for four times in one run and calibrate the earphone at the intervals. The initial distance between the earphone and the smartphone is 10 cm. The average running time is 240.08 seconds for one run. We compute the tracking error after aligning the captured trace with the ground truth using dynamic time warping [31]. By default, we use a wooden table with metal supports in the experiment.

5.2.2 Comparison with One-shot Calibration. We use the earphone to draw three different shapes, *i.e.*, rectangle, triangle and circle, and compare the tracking accuracy between MagSound and One-shot. In each run, we draw the shape four times, and calibrate the earphone at the interval. Figure 15 and Figure 16 show the trajectories using MagSound and One-shot, respectively. We also draw the trajectories whose coordinates are computed by solving geometric equations in Figure 17, which is denoted by *Geo-one-shot*. When there is no real solution for the geometric equations, we take the last coordinate as the current coordinate.

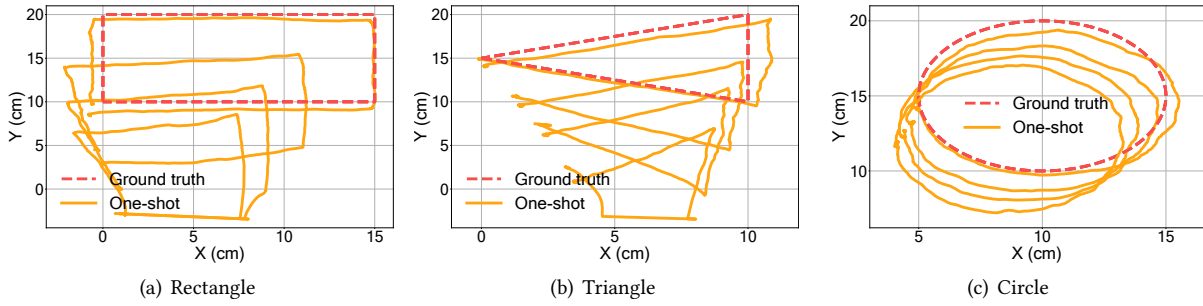


Fig. 16. Trajectories of One-shot

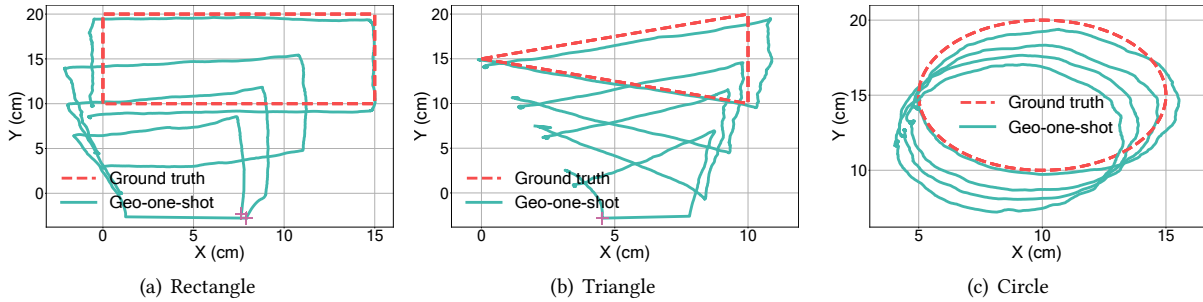


Fig. 17. Trajectories of Geo-one-shot (pink pluses represent the point where Geo-one-shot has no real solution)

We can see that the trace by One-shot gets distorted and drift over time, while the trace by MagSound keeps relatively stable. Meanwhile, using Geo-one-shot, the trajectories are similar to One-shot, but lose some sample points. Particularly, using One-shot or Geo-one-shot, the last trace in the trajectories of rectangle or triangle is partially below the x-axis, where the case that Geo-one-shot has no real solution happens. The proportion of this case is 19.87% and 8.14% for rectangle and triangle, respectively. We label this case with a pink plus in Figure 17(a) and Figure 17(b). The results show that it is more robust to calculate coordinates in an optimization way than solving systems of geometric equations when there is a significant clock offset.

In total, MagSound achieves 2D tracking error of 0.67 cm for rectangle, 0.38 cm for triangle, and 0.59 cm for circle. The median tracking error between MagSound and One-shot is compared in Figure 18. MagSound improves the accuracy by at least 27.19% to One-shot. Specifically, in Figure 19 we compare the performance of each trace for the circle trajectories where the One-shot performs the best. As we draw more traces, the error of One-shot increases while the error of MagSound even gets reduced. Since we collect the clock offset data for function fitting at intervals, as time goes by, we would have more data to fit and achieve smaller fitting error.

5.2.3 Impact of Difference Devices. We use different earphones and smartphones to evaluate the impact of different devices. As is shown in Figure 20, the smartphones are Google Pixel 4 (S1), Xiaomi Redmi K40 (S2) [20], and Samsung S7 (S3) [17]. The wireless earphones are Samsung Galaxy Buds Pro (E1), Xiaomi Redmi Buds 3 (E2), and another pair of Samsung Galaxy Buds Pro (E3). For experiments on S1-E3, we use the same pretrained neural network model as S1-E1 to evaluate the device-to-device variation. The median error of MagSound is similar

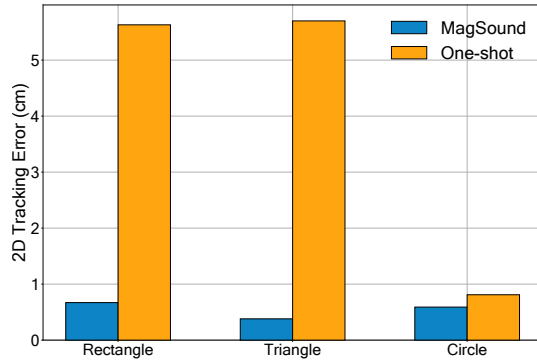


Fig. 18. Tracking errors of different shapes

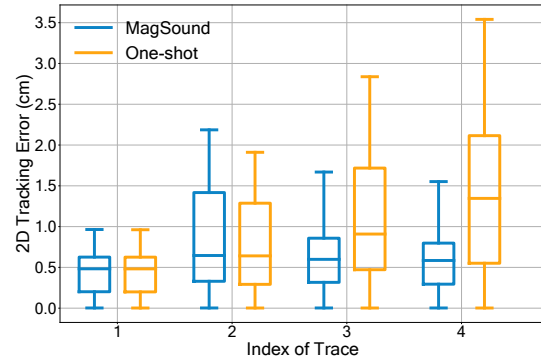


Fig. 19. Tracking errors per circle



Fig. 20. Different devices

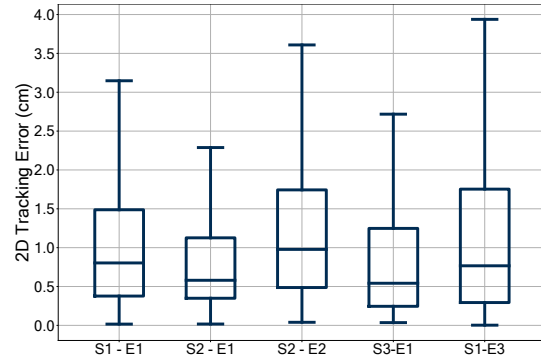


Fig. 21. 2D tracking error of different devices

for all device pairs, ranging from 0.54 cm to 0.98 cm. Figure 21 compares the tracking error between different devices, and we observe that MagSound works well even for devices from different suppliers, *e.g.*, Xiaomi K40 and Samsung Buds Pro (S2-E1). The performance of S1-E1 and S1-E3 is close to each other, which verifies that the pretrained machine learning model is generic to devices of the same model.

5.2.4 Impact of Ambient Sound and Magnetic Interference. As is shown in Figure 22, we evaluate the impact of ambient sound and magnetic interference on MagSound in seven scenarios. For ambient sound, we perform evaluation in three different environments: laboratory (noise at 54.5 dBA), cafeteria (noise at 52.7 dBA), and corridor (noise at 64.7 dBA). We measure the noise using a sound level meter. In the cafeteria, people eat snacks and chat softly. In the corridor, people make footsteps and talk loudly. As is shown in Figure 23, the error distribution of each environment is close to each other. The median error of the three cases are 0.67 cm, 0.45 cm, and 0.88 cm. We attribute this result to the ZC sequence, which helps us distinguish transmitted signal and noise.

For magnetic interference, we perform evaluation in four different scenarios. Firstly, to test the hard iron effect [1], we place a strong magnet at 15 cm from the smartphone, which is a common distance in magnet-based sensing [9]. Secondly, to test the soft iron effect [54], we place three ferrite rods randomly around the smartphone, and the nearest ferrite rod is 5 cm to the smartphone as in [64]. Thirdly, we place an electric fan under the wooden table, directly below the smartphone, and keep the fan running during the experiment to test the impact of electromagnetic interference since the electric fan is a typical home appliance driven by motors. The vertical

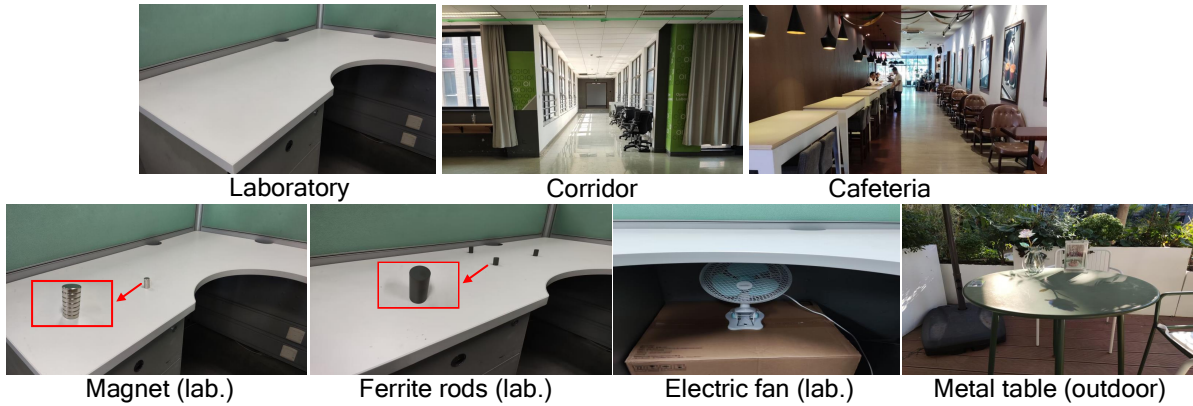


Fig. 22. Different environments and magnetic interference

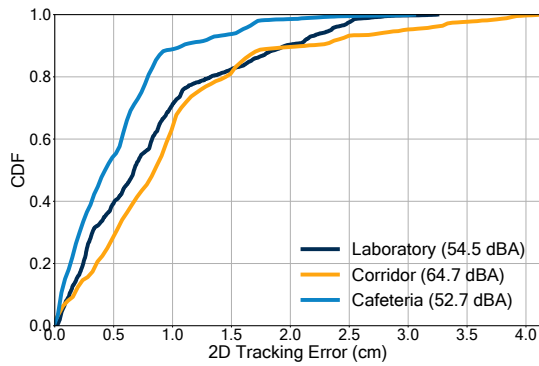


Fig. 23. Impact of different environments

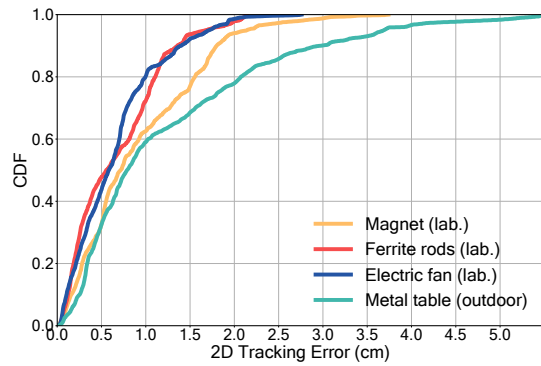


Fig. 24. Impact of magnetic interference

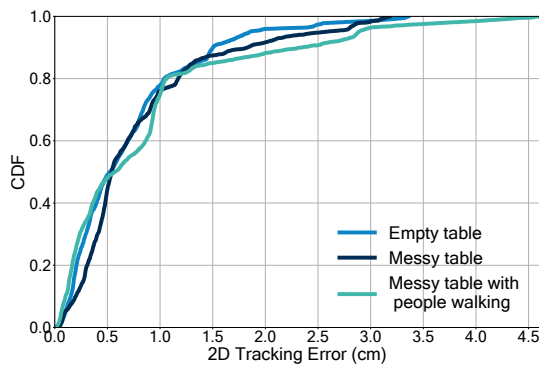


Fig. 25. Multipath environments

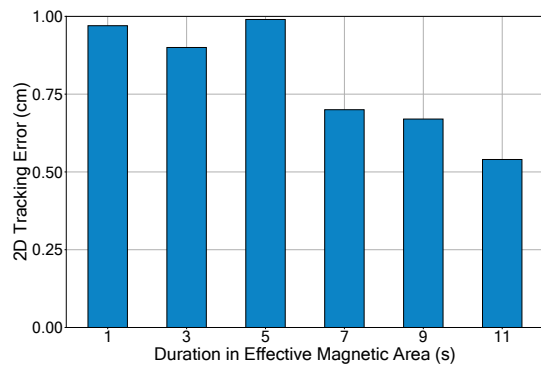


Fig. 26. Duration in effective magnetic area

distance between the top of the electric fan and the mobile phone is less than 10 cm. Finally, we perform the evaluation on an iron metal table of an outdoor coffee shop. As is shown in Figure 24, the error distribution of

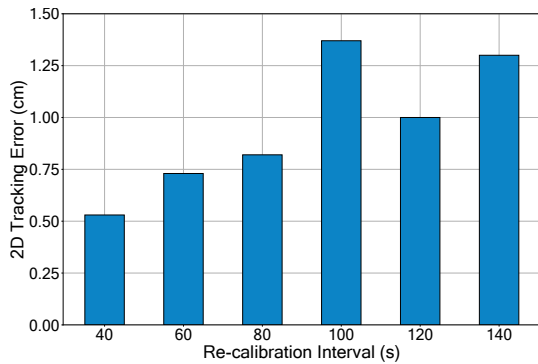


Fig. 27. Impact of correction interval

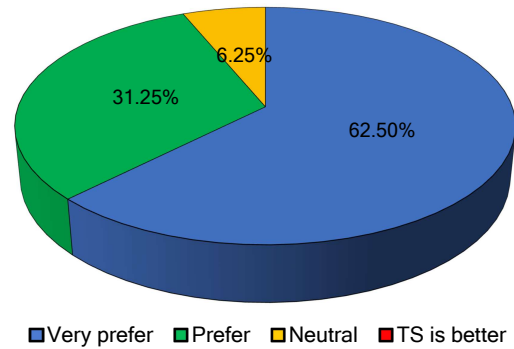


Fig. 28. User preference for MagSound

each environment is close to each other. The median error of the four cases are 0.72 cm, 0.55 cm, 0.58 cm, and 0.79 cm, respectively. We observe that neither hard iron effects nor soft iron effects have a significant impact on the accuracy. We attribute the robustness to hard iron effects to that we have removed the background magnetic field before the model inference. Besides, we find that the range of electromagnetic interference for common household appliances is limited within a few centimeters. Furthermore, the induced magnetic field strength around soft iron is too weak to influence the smartphone. Although the iron metal table hinders the movement of the earphone due to the mutual attraction of the two, we find that MagSound can still re-calibrate the tracking and maintains millimeter-level accuracy.

5.2.5 Impact of Multipath. We evaluate the impact of multipath in three different cases. The first case is an empty table, which corresponds to no environmental multipath. For static multipath, we place bags and books 10-15 cm away from the smartphone or the earphone to make a messy table. Further, we ask volunteers to walk around the messy table at 50-70 cm to introduce dynamic multipath. The median error of the three cases are 0.53 cm, 0.54 cm, and 0.58 cm, respectively. The error distribution of each case is shown in Figure 25. We observe that multipath does have an effect on the accuracy, especially for extreme errors. However, the error distribution of MagSound is similar in the three cases, which verifies its robustness.

5.2.6 Impact of Duration in Effective Magnetic Area. We evaluate the impact of duration in effective magnetic area from 1 second to 11 seconds in intervals of 2 seconds. We start timing when the earphone moves into the area, and we remove the earphone from the area when the time is up. Note that when the duration is less than or equal to 1 second, users only need to move in and out of the earphone, and the time is negligible. We plot the median tracking error in Figure 26. We can see that the accuracy improves with more duration in the effective magnetic area. This is because more duration means more clock offset data to fit. Besides, when the duration is more than 7 seconds, the performance improvement becomes small, which means MagSound do not need too long duration in the effective magnetic area.

5.2.7 Impact of Correction Interval. We evaluate the impact of correction interval from 40 seconds to 140 seconds in intervals of 20 seconds. Figure 27 shows the median tracking error under different correction intervals. We can see that the performance of MagSound exhibits a slow decline as the interval increases. Nevertheless, MagSound maintains millimeter-level accuracy within an interval of 100 seconds, which is expected to meet the requirements of applications such as drawing a fine-grained trajectory or inputting a few words. As discussed in Section 5.2.2, when the number of re-calibration increases, the performance of MagSound is expected to be better.

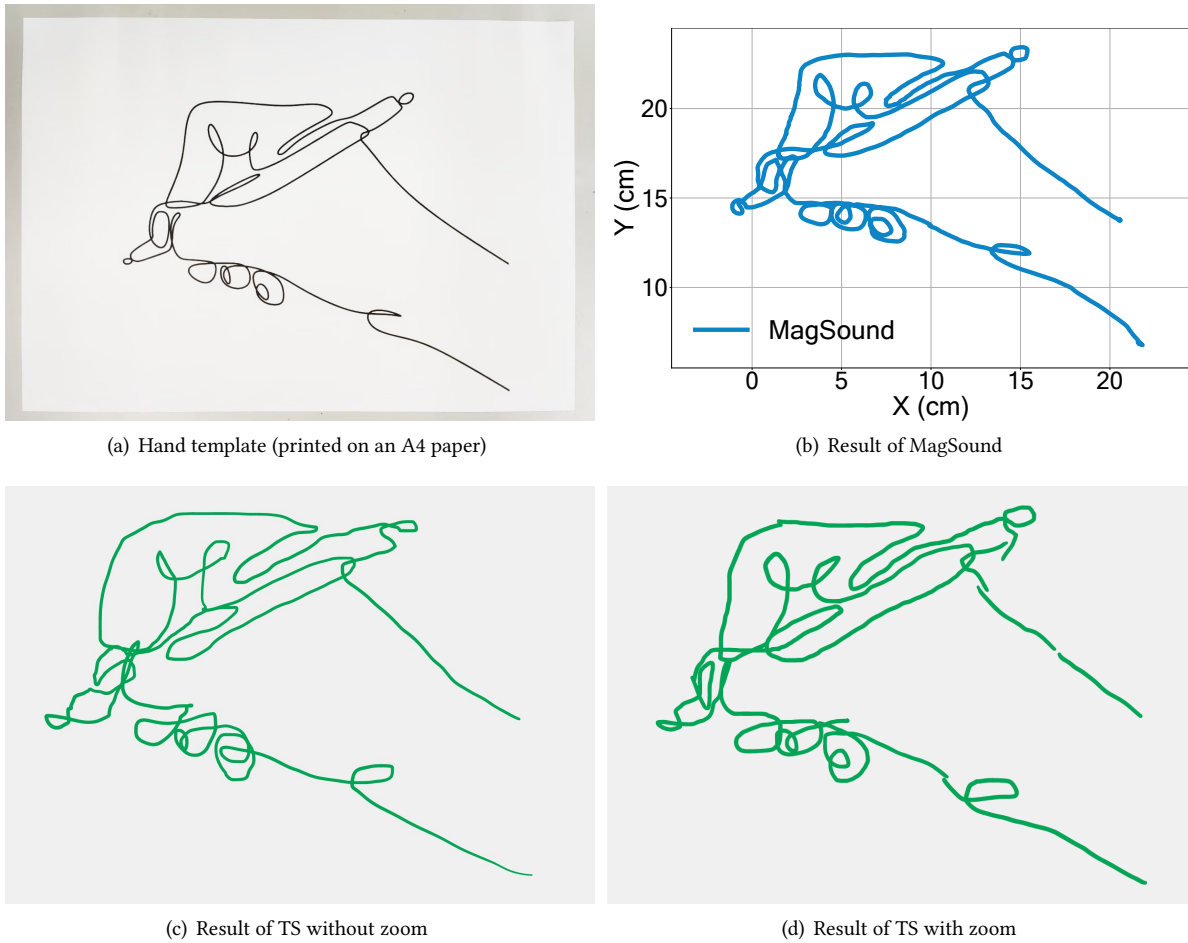


Fig. 29. Contrast of MagSound and touch screen based drawing (TS) on fine-grained drawing

5.3 Potential Applications

One of the potential applications of MagSound is fine-grained drawing on smart devices. Currently, many smart devices, including smartphones and smartwatches, are not equipped with a stylus. However, when we try to do fine-grained drawing on the touch screen, the fat finger problem [56] severely affects the painting accuracy. When drawing on the touch screen, the drawing would get distorted due to the occlusion of the target by the finger. Therefore, targets on the touch screen have a certain minimum size to be reliably selected by the finger. According to [25, 26, 60], the value for this minimum target size varies from 1.05 cm to 2.60 cm depending on different study conditions. Besides, in practice, users may divide the target trajectory, enlarge each part, and then draw the trace in sequence. This solution does improve performance to some extent. However, frequent zoom-in and zoom-out seriously affect the user experience, making the process laborious and disorienting to users [60]. Moreover, the drawing may break at the junction of each segment. What's worse, the finger pitch and finger roll

angles influence the sensing accuracy of the touch screen, too [27]. Meanwhile, MagSound avoids the fat finger problem, as MagSound does not need the touch screen and has a larger sensing range.

To illustrate the fat finger problem, we recruit 16 volunteers to perform fine-grained drawing using the Microsoft Whiteboard app [16] on a Google Pixel 4 smartphone. Figure 29(a) shows the drawing template of the experiment. We first set the template picture the same size as the screen on the canvas, and then we ask the user to draw according to the template. We call such a drawing method as Touch Screen based drawing (TS). We collect the user’s paintings with and without zooming, respectively. For comparison, we also ask the user to draw using MagSound, where the template is printed on an A4 paper. Figure 29(b), Figure 29(c), and Figure 29(d) show the example results corresponding to the three methods.

Through the experiment, we observe that the details of the drawing using TS without zoom are severely distorted. Moreover, when using TS with zoom, we observe evident breakage of the trajectory at the junction of the adjacent parts. In contrast, the result of MagSound is the smoothest and is the closest to the template among the three methods. After the experiment, we asked the volunteers “Do you prefer MagSound to TS?” with four options: very prefer, prefer, neutral and TS is better. Figure 28 presents the feedback from participants. We can see that 93.73% of the participants prefer or very prefer MagSound to TS. Most users express their complaints about the fat finger problem of TS, and only one user holds the view that without considering the time cost, he can draw better using TS with zoom.

Thus, compared with traditional touch screen based methods, MagSound has advantages on drawing accuracy. Moreover, as the size of the template far exceeds the size of the touch screen of common smart devices, such as smartphones and smartwatches, earphone tracking based interactions can push the limit of traditional touch screen based interactions in a larger plane.

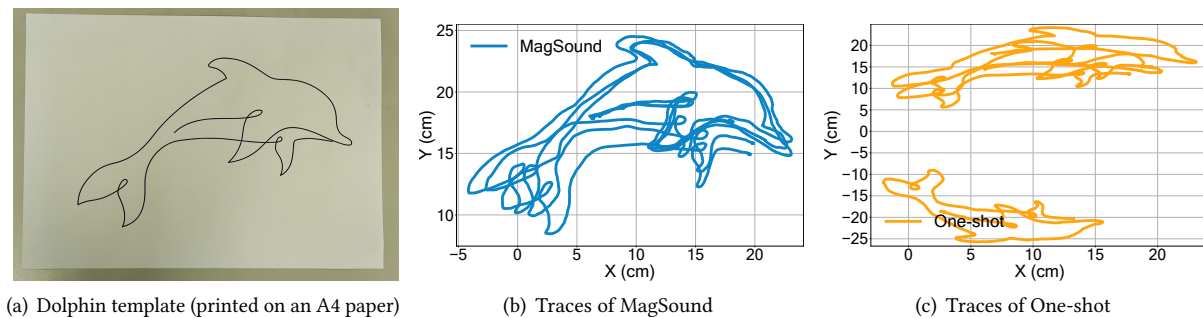


Fig. 30. Fine-grained drawing of a dolphin

To evaluate the re-calibration of MagSound, we show another complex drawing sample of a dolphin. We draw the trace three times in total, and calibrate the earphone at intervals. As is shown in Figure 30, we can see that the three traces by MagSound kept close to the original template, while those by One-shot become twisted.

The result shows that MagSound can track the earphone in a fine-grained manner and enable the smart devices to perform applications such as fine-grained drawing. Further, we believe that the ability to track the complex movements of the earphone will enable the earphone to be adapted for VR game interactions, *e.g.*, controlling the movement trajectory of a character. In fact, there have been some novel Mobile Motion Games (MMG), such as SwordFight [74] and Big Stomach Challenge [33].

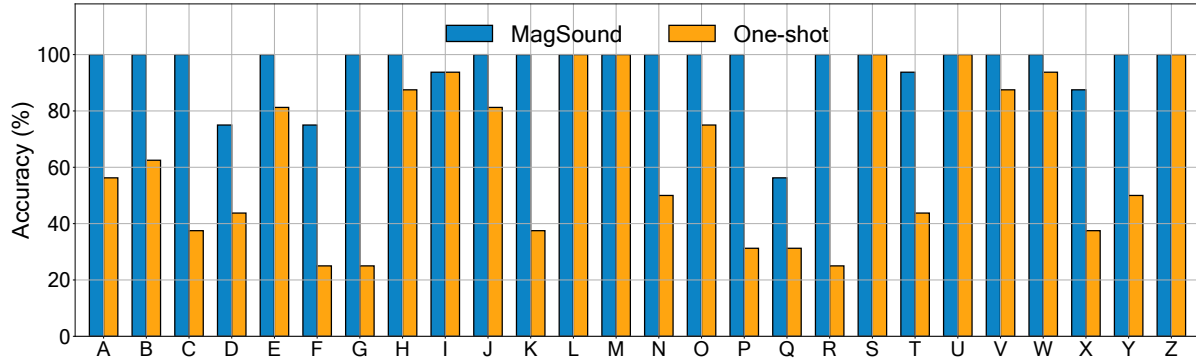


Fig. 31. Letter recognition accuracy



Fig. 32. MagSound on letter 'Q'

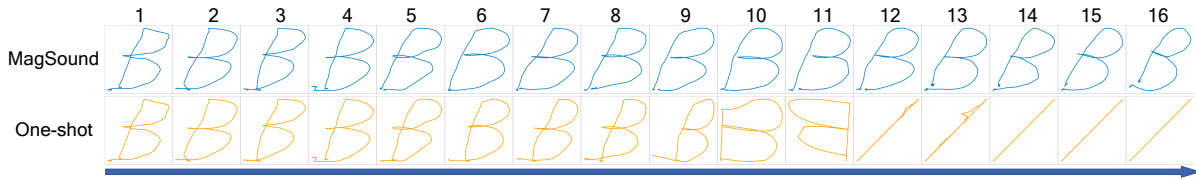


Fig. 33. Letter trace example of 'B'

5.4 Case Study of Handwriting Recognition

5.4.1 Setup. We also perform a case study on handwriting recognition. Compared with 2D tracking tasks, handwriting recognition focuses more on the maintenance of relative shape rather than absolute distance. We invite volunteers to write 26 uppercase letters using the earphone, each letter 16 times. In the scenario, the user would choose the recognized word every time a word is written, and we can re-calibrate the tracking at the same time. According to [5, 24], the average word length is 4 letters. Therefore, the tracking would be re-calibrated every 4 letters in one run. We input the captured trace into the recognition interface developed by Google Input Method Editors (IME) [23]. Google IME takes trajectory coordinates as input, and returns inferred letters. A letter trajectory is recognized, if Google IME returns the corresponding ground truth. We use recognition accuracy as the metric for evaluation. Given the same recognition interface, the accuracy of the input trajectory determines the recognition accuracy.

5.4.2 Comparison with One-shot Linear Calibration. The proposed MagSound can improve the recognition accuracy to One-shot significantly. In Figure 31, we compare the recognition accuracy of each letter between MagSound and One-shot. We can see that the accuracy of MagSound is higher than or equal to One-shot on every letter. However, the accuracy of letter 'Q' is poor for both methods. As is shown in Figure 32, the trajectories obtained by MagSound are actually easy to recognize. A possible reason is that the stroke order will affect the

(a) MagSound

(b) One-shot

Fig. 34. Comparison on 'Nice to meet you!'

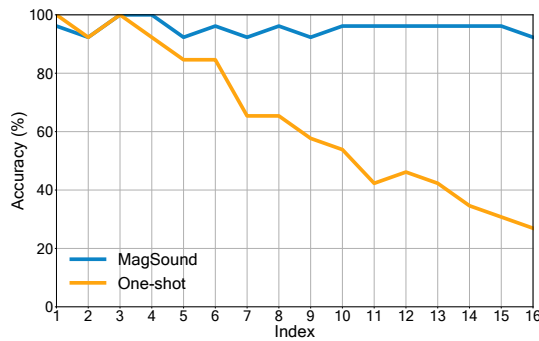


Fig. 35. Accuracy over time

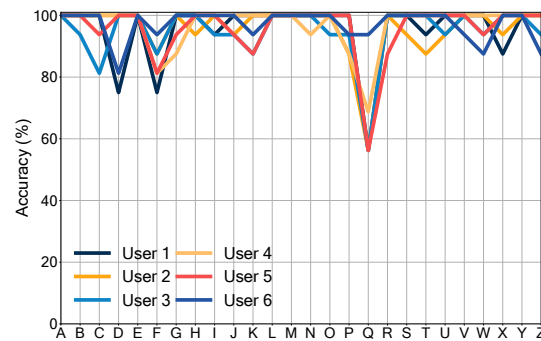


Fig. 36. Impact of different users

recognition of the IME. If the stroke order is different from the stroke order referenced by the IME, the error rate will increase. In general, the average accuracy of MagSound and One-shot is 95.43% and 63.70%, respectively, which means that MagSound brings an improvement of 49.81% to One-shot. We believe the improvements are based on the time stability of MagSound. To illustrate, we plot the trend of accuracy across all letters over time in Figure 35. We can see that the accuracy between MagSound and One-shot is close at the first 4 letters. Then, the accuracy of One-shot decreases over time while MagSound keeps the accuracy at a high level due to re-calibration. Further, we draw the trace of letter 'B' in Figure 33. We observe that the trace under MagSound maintains the relative shape, while the trace under One-shot get distorted since the tenth letter. Finally, we compare MagSound and One-shot on a real handwriting sentence in Figure 34. We can see that the trace of MagSound keep stable, while the trace of One-shot gets distorted. The results shed light on the potential of MagSound to enable HCI applications such as handwriting input.

5.4.3 Impact of Writing Habit. To evaluate the impact of different writing habits, we invite 6 volunteers to write uppercase letters using MagSound. Each volunteer writes 16 times in one run for each letter, and the tracking gets

Table 3. Processing delay of each component

Component	Delay (ms)
Magnetic coordinate	1.64
Acoustic coordinate	11.23
Calibration and fusion	0.42
Other	0.19

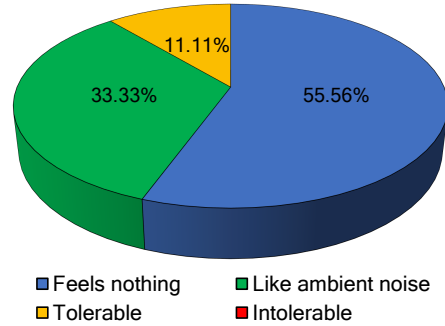


Fig. 37. Audibility to users

re-calibrated every four times. Consequently, we collect a total of 2496 letters from different users for evaluation. Figure 36 compares the accuracy of each volunteer on each letter. We observe that the curves are similar, which verifies the robustness of MagSound across different users.

5.5 System Performance and User Study

In MagSound, the average signal processing latency per frame is 13.48 ms. Since the frame length is 40 ms, the relatively low latency enables MagSound to run in real time. In Table 3, we draw the delay of each component in MagSound. We can see that the component of acoustic coordinate costs the most time, which is mainly caused by the padding and demodulation of the ZC sequence. Furthermore, the low latency of the calibration and fusion components validates the low computational cost of the optimization.

Finally, we invite 18 volunteers from age 22 to 47 to test the audibility of MagSound on different devices. The earphone transmits signals for 40-50 cm to the user, which follows the common usage of earphone tracking. We set four noise levels from low to high, including feels nothing, heard something but like normal noise, a bit noisy but tolerable and totally intolerable. As is shown in Figure 37, we find that none of the users feels intolerable to the acoustic signal transmitted by MagSound, and most volunteers feels nothing or can not distinguish the transmitted signal from ambient noise. The result is consistent with our preliminary study in Section 3.2.1. Therefore, the acoustic signal transmitted in MagSound would not disturb people's daily life. We attribute the result to the weak frequency response of wireless earphones.

6 DISCUSSION

6.1 Receiving Acoustic Signal Using Microphones in Wireless Earphones

Wireless earphones nowadays are usually equipped with microphones for telephone calls. Since the speaker on wireless earphones transmits weak signals, we also studied the possibility to take the smartphone as the transmitter and the earphone as the receiver. Nevertheless, we find that the built-in microphones on most COTS wireless earphones can only support sound signals with a frequency below 10 kHz [45]. One possible reason is that an upper limit of 10 kHz is enough for the designed applications, such telephone call or active noise cancellation. To transmit higher frequency signals, we choose the earphone as the transmitter and the smartphone as the receiver.

6.2 Splitting Strokes

For applications such as handwriting input, it is important to divide a writing trace into several strokes. We note that most wireless earphones have programmable buttons reserved for volume control or song switch. It is convenient for users to mark the beginning or end of a stroke by touching the button. However, the challenge is that users may touch the button by accident when holding the earphone. We leave this as future work to advance the prototype into practical application.

6.3 3D Earphone Tracking on COTS Smart Devices

Most of the COTS smart devices, such as smartphones and smartwatches, have one or two microphones, which means that we have at most two equations that we can use to determine the coordinates of the earphone. Thus, we evaluate the performance of MagSound mainly in the 2D tracking scenario. However, with the development of technology, when the microphones of smart devices are expanded to equal or more than three, the system can be easily adapted to the 3D tracking scenario in a manner similar to what we discussed in Section 4.3.3. Without loss of generality, supposing that we have three microphones on a smart device, we would have a system of three geometric equations:

$$\begin{cases} \text{dist}(\mathbf{P}, \mathbf{P}^1) = d_1 \\ \text{dist}(\mathbf{P}, \mathbf{P}^2) = d_2 \\ \text{dist}(\mathbf{P}, \mathbf{P}^3) = d_3 \end{cases}, \quad (17)$$

where the symbols share the same meaning as those in Section 4.3.3. To solve the earphone coordinate \mathbf{P} , the objective function to optimize become

$$\min_{\mathbf{P}} \left\{ (\text{dist}(\mathbf{P}, \mathbf{P}^1) - d_1)^2 + (\text{dist}(\mathbf{P}, \mathbf{P}^2) - d_2)^2 + (\text{dist}(\mathbf{P}, \mathbf{P}^3) - d_3)^2 \right\}. \quad (18)$$

By optimizing the objective function, we would obtain the 3D coordinate of the earphone. If we have more than three microphones, we can also utilize the extra microphones to improve the robustness to tracking failure [62].

Since the magnetometer in smart devices collects data from the three axes of x , y , and z , there have been some existing works [30, 69] that track a magnet in 3D space using the built-in magnetometer in COTS smart devices. Thus, 3D earphone tracking based on magnetic field strength still works in MagSound, which enables users to re-calibrate 3D acoustic signal based tracking. Nevertheless, a practical challenge is that collecting magnetic field strength fingerprints in a 3D space is more difficult than in a 2D plane. One possible solution is to employ off-the-shelf robots, *e.g.*, Makerbot [72], to place the earphone in a 3D space precisely.

6.4 Limitations

For some devices, such as AirPods Pro, we observe that its clock drift function changes rapidly, yielding 1D tracking error of meter-level in 240 seconds using one-shot calibration. We also find that Bluetooth Codecs may cause errors due to audio compression. Although MagSound can reduce the error into centimeter-level for AirPods Pro, the tracking system have to get re-calibrated frequently. Besides, due to the weak frequency response, the sensing range of earphone based tracking is limited to one meter. However, we argue that the sensing range is enough for potential applications.

7 CONCLUSION

In this paper, we present MagSound, which can re-calibrate wireless earphone tracking in one run without modified or specified hardware. Based on the analysis of the unique characteristics of earphones, we perform integrated magnetic and acoustic sensing and design a robust signal model for earphone sensing. The proposed MagSound combines the advantages of both magnetic sensing and acoustic sensing to address challenges such

as clock offset and low SNR in earphone tracking. We implement a prototype of MagSound and evaluate the system in applications such as 2D tracking and handwriting recognition. Comprehensive experiments manifest that MagSound can maintain the tracking accuracy at millimeter-level over time, and achieve 95.43% accuracy in handwriting recognition. We believe that MagSound can propel the wireless earphone into a pervasive acoustic sensing platform.

ACKNOWLEDGMENTS

We would like to thank our anonymous reviewers for their valuable comments. This work was supported in part by the National Natural Science Foundation of China under Grant 61872178, 62272223, 62272213, in part by the National Natural Science Foundation of China under Grant 61832005, in part by the Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, and in part by the Jiangsu High-level Innovation and Entrepreneurship (Shuangchuang) Program.

REFERENCES

- [1] Sudhakar A.N., Markandeya R., Srinivasa Rao B., Kumar Pandey Ajoy, and Kaushik D. 2022. Effect of alloying elements on the microstructure and mechanical properties of high chromium white cast iron and Ni-Hard iron. *Materials Today: Proceedings* 61 (2022), 1006–1014. <https://doi.org/10.1016/j.matpr.2021.10.284>
- [2] Android. 2022. Magnetometer accuracy in Android. https://developer.android.com/guide/topics/sensors/sensors_position
- [3] Inc Apple. 2019. AirPods Pro. <https://www.apple.com/shop/product/MLWK3AM/A/airpods-pro>
- [4] Glen Ballou. 2013. *Handbook for sound engineers*. Taylor & Francis.
- [5] Vladimir V. Bochkarev, Anna V. Shevlyakova, and Valery D. Solovjev. 2012. Average word length dynamics as indicator of cultural changes in society. arXiv:1208.6109 [cs.CL]
- [6] Yanling Bu, Lei Xie, Yafeng Yin, Chuyu Wang, Jingyi Ning, Jiannong Cao, and Sanglu Lu. 2022. Handwriting-Assistant: Reconstructing Continuous Strokes with Millimeter-Level Accuracy via Attachable Inertial Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 5, 4, Article 146 (dec 2022), 25 pages. <https://doi.org/10.1145/3494956>
- [7] Gaoshuai Cao, Kuang Yuan, Jie Xiong, Panlong Yang, Yubo Yan, Hao Zhou, and Xiang-Yang Li. 2020. EarphoneTrack: Involving Earphones into the Ecosystem of Acoustic Motion Tracking. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (SenSys '18)*. Association for Computing Machinery, New York, NY, USA, 95–108. <https://doi.org/10.1145/3384419.3430730>
- [8] Yifeng Cao, Ashutosh Dhekne, and Mostafa Ammar. 2021. ITrackU: Tracking a Pen-like Instrument via UWB-IMU Fusion. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (Virtual Event, Wisconsin) (MobiSys '21)*. Association for Computing Machinery, New York, NY, USA, 453–466. <https://doi.org/10.1145/3458864.3467885>
- [9] Dongyao Chen, Mingke Wang, Chenxi He, Qing Luo, Yasha Iravantchi, Alanson Sample, Kang G. Shin, and Xinbing Wang. 2021. MagX: Wearable, Untethered Hands Tracking with Passive Magnets. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (New Orleans, Louisiana) (MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 269–282. <https://doi.org/10.1145/3447993.3483260>
- [10] Ke-Yu Chen, Shwetak N. Patel, and Sean Keller. 2016. Finexus: Tracking Precise Motions of Multiple Fingertips Using Magnetic Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 1504–1514. <https://doi.org/10.1145/2858036.2858125>
- [11] Haiming Cheng and Wei Lou. 2022. PD-FMCW: Push the Limit of Device-Free Acoustic Sensing using Phase Difference in FMCW. *IEEE Transactions on Mobile Computing* (2022), 1–1. <https://doi.org/10.1109/TMC.2022.3162631>
- [12] Romit Roy Choudhury. 2021. Earable Computing: A New Area to Think About (*HotMobile '21*). Association for Computing Machinery, New York, NY, USA, 147–153. <https://doi.org/10.1145/3446382.3450216>
- [13] Tai L Chow. 2006. *Introduction to electromagnetic theory: a modern perspective*. Jones & Bartlett Learning.
- [14] International Electrotechnical Commission. 2014. Digital Audio Interface – Part1: General. IEC 60958-1.
- [15] Google Corporation. 2019. Google Pixel 4. <https://www.amazon.com/Google-GA01187-US-Pixel-Black-Unlocked/dp/B07YMNXL3>
- [16] Microsoft Corporation. 2022. Microsoft Whiteboard. <https://app.whiteboard.microsoft.com/>
- [17] Samsung Corporation. 2016. Samsung S7. <https://www.amazon.com/Samsung-Galaxy-G930A-32GB-Unlocked/dp/B07DHTFD1L>
- [18] Samsung Corporation. 2021. Samsung Galaxy Buds Pro. <https://www.samsung.com/global/galaxy/galaxy-buds-pro/>
- [19] Xiaomi Corporation. 2021. Redmi Buds 3. <https://www.mi.com/global/product/redmi-buds-3/>
- [20] Xiaomi Corporation. 2021. Xiaomi Redmi K40. <https://www.mi.com/redmik40>
- [21] Dian Ding, Lanqing Yang, Yi-Chao Chen, and Guangtao Xue. 2022. Handwriting Recognition System Leveraging Vibration Signal on Smartphones. *IEEE Transactions on Mobile Computing* (2022), 1–1. <https://doi.org/10.1109/TMC.2022.3148172>

- [22] Linfei Ge, Qian Zhang, Jin Zhang, and Qianyi Huang. 2020. Acoustic Strength-Based Motion Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 4, 4, Article 127 (dec 2020), 19 pages. <https://doi.org/10.1145/3432215>
- [23] Google. 2022. Google Input Tools. <https://www.google.com/inputtools/request>
- [24] Peter Grzybek. 2007. History and methodology of word length studies. *Contributions to the Science of Text and Language* (2007), 15–90.
- [25] Lawrence Alan Gust. 2006. Compact optical pointing apparatus and method. US Patent 7,102,617.
- [26] Anthony D Hall, James B Cunningham, Richard P Roache, and Julie W Cox. 1988. Factors affecting performance using touch-entry systems: Tactual recognition fields and system accuracy. *Journal of applied psychology* 73, 4 (1988), 711.
- [27] Christian Holz and Patrick Baudisch. 2010. The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (CHI '10). Association for Computing Machinery, New York, NY, USA, 581–590. <https://doi.org/10.1145/1753326.1753413>
- [28] Guan-Chyun Hsieh and James C Hung. 1996. Phase-locked loop techniques. A survey. *IEEE Transactions on industrial electronics* 43, 6 (1996), 609–615.
- [29] Xingman Huang and Xiaomei Wu. 2019. An Electromagnetic Tracking Method Based on Phase Difference Detection. *IEEE Transactions on Magnetics* 55, 9 (2019), 1–8. <https://doi.org/10.1109/TMAG.2019.2915264>
- [30] Sungjae Hwang, Andrea Bianchi, Myungwook Ahn, and Kwangyun Wohn. 2013. MagPen: Magnetically Driven Pen Interactions on and around Conventional Smartphones. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Munich, Germany) (*MobileHCI '13*). Association for Computing Machinery, New York, NY, USA, 412–415. <https://doi.org/10.1145/2493190.2493194>
- [31] F. Itakura. 1975. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23, 1 (1975), 67–72. <https://doi.org/10.1109/TASSP.1975.1162641>
- [32] jaakkopasanen. 2022. AutoEq. <https://github.com/jaakkopasanen/AutoEq/>
- [33] Ruinan Jin, Chao Cai, Tianping Deng, Qingxia Li, and Rong Zheng. 2021. MotionBeep: Enabling Fitness Game for Collocated Players With Acoustic-Enabled IoT Devices. *IEEE Internet of Things Journal* 8, 13 (2021), 10755–10765. <https://doi.org/10.1109/JIOT.2021.3050436>
- [34] Yincheng Jin, Yang Gao, Xiaotao Guo, Jun Wen, Zhengxiong Li, and Zhanpeng Jin. 2022. EarHealth: An Earphone-Based Acoustic Otoscope for Detection of Multiple Ear Diseases in Daily Life. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) (*MobiSys '22*). Association for Computing Machinery, New York, NY, USA, 397–408. <https://doi.org/10.1145/3498361.3538935>
- [35] WT Kaune, MC Miller, MS Linet, EE Hatch, RA Kleinerman, S Wacholder, AH Mohr, RE Tarone, and C Haines. 2002. Magnetic fields produced by hand held hair dryers, stereo headsets, home sewing machines, and electric clocks. *Bioelectromagnetics: Journal of the Bioelectromagnetics Society, The Society for Physical Regulation in Biology and Medicine, The European Bioelectromagnetics Association* 23, 1 (2002), 14–25.
- [36] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89. <https://doi.org/10.1109/MPRV.2018.03367740>
- [37] Nokia Bell Lab. 2022. eSense Website. <https://www.esense.io/>
- [38] Roberto Lanza and Antonio Meloni. 2006. *The Earth's Magnetic Field*. Springer.
- [39] Sinjin Lee, Kevin Fu, Tadayoshi Kohno, Benjamin Ransford, and William H. Maisel. 2009. Clinically significant magnetic interference of implanted cardiac devices by portable headphones. *Heart Rhythm* 6, 10 (2009), 1432–1436. <https://doi.org/10.1016/j.hrthm.2009.07.003>
- [40] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. 2022. EarIO: A Low-Power Acoustic Sensing Earable for Continuously Tracking Detailed Facial Movements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 6, 2, Article 62 (jul 2022), 24 pages. <https://doi.org/10.1145/3534621>
- [41] Zeshui Li, Haipeng Dai, Wei Wang, Alex X. Liu, and Guihai Chen. 2018. PCIAS: Precise and Contactless Measurement of Instantaneous Angular Speed Using a Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 4, Article 177 (dec 2018), 24 pages. <https://doi.org/10.1145/3287055>
- [42] Qianru Liao, Yongzhi Huang, Yandao Huang, Yuheng Zhong, Huitong Jin, and Kaishun Wu. 2022. MagEar: Eavesdropping via Audio Recovery Using Magnetic Side Channel. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) (*MobiSys '22*). Association for Computing Machinery, New York, NY, USA, 371–383. <https://doi.org/10.1145/3498361.3538921>
- [43] Kang Ling, Haipeng Dai, Yuntang Liu, Alex X. Liu, Wei Wang, and Qing Gu. 2022. UltraGesture: Fine-Grained Gesture Sensing and Recognition. *IEEE Transactions on Mobile Computing* 21, 7 (2022), 2620–2636. <https://doi.org/10.1109/TMC.2020.3037241>
- [44] Jianwei Liu, Wenfan Song, Leming Shen, Jinsong Han, Xian Xu, and Kui Ren. 2021. MandiPass: Secure and Usable User Authentication via Earphone IMU. In *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS '21)*. 674–684. <https://doi.org/10.1109/ICDCS51616.2021.00070>
- [45] Ruofeng Liu, Wenjun Jiang, and Xun Chen. 2021. Acoustic Ruler Using Wireless Earbud. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services* (Virtual Event, Wisconsin) (*MobiSys '21*). Association for Computing Machinery, New York, NY, USA, 501–502. <https://doi.org/10.1145/3458864.3466905>

- [46] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-Precision Acoustic Motion Tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (New York City, New York) (*MobiCom '16*). Association for Computing Machinery, New York, NY, USA, 69–81. <https://doi.org/10.1145/2973750.2973755>
- [47] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. FingerIO: Using Active Sonar for Fine-Grained Finger Tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). Association for Computing Machinery, New York, NY, USA, 1515–1525. <https://doi.org/10.1145/2858036.2858580>
- [48] International Commission on Non-Ionizing Radiation Protection et al. 1998. Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health physics* 74, 4 (1998), 494–522. <https://doi.org/10.1097/hp.0000000000001210>
- [49] Farshid Salemi Parizi, Eric Whitmire, and Shwetak Patel. 2019. AuraRing: Precise Electromagnetic Finger Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 3, 4, Article 150 (dec 2019), 28 pages. <https://doi.org/10.1145/3369831>
- [50] B.M. Popovic. 1992. Generalized chirp-like polyphase sequences with optimum correlation properties. *IEEE Transactions on Information Theory* 38, 4 (1992), 1406–1409. <https://doi.org/10.1109/18.144727>
- [51] Michael JD Powell. 1964. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal* 7, 2 (1964), 155–162. <https://doi.org/10.1093/comjnl/7.2.155>
- [52] Michael JD Powell. 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*. Springer, 51–67.
- [53] RTINGS. 2019. RTINGS website. <https://www.rtings.com/headphones/tests/sound-quality/raw-frequency-response>
- [54] W M Saslow. 1991. How a superconductor supports a magnet, how magnetically soft² iron attracts a magnet, and eddy currents for the uninitiated. *American Journal of Physics; (USA)* 59, 1 (1 1991). <https://doi.org/10.1119/1.16700>
- [55] Thomas Schmid, Zainul Charbiwala, Jonathan Friedman, Young H. Cho, and Mani B. Srivastava. 2008. Exploiting Manufacturing Variations for Compensating Environment-Induced Clock Drift in Time Synchronization (*SIGMETRICS '08*). Association for Computing Machinery, New York, NY, USA, 97–108. <https://doi.org/10.1145/1375457.1375469>
- [56] Katie A Siek, Yvonne Rogers, and Kay H Connelly. 2005. Fat finger worries: how older and younger users physically interact with PDAs. In *IFIP Conference on Human-Computer Interaction*. Springer, 267–280. https://doi.org/10.1007/11555261_24
- [57] Steven W Smith et al. 1997. The scientist and engineer’s guide to digital signal processing.
- [58] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. VSkin: Sensing Touch Gestures on Surfaces of Mobile Devices Using Acoustic Signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking* (New Delhi, India) (*MobiCom '18*). Association for Computing Machinery, New York, NY, USA, 591–605. <https://doi.org/10.1145/3241539.3241568>
- [59] Acoustical Terminology. 2006. American national standard. *ANSI S1* (2006), 1–1994.
- [60] Daniel Vogel and Patrick Baudisch. 2007. Shift: A Technique for Operating Pen-Based Interfaces Using Touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '07*). Association for Computing Machinery, New York, NY, USA, 657–666. <https://doi.org/10.1145/1240624.1240727>
- [61] Haoran Wan, Lei Wang, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, and Wei Wang. 2022. VECTOR: Velocity Based Temperature-Field Monitoring with Distributed Acoustic Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 6, 3, Article 144 (sep 2022), 28 pages. <https://doi.org/10.1145/3550336>
- [62] Anran Wang and Shyamnath Gollakota. 2019. MilliSonic: Pushing the Limits of Acoustic Motion Tracking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3290605.3300248>
- [63] Lei Wang, Xiang Zhang, Yuanshuang Jiang, Yong Zhang, Chenren Xu, Ruiyang Gao, and Daqing Zhang. 2021. Watching Your Phone’s Back: Gesture Recognition by Sensing Acoustical Structure-Borne Propagation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 5, 2, Article 82 (jun 2021), 26 pages. <https://doi.org/10.1145/3463522>
- [64] Mingke Wang, Qing Luo, Yasha Iravantchi, Xiaomeng Chen, Alanson Sample, Kang G. Shin, Xiaohua Tian, Xinbing Wang, and Dongyao Chen. 2022. Automatic Calibration of Magnetic Tracking. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking* (Sydney, NSW, Australia) (*MobiCom '22*). Association for Computing Machinery, New York, NY, USA, 391–404. <https://doi.org/10.1145/3495243.3558760>
- [65] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-Free Gesture Tracking Using Acoustic Signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking* (New York City, New York) (*MobiCom '16*). Association for Computing Machinery, New York, NY, USA, 82–94. <https://doi.org/10.1145/2973750.2973764>
- [66] Yuntao Wang, Jiexin Ding, Ishan Chatterjee, Farshid Salemi Parizi, Yuzhou Zhuang, Yukang Yan, Shwetak Patel, and Yuanchun Shi. 2022. FaceOri: Tracking Head Position and Orientation Using Ultrasonic Ranging on Earphones. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). Association for Computing Machinery, New York, NY, USA, Article 290, 12 pages. <https://doi.org/10.1145/3491102.3517698>
- [67] Dan-Ping Xu, Han-Wen Lu, Yuan-Wu Jiang, Hyung-Kyu Kim, Joong-Hak Kwon, and Sang-Moon Hwang. 2017. Analysis of Sound Pressure Level of a Balanced Armature Receiver Considering Coupling Effects. *IEEE Access* 5 (2017), 8930–8939. <https://doi.org/10.1109/ACCESS.2017.2696565>

- [68] Zhijian Yang, Yu-Lin Wei, Sheng Shen, and Romit Roy Choudhury. 2020. Ear-AR: Indoor Acoustic Augmented Reality on Earphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (London, United Kingdom) (MobiCom '20)*. Association for Computing Machinery, New York, NY, USA, Article 56, 14 pages. <https://doi.org/10.1145/3372224.3419213>
- [69] Sang Ho Yoon, Ke Huo, and Karthik Ramani. 2016. TMotion: Embedded 3D Mobile Input Using Magnetic Sensing Technique. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (Eindhoven, Netherlands) (TEI '16)*. Association for Computing Machinery, New York, NY, USA, 21–29. <https://doi.org/10.1145/2839462.2839463>
- [70] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-Grained Acoustic-Based Device-Free Tracking. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (Niagara Falls, New York, USA) (MobiSys '17)*. Association for Computing Machinery, New York, NY, USA, 15–28. <https://doi.org/10.1145/3081333.3081356>
- [71] Hans-Jurgen Zepernick and Adolf Finger. 2013. *Pseudo random signal processing: theory and application*. John Wiley & Sons.
- [72] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan, and Gregory D. Abowd. 2017. SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 1, 2, Article 30 (jun 2017), 25 pages. <https://doi.org/10.1145/3090095>
- [73] Qian Zhang, Dong Wang, Run Zhao, Yinggang Yu, and JiaZhen Jing. 2021. Write, Attend and Spell: Streaming End-to-End Free-Style Handwriting Recognition Using Smartwatches. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 5, 3, Article 138 (sep 2021), 25 pages. <https://doi.org/10.1145/3478100>
- [74] Zengbin Zhang, David Chu, Xiaomeng Chen, and Thomas Moscibroda. 2012. SwordFight: Enabling a New Class of Phone-to-Phone Action Games on Commodity Phones. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (Low Wood Bay, Lake District, UK) (MobiSys '12)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/2307636.2307638>
- [75] Tianyue Zheng, Chao Cai, Zhe Chen, and Jun Luo. 2022. Sound of Motion: Real-time Wrist Tracking with A Smart Watch-Phone Pair. In *IEEE Conference on Computer Communications (INFOCOM '22)*, 110–119. <https://doi.org/10.1109/INFOCOM48880.2022.9796731>