# Accurate Ranging on Acoustic-enabled IoT Devices

Chao Cai, *Student Member, IEEE,* Menglan Hu, Xiaoqiang Ma, *Member, IEEE,* Kai Peng, and Jiangchuan Liu, *Fellow, IEEE*

***Abstract*—The enabling Internet-of-Things technology has inspired many innovative sensing mechanisms by repurposing the onboard sensors. Leveraging the built-in acoustic sensors for ranging is among one of the interesting applications. However, among the few studies on acoustic ranging, the one-way sensing method suffers from synchronization errors and requires cumbersome kernel modifications; the other two-way approaches overcome these shortcomings, but they are sensitive to system delays. In this case, this paper proposes a novel lightweight one-way sensing paradigm without the above drawbacks. The key insight of our work is to perform ranging by estimating the propagation time of acoustic signals via Linear Frequency Modulation (LFM) signal mixing. Such a signal mix operation can translate range estimation into fine-grain frequency estimation, thereby enhancing ranging accuracy. In addition, our system can have multiple receivers co-exist and thus the measurement dimensions are boosted. We have implemented and evaluated our system prototype in real-world settings. The prototype demonstrated centimeter-level ranging performance.**

***Index Terms*—Acoustic, sensor calibration, sensor signal processing, ranging, linear frequency modulated signal, internet-of-things.**

## I. Introduction

The Internet-of-Things (IoT) technology [22], [36], [12] enables physical devices to sense, process, and share data in a connected network. It facilitates everyday objects with digital intelligence [21], [44], [31], leading to many novel applications that bring great convenience. The IoT devices, in particular, mobile devices, have been rapidly changing our lives. Mobile IoT platforms like smartphones and wearables have become an indispensable part of our daily life. Recent statistics [4] show that the number of smartphones will reach 5.07 billion in 2019. Beyond communication, mobile IoT devices like smartphones have already become new sensing platforms embedded with rich built-in sensors, e.g., cameras, accelerometers, gyroscopes, barometers, and acoustic sensors. These on-board sensors have enabled a myriad of mobile computing supported applications [35], [34], [27]. Meanwhile, there have been increasing research interests on exploiting the on-board sensors besides their primary usages, enhancing the functionality. For instance, the default function for a camera is to take photos, but researchers find that cameras can also be used for visible light communication [52]. Also, accelerometers and gyroscopes are mainly used for attitude calculation, but they can also be helpful in activity recognition [23].

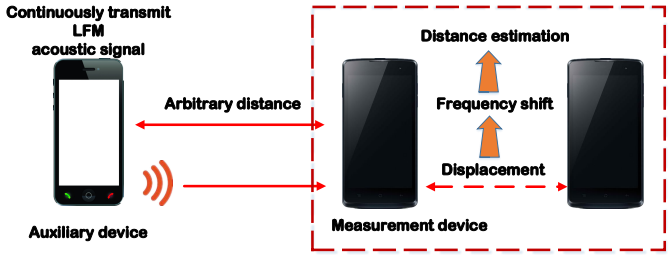



Fig. 1: The auxiliary device repeatedly triggers LFM signal and silent slot interactively. The measurement device tracks the frequency shift to perform ranging.

While acoustic sensors on commodity IoT devices are originally used to record and play audio signals, in this paper, we propose to use them to perform ranging. Such kinds of applications not only make acoustic-enabled IoT devices an alternative to traditional measurement tools, but also enable lots of context-aware computing [10] such as distance/size measurements [35], [46], tracking [34], [49], [55], [47], localization [24], indoor outdoor detection [8], [11], efficient network management [32], and content sharing [13], [14].

There have been some initial attempts to perform ranging via acoustic sensors. Peng et al. [35] presented the first work which obtained inter-device distance by estimating the travel time of acoustic signals. A commercial product [42] for iOS devices also followed this scheme. However, this scheme cannot work without the involvement of a central server. Besides, the two-way sensing approach in this scheme though avoids tight synchronization requirements but can be impaired by the unpredictable system delay, leading to severe ranging errors. Another stand-alone work [46] sidestepped the above problem by scheduling associated tasks at the kernel space. This approach synchronizes the devices through radio signals and performs ranging by measuring the time-of-flight of the acoustic signals. However, this one-way sensing scheme needs kernel hacking which is non-trivial and is device-dependent, limiting wide-scale adoption. In addition, the reported performance is inaccurate, and the ranging errors are up to 20 cm. There also exist alternative solutions on mobile platforms (e.g., iPin[3]), but such solutions demand additional hardware.

To this end, this paper aims to design an accurate stand-alone ranging scheme on commodity mobile devices with embedded acoustic sensors, namely, microphones and speakers. Also, the new scheme shall enjoy easy installation with no extra burden of kernel modifications and be free of the fragile synchronization tricks. To achieve the design goal,

C. Cai, M. Hu, X. Ma, and K. Peng are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. E-mail: {cchust, humenglan, maxiaoqiang, pkhust}@hust.edu.cn. K. Peng is the corresponding author.

J. Liu is with the College of Natural Resources and Environment, South China Agricultural University, Guangzhou, China. E-mail: csljc@ieee.org.

This work was supported in part by the National Natural Science Foundation of China under Grant 61502193, 61872415, 61872416, 61671216, 61702204, 61871436, 61471408, and 91538203.

we adopt a one-way sensing mechanism. The basic idea is to perform ranging by estimating the propagation time of acoustic signals via Linear Frequency Modulated (LFM) signal mixing [26], [27]. The signal mixing approach correlates the displacement to the proportional shifted frequency at the receiver end as depicted in Fig. 1. Hence, it eliminates synchronization requirements. Such a scheme can also lead to simultaneous multiple dimension measurements. Moreover, the signal mixing can translate the time estimation which is constrained by the sampling rate into fine grain frequency estimation, thus achieving much higher accuracy.

To implement the idea, we shall tackle several fundamental challenges. First, the signal mixing operation requires the transmitter and receiver to share the same clock. But due to the instability of local oscillator, the sampling clock drifts even within a few seconds [43], resulting in slow frequency drifting, which finally leads to ranging errors. In this case, a compensation technique is desired. The second challenge stems from degraded SNR of signal mixing operation at varying distances. In our system, the auxiliary device incessantly transmits LFM signals with "silent periods" in between and the measurement device periodically extracts a fixed number of samples for distance estimation. At a certain distance, the extracted signal would be a partial LFM signal, resulting in low frequency peaks and posing challenges for frequency estimation. We refer this phenomenon to as degraded SNR. The SNR degradation is further enhanced due to the fact that the acoustic signal experiences path losses. As a result, the SNR degradation inevitably introduces distance estimation errors. The third challenge is the multipath effect. Ranging requires Line-of-Sight (LOS) signals. However, multiple acoustic echoes in indoor environments can deteriorate the performance of LOS signal detection thus result in ranging errors. The problem can be more complicated when multiple echoes are coherent.

In the design of our ranging scheme, we have addressed the aforementioned challenges. First, to tackle the slow frequency drifting problem, we propose an Orthogonal Frequency Division Multiplexing (OFDM) [33], [9], [15] based solution. The slow frequency drifting (also called Sampling Frequency Offset (SFO) [48]) manifests itself as the slope of phase across each subcarrier [37]. Therefore, by tracking the change rate of the phase slope across different subcarriers, we can estimate the frequency drifting rate and compensate for it. Second, we propose an adaptive sampling mechanism which adaptively chooses the start point for a sampling period, keeping the relatively same amount of LFM samples among different distances. Accordingly, the SNR degradation is only affected by path losses. Third, we propose to use the Modified MUltiple SIgnal Classification (MMUSIC) [17] algorithm which decorrelates the coherence among multiple echoes to resolve the multipath effect and thus enhance the ranging accuracy. We summarize our contributions as follows:

- We propose a standalone software-oriented acoustic ruler on commodity hardware. It adopts one-way sensing and can simultaneously measure multiple dimensions.
- We design a lightweight OFDM-based method to compensate the ranging errors. The proposed approach achieves much higher resolution in frequency drift estimation.
- We develop an adaptive mechanism to improve SNR under dynamic scenarios.
- We implement and evaluate a system prototype on commercial-off-the-shelf devices, demonstrating centimeter-level ranging accuracy.

The rest of the paper is organized as follows: Section II discusses related work. Section III presents the system overview. In Section IV, we first outline the basic principles of the ruler, then elaborate the OFDM based frequency estimation method in the calibration phase, and finally describe the approaches to enhance the robustness in the distance estimation phase. Section V evaluates the performance of our ruler system under various settings, with conclusions following in Section VI.

## II. Related Work

With the fast development of mobile technology, many sensing tasks have been deployed on the mobile terminals, leading to tremendous interesting applications. Leveraging acoustic sensors for range-based applications are among one of them.

BeepBeep [35] was the pioneering work that leveraged acoustic signal for accurate ranging. It obtains inter-device distances by estimating the propagated time of acoustic signals. BeepBeep avoids tight synchronization via a two-way sensing paradigm. In BeepBeep, one device first emits a chirp signal. Upon detection, the other device waits for an arbitrary period of time and emits another chirp signal. Both transceivers then calculats the time difference between the events of transmission and detection locally. A central server is involved to compute the final results. BeepBeep reported centimeter-level ranging accuracy. However, the performance of BeepBeep can be distorted by system delay which exhibits unpredictability. RFBeep [46] bypassed the problem by scheduling tasks directly at the kernel space. RFBeep exploits the differences in propagation characteristics between radio and acoustic signals to perform ranging. Nevertheless, this scheme needs kernel hacking which is non-trivial and is device-dependent. The reported accuracy is also not satisfactory as ranging errors can be up to 20 cm.

Range information can be used for precise tracking. AAMouse [54] was a smartphone-based tracking system via Doppler effect estimation. However, AAMouse suffers from integration errors in the long run. In this case, [26] and [27] solved it via a combination of Doppler effect estimation and signal mixing operation. Similar to our paper, they also adopted one-way sensing paradigm and achieved accurate tracking with millimeter-level accuracy. However, they applied traditional Frequency Modulated Continuous Wave (FMCW) waveform which was subject to Inter-Symbol-Interference (ISI). Also, the systems in [26], [27] used traditional MUSIC which would perform worse when multiple reflections are coherent. In contrast, our work has addressed these issues. We mitigate ISI by inserting guard interval between successive transmissions. Moreover, we de-correlate the coherence between multiple echoes by applying MMUSIC.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2879371, IEEE Internet of Things Journal



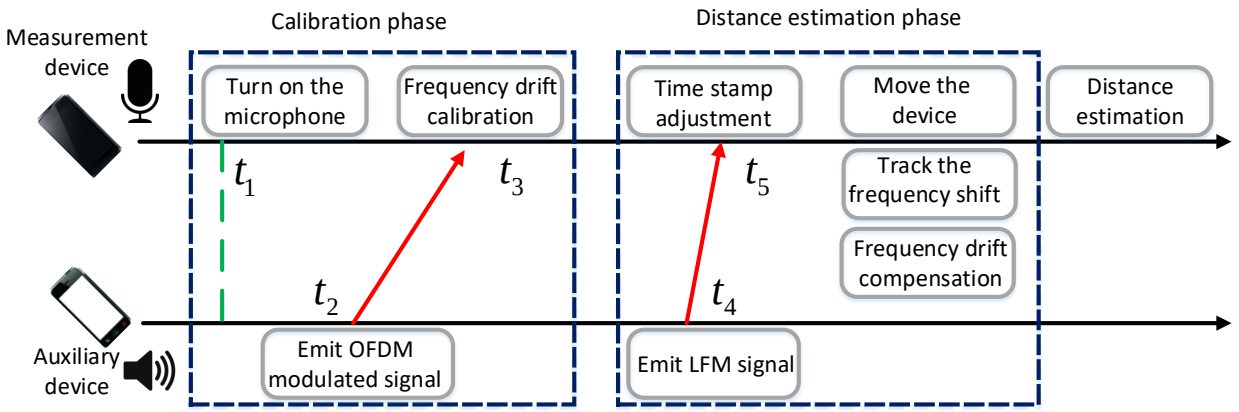



Fig. 2: Measurement procedure.

Range-based localization using acoustic sensing raises much interest recently. The authors in [24] built a localization network based on acoustic [35] and WiFi signals. Ranges among multiple smartphone pairs were utilized to form a rigid graph, and the WiFi signal strength was used to find the best-matched location. More range-based localization systems [25], [19], [18], [50] were built on Time-of-Arrival (ToA) [25] or Time-Difference-of-Arrival (TDoA) estimation [19], [18], [50], either synchronously [25], [19], [18] or asynchronously [50]. These systems usually involved custom-built anchors but often achieved high precision.

Around device interaction such as device-free gesture tracking is also based on range information. FingerIO [34] was a recently proposed device-free gesture tracking system. It utilized the mobile devices as an active sonar that was capable of tracking the moving fingers with 7 mm accuracy at a frame rate of 169 fps. LLAP [49] was another work for gesture tracking. Built on the coherent detection structure, LLAP employed the phase of acoustic echoes for finger localization and tracking. LLAP reported a tracking accuracy of 3.5 mm for 1D hand movement and 4.57 mm for 2D drawing with less than 15 ms latency. However, both FingerIO and LLAP are not resilient to nearby interferences. To cope with these inferences, another work Strata [55] applied more robust channel estimation methods. It also adopted coherent detector structure and outperformed the above two studies.

## III. Acoustic Ruler System Overview

Our acoustic ruler adopts one-way sensing paradigm. It is free of any fragile synchronization tricks which can be impaired by the uncertainty of system delay. Moreover, our approach requires no extra burden of kernel modifications and thus can be easily implemented across different platforms. Meanwhile, it is a stand-alone application which enables multiple dimension measurements simultaneously. The proposed scheme enjoys simple design and entails two mobile devices: an auxiliary device and a measurement device. The auxiliary device is responsible for emitting specially modulated signals. The measurement device processes the received signal, compensates the frequency drift, and estimates the distance on its own. The workflow of the system is composed of two phases: the *calibration phase* and the *distance estimation phase* as depicted in Fig. 2.

The calibration phase starts with the microphone of the measurement device turned on at time $t_1$. At time $t_2$, the auxiliary device beams an OFDM signal. This OFDM signal, which is detected at $t_3$, lasts for three seconds and we insert a Zadoff-Chu (ZC [41]) sequence at the beginning for precise synchronization. Since this OFDM signal is known to the receiver, the measurement device can estimate the drifting rate by inspecting slope change across each subcarrier. We refer this step as calibration phase. The calibration phase may not be needed for short-term measurements which last for seconds, but it is compulsory to improve ranging accuracy at a long run. It is also worth noting that the calibration phase is only needed to be done once for a specific transceiver pair.

After calibration (at time $t_4$), signals emitted from the auxiliary device alternate between LFM and the "silent period" (no signal transmission). The measurement device first adjusts the sampling time stamp to ensure that the extracted samples in each cycle contain a LFM signal. This adjustment is achieved via correlations and required only once. In addition, the performance of this step incurs no impact on the final ranging performance. After the time stamp adjustment, we can move the measurement device for distance estimation. When the measurement device moves, it extracts the same amount of samples in each cycle and mixes them with the known LFM signal. A fast Fourier transform is then applied to inspect the major frequency component of the mixed results. By accurately tracking the frequency shift, the range information can be obtained.

Our proposed acoustic ruler design enjoys multifold flexibility. During the distance estimation phase, both devices can move. In addition, multiple measurement devices can co-exist, which implement simultaneous multiple dimensional measurements. Further, the signal generation block on the auxiliary device can be replaced by playing already-generated audio files. In this case, speaker-only devices can serve as the auxiliary device. Likewise, microphone-only devices can function as the measurement device. Moreover, various devices, including smartphones, wearables, pads, laptops and even traditional PCs can be flexibly utilized to implement our

design.

## IV. Acoustic Ruler Design

### A. Principal of acoustic ruler

The fundamental principle of the acoustic ruler is that it converts distances to the proportional frequency shifts via LFM signal mixing operation. The LFM signal, which is commonly known as chirp signal, can be represented as $s = \cos\left(2\pi f_{\min} t + \pi k t^2\right)$, where $f_{min}$ is the initial frequency and $k$ denotes the modulation coefficient. The modulation coefficient can be further obtained by $k = \frac{B}{T}$, where $B$ is the bandwidth and $T$ is the duration. When this LFM signal travels through the air and reaches the receiver end, it experiences a propagation delay $\Delta t$ and a path loss $\alpha$. Thus, at the receiver end, the signal becomes:

$$r = \alpha \cos\left(2\pi f_{\min}\left(t - \Delta t\right) + \pi k (t - \Delta t)^2\right) \tag{1}$$

The receiver mixed the delayed and attenuated version of $r$ with the known signal $s$. That is:

$$\begin{aligned} r_{\text{mix}} &= r \cdot s \\ &= \alpha \cos\left(2\pi f_{\min}\left(t - \Delta t\right) + \pi k (t - \Delta t)^2\right) \cdot \\ &\qquad \cos\left(2\pi f_{\min} t + \pi k t^2\right) \\ &= \tfrac{\alpha}{2} \cos\left(2\pi f_{\min} \Delta t + \pi k \left(2t\Delta t - \Delta t^2\right)\right) + (terms\ with \\ &\qquad high\ frequency) \end{aligned} \tag{2}$$

After a low pass filter, the mixed result becomes:

$$r_{filtered} = \frac{\alpha}{2} \cos\left(2\pi f_{\min} \Delta t + \pi k \left(2t\Delta t - \Delta t^2\right)\right) \tag{3}$$

We assume that the measurement device is at distance $d$ from the initial location. Then $\Delta t$ can be given by $\Delta t = \frac{d}{c}$, where $c$ is the velocity of sound. Plugging $\Delta t$ into Eq. 3 we can get:

$$r_{filtered} = \frac{\alpha}{2} \cos\left(2\pi f_{\min} \frac{d}{c} + \pi k \left(2t\frac{d}{c} - \frac{d}{c^2}^2\right)\right) \tag{4}$$

We now look into the frequency components of Eq. 4, which can be obtained by taking the derivative of the phase component in Eq. 4 with respect to $t$. Via the above operation, all the constant terms are removed and the remaining frequency component is obtained as $f$:

$$f = \frac{1}{2\pi}\frac{\partial \theta}{\partial t} = \frac{kd}{c} \tag{5}$$

Accordingly, the distance can be given by: $d = \frac{fc}{k} = \frac{fcT}{B}$. In our system, the LFM signal spans $4$ kHz bandwidth with a duration of $T = 0.04$ s. If we have $1$ Hz frequency estimation resolution, the distance resolution would be up to $\frac{1 \times 340 \times 0.04}{4000} = 3.4$ mm, which is accurate enough for most daily use.

The signal triggered at the transmitter side is designed carefully. We emit the LFM signal periodically but keep the channel silent for a while as depicted in Fig. 3. Such signal design is different from commonly adopted FMCW [26], [27] which is composed of LFM signals in tandem. Our scheme has multiple benefits. First, the "silent period" can serve as a guard interval to achieve high SNR. As the acoustic speakers have

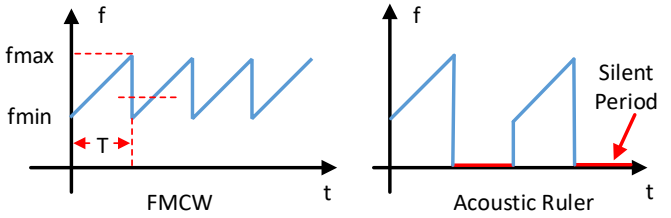



Fig. 3: Tradition FMCW and the signal type of acoustic ruler.

the "ringing" effect [38], [33] thus may cause inter-symbol-interference if LFM signals are incessantly triggered. Second, it can mitigate the multipath effect. The underlying fact is that the signal strength of acoustic reflections experiences a 20dB loss after 25ms [45]. Therefore, a guard interval can wait for any reverberations to decay significantly.

Our acoustic ruler translates frequency shifts to displacements. Thus, the ranging accuracy highly depends on the granularity of the frequency estimation method. In our system, we apply the Fast Fourier Transform (FFT) approach, the accuracy of which depends on the number of samples. However, in each cycle of the measurement phase, we only extract $1920$ samples. Directly performing FFT on such limited number of measurements achieves a $\frac{48000}{1920} = 25$Hz resolution which is not accurate enough. Consequently, we pad zero samples to the original ones, making it a pseudo-48000-sequence and thus achieving 1Hz resolution.

The above principal assumes that the transmitter and receiver (i.e., the auxiliary device and the measurement device) share the same clock, which may not hold in reality. The clock shifts slowly, leading to slow frequency drifting problem. In the next section, we will present how to compensate for this problem.

### B. Calibration Phase

The goal of calibration is to compensate ranging errors introduced by SFO [16], [26], [29], [30], [43]. SFO, incurred by the instability of local oscillator, is a common problem in communication systems. It exhibits unpredictability and generates accumulation error. SFO provokes ranging errors by distorting the extracted waveform at the receiver end. Normally, the waveform in each cycle is shaped by $f_s \cdot T = 1920$ samples when $f_s = 48000$ Hz and $T = 0.04$ s. However, SFO can disrupt the equilibrium state, making the waveform be shaped by less or more samples. Consequently, this distortion is equivalent to slowly shift our sampling window, result in frequency shift, thus leading to distance estimation error. The errors caused by SFO gradually increase [26], [43] if not compensated, leading to intolerable errors in the long run.

Intuitively, we can compare the time experienced by a set of samples with the expected value in the long run [43]. However, such a method requires a precise clock and is time-consuming. In contrast, our proposed OFDM-based estimation method can "magnify" the changes caused by SFO and make it detectable in sub-sample time. To help understand how it works, we first present some basic knowledge on OFDM.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2879371, IEEE Internet of Things Journal



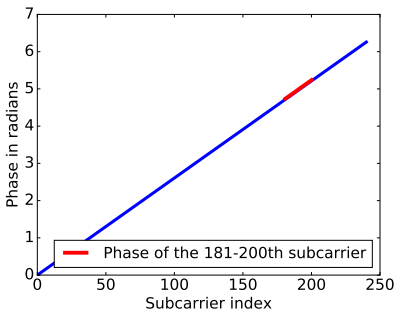



Fig. 4: Phase shift across each subcarrier when one sample drift happens.

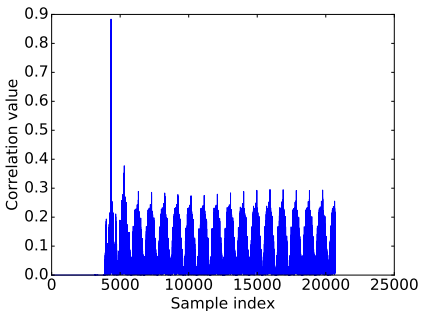



Fig. 5: Precisely synchronization using ZC signal.

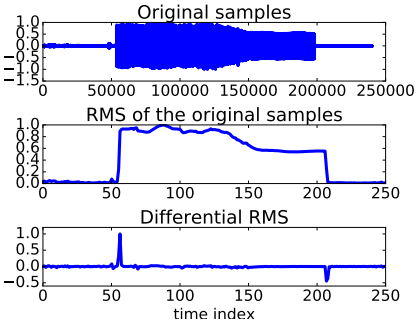



Fig. 6: Signal processing at receiver end.

Orthogonal Frequency Division Multiplexing (OFDM) [9], [15] is a highly efficient modulation technique that is commonly used in modern digital wireless communication systems including LTE and WiFi. The whole structure of the OFDM-based communication system is quite complicated and cannot be addressed in this paper. As a result, to simplify the explanation, we mainly focus on the contents that are most relevant to our proposal. OFDM deploys the data bits on orthogonal subcarriers and transmit them concurrently. Such an operation is efficiently achieved by employing the Fourier transform. Assuming we have $N$ subcarriers, each subcarrier carries the data bit $X_i, 1 \leq i \leq n$. The ready-to-transmit time-domain OFDM signal is generated by performing the inverse fast Fourier transform (IFFT) over the data bits:

$$\mathrm{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{i2\pi kn/N}, k = 0, 1, ..., N-1, \quad (6)$$

where $x_k$ is the time domain discrete samples that are ready to be transmitted. At the receiver side, the data bits can be recovered by performing FFT:

$$\mathrm{X}_n = \sum_{n=0}^{N-1} x_k e^{-i2\pi kn/N}, n = 0, 1, ..., N-1 \quad (7)$$

The above process assumes that the transceivers are perfectly synchronized. However, in practice, this may not hold true. As a result, sampling offset errors would appear. To address this problem, a common practice is to pad cyclic suffice at the end of the OFDM symbol [9], [15]. The cyclic suffice is a copy of the first few data bits in $X_n$. To understand why it works, say the received symbol is offset by $W$ samples, the new time-domain sequence can then be represented by $X_{(k+W)modN}$. Now performing FFT on the offset samples, we then obtain the recovered bits by:

$$\begin{aligned} \mathrm{x}_n^W &= \sum_{n=0}^{N-1} x_{(k+W) \bmod N} e^{-i2\pi kn/N} \\ &= X_n e^{i2\pi Wn/N} \end{aligned} \quad (8)$$

From Eq. 8, we can observe that the original data bits can be recovered but with additional phase shift that depends on the offset $W$. Moreover, the additional phase increases linearly with the subcarrier index. For example, one sample offset error leads to the phase shifting linearly from 0 to $2\pi$ across the subcarriers as depicted in Fig. 4. More generally, $W$ sample offsets result in the phase shifting linearly from 0 to $2W\pi$ across the subcarriers.

In summary, if the data bits are known to the receiver in prior, the offset error $W$ can then be obtained by inspecting the phase shift across each subcarrier. Furthermore, $W$ can be fractional which may be caused by the SFO since SFO can introduce an equivalent minor shifting effect as we stated previously. Accordingly, SFO can be assessed and if considering time, the shifting rate can also be obtained. The above OFDM properties serve as the fundamental principle for our estimation strategy.

However, implementing a standard acoustic-based OFDM system is nontrivial. A standard OFDM system [28] involves too many components and a straightforward implementation on mobile devices may be impossible. In addition, some of the signal processing modules [40] like Carrier Frequency Offset (CFO) correction and Forward Error Correction (FEC) may be unnecessary. In this case we propose a simplified structure, which is introduced as follows.

Our OFDM symbol is designed as follows: An identical PN sequence is consecutively repeated with a Zadoff-Chu (ZC) sequence [41] inserted at the head of it as depicted in Fig. 7. The structure of PN sequence repetition is equivalent to add cyclic prefix or cyclic suffix. The use of PN sequence here is to avoid high Peak-to-Average Power Ratio (PAPR) [9], [15], which may saturate the device thus distort the signal. The Generator polynomial parameter for the PN sequence is $z^6 + z^4 + z^2 + 1$, by which we generate known messages to the receiver. The message is Binary Phase Shift keying (BPSK) modulated and in total occupies 3 seconds. Zadoff-Chu sequence is used for precise synchronization. ZC sequence serves as the Primary Synchronization Sequence (PSS) in

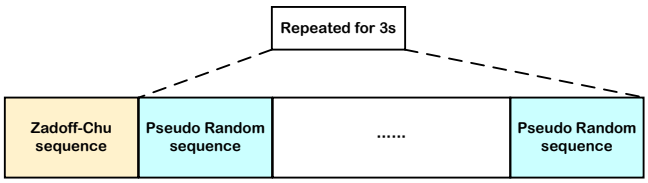



Fig. 7: The structure of the OFDM signal.

LTE system. It is phase-modulated and is orthogonal to its delayed version [41]. Therefore, the ZC sequence assisted synchronization method can achieve much better performance than its well-known counterpart, chirp-based approach. Fig. 5 depicts the correlation results using ZC signal. Apparently, the maximum peak has no comparable "sidelobes" [35], making it easier to perform synchronization.

We design a light-weight strategy to detect the OFDM signals at the receiver end. Our detection scheme consists of two stages. The first stage coarsely locates point-of-interest samples via Root Mean Square (RMS) associated operations. Then the results are refined by ZC synchronization at the second stage. We now elaborate on more details. At the first stage, once the audio samples are captured, we first calculate their RMS [53] after a high pass filter. The RMS of a set of measurements can be defined by the following equation:

$$rms(X) = \sqrt{\frac{\sum_{i=1}^{N} x_i}{N}}, \tag{9}$$

where $X$ denotes the measurements and $N$ represents its length. Then, we can write the differential RMS as:

$$\Delta rms = rms(j) - rms(j-1), j \in [1, N] \tag{10}$$

The original captured samples, RMS, and differential RMS have been depicted in Fig. 6. Evidently, the start and end of the calibration segment can be coarsely derived from the peak and valley of the differential RMS results. These results are to be refined for a more precise extraction. At the second stage, we use time-domain ZC signal to correlate the received samples and accordingly slide the coarse start to a more precise location based on the correlation results, refining calibration segment extraction.

After the successful extraction of the OFDM signals, we slice the signals into different segments. Each segment contains $N$ samples, $\{X_1, X_2, ..., X_N\}_{t_j}$, where $N$ denotes the number of subcarriers, and $t_j$ is the time stamp with index $j$. We equalize each segment with the first one $\{X_1, X_2, ..., X_N\}_{t_1}$ then extract the phase $\{\theta_1, \theta_2, ..., \theta_N\}_{t_j}$. We interpolate the phase across each subcarrier to obtain the slope $k_{t_j}$. The drifting rate can be finally given by:

$$r = \frac{k_{t_j} - k_{t_i}}{t_j - t_i} \tag{11}$$

One can also change the time difference $t_j - t_i$ to scale the estimation granularity. Note that the above estimation method assumes the frequency drift rate is constant. This assumption is valid within several minutes [26]. Considering that the measurement process normally lasts for only several seconds, our compensation method is accurate enough to equalize the errors. It is also worth noting that the transceiver pairs should be immobile during calibration.

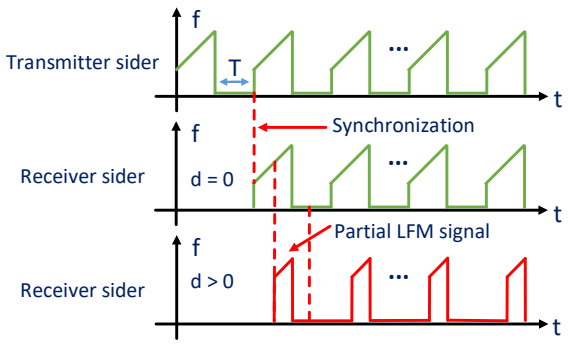



Fig. 8: Signal processing at transceiver side.

### C. Distance Estimation Phase

In the distance estimation phase, the auxiliary device triggers LFM signals with silent periods in between. The measurement device estimates the range information via frequency estimation. In this phase, there are two intermediate processing modules, namely the adaptive sampling strategy and the multipath suppression algorithm, to improve the performance. We now present more details.

*1) Adaptive Sampling for Distance Estimation:* Adaptive sampling is used to cope with the degraded SNR. To understand why it helps, we first need to dive into the details of the operations at the receiver side. The receiver, i.e., the measurement device, periodically captures a fixed number of samples and mixes them with the known chirp signal, performing FFT and estimating the range information. The signals for sampling depicted in Fig. 8 are LFM signals and "silent period". They both last for $T = 0.4$ s and alternate on the time axis. Initially, the measurement device uses the matched filter to detect the presence of the LFM signal, knowing the start of distance estimation phase as well as adjusting the sampling time stamp. Later on, the measurement device retrieves 1920 samples every $2T = 0.08$ s and mix them with the known chirp signal for distance estimation. If the user stays still, the measurement device can capture the entire LFM signal in each cycle and the major frequency component of the mixed results, having the highest peak on the spectrum, would be direct current. When the distance becomes larger (e.g., $d > 0$ depicted in Fig. 8), the extracted signal for mixing is clipped. Consequently, the highest peak shifts proportionally, but the magnitude of which begins to decrease reversely. Our simulation results depicted in Fig. 9 show this phenomenon, which we refer to degraded SNR. If considering the path loss effect, the frequency peaks may even be submerged under noise level when the distance is large. At some extreme case where $d = cT = 340 \times 0.04 = 13.6$ m, ranging cannot work since the captured samples contain no LFM signals. The degraded SNR phenomenon inevitably affects the ranging performance.

One may wonder why we do not use the conventional FMCW signal to handle the degraded SNR problem. Under such circumstance, the measurement device samples a cyclic shift of the LFM signals in each time slot and thus the problem is solved. However, such an adoption would lead to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2018.2879371, IEEE Internet of Things Journal



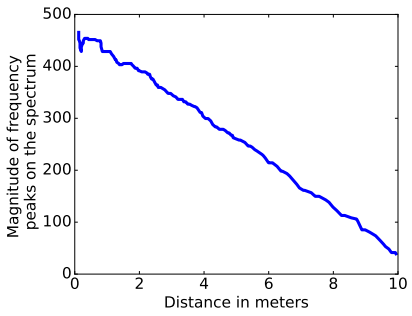



Fig. 9: The magnitude of the frequency peaks on the spectrum drop when the distance becomes large

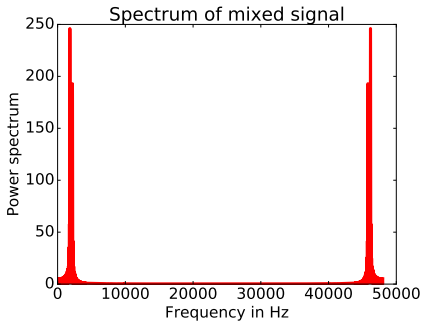



Fig. 10: Using FMCW signal introduce also identical peaks on the spectrum at certain distance

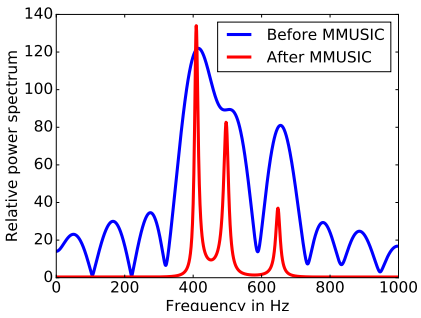



Fig. 11: Comparison of pseudo-spectrum before and after MMUSIC

another problem. When the captured samples are constructed by a cyclic shift around $N/2$, where $N$ is the length of the LFM signals, there are two comparable peaks on the spectrum at certain distances as depicted in Fig. 10. This property obfuscates our decision on estimating the shifted frequency, resulting in ranging errors. In addition, as we have already mentioned in Section IV-A that traditional FMCW method cannot cope with the "ring" effect which causes ISI and have much more stronger multipath interferences.

The solution to degraded SNR is our adaptive sampling strategy. Instead of retrieving 1920 samples every $2T$ interval, we dynamically change this interval according to the current distance. Therefore, when the distance is larger than a specific threshold, we shorten the interval than $2T$ for one sampling period and make it normal ($2T$) in the following cycles. Such a minor time stamp adjustment can guarantee that the captured samples contain LFM signals as many as possible, combating the degraded SNR at varying distances. The adaptive sampling strategy can be described using the following equation:

$$t_{interval} = \begin{cases} 2T, f_x \leq f_0 \\ 2T - 2\frac{(f_x - f_0)T}{B}, f_x > f_0 + \delta \end{cases}, \tag{12}$$

where $f_0 = \frac{d_{thr}B}{cT}$, $d_{thr}$ is the threshold distance where the SNR is deemed as acceptable and $\delta$ is a value that aims to avoid frequent interval changes. During our experiment, we found that our system achieves acceptable performance when $d_{thr} = 2$ and $\delta = 100$. We should keep a track of the subtracted frequency components and add them to the final result.

*2) Multipath Effect:* Except for LOS path, the LFM signal emitted from the transmitter may propagate from different reflected paths arriving at receiver ends. This phenomenon, which is called multipath effect, exists in many wireless [51] or acoustic sensing systems [49]. The multiple reflected signals introduces multiple frequency components in the mixed results, making it hard for the FFT-based method to differentiate them, leading to distance estimation error. To this end, we propose to leverage MMUSIC [17] to resolve the multipath problem, enhancing ranging accuracy.

Based on Eq. 4, the captures signals after mixing have multiple frequency components which can be denoted by [27], [39]:

$$s(f_i) = \sum_{i=1}^{D} \cos\left(2\pi f_i \frac{k}{f_s} + \theta_i\right), \tag{13}$$

where $D$ is the number of paths and $\theta_i$ is the phase which can be ignored. $k$ is the index for the samples, $f_s$ is the sampling rate, and $\frac{k}{f_s}$ is a discrete form for $t$ in Eq. 4. $f_i$ denotes the frequency component corresponding to each multiple reflected signal, and it is proportional to its propagating path. The above expression follows the basic MUSIC model. To apply MUSIC, we need to obtain the auto-correlation matrix $R$ from the mixed results $s$ as $S^H S$, where $S$ is the Toeplitz matrix formed by $s$ with the size $(N - M + 1) \cdot M$ and $N$ is the length of the samples. Since $s$ is the real signal, $R$ is a Hermitian matrix, the size of which is $M$. Following that, the classic MUSIC algorithm applies eigenvalue decomposition to $R$. However, as we stated in the previous section, the performance of the classic MUSIC algorithm can be affected by the signal coherence. In this case, more sophisticated method are needed.

To this end, we propose to use a Modified version of MUSIC [17], to de-correlate the signal coherence and thus enhance the accuracy. The MMUSIC de-correlate the signal coherence by introducing a conjugate reconstruction. That is:

$$\hat{R} = R + JR'J, \tag{14}$$

where $J$ is an $M$-th order anti-matrix and $R'$ is the conjugation of $R$. After that, we then apply eigenvalue decomposition on $\hat{R}$, and pick up the smallest $M_n$ eigenvectors corresponding to the smallest eigenvalues, forming a noise subspace $R_n$, which is thought to be orthogonal to the signal subspace, $T(f_i)$:

$$R_n \cdot T(f_i) = 0, \tag{15}$$

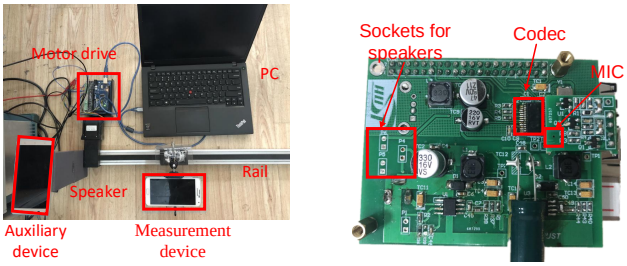
where $T(f_i)$ is the steering vector defined as:

$$T = [1, e^{j2\pi f/f_s}, ..., e^{j2\pi f(M-1)/f_s}]^T \tag{16}$$

Hence we can obtain each $f_i$ by locating the peaks in the pseudo-spectrum as:

$$P = \frac{1}{T^H R_n {R_n}^H T} \tag{17}$$

The legitimate LOS signal, which reflects ground truth range information, has a corresponding peak located on the leftmost




(a) Testbed setup

(b) Customized acoustic board with raspberry Pi

Fig. 12: Testbed setup

side of the spectrum $P$. The resolution of the above algorithm is reported to be 1.2 Hz [27], which is high enough to resolve the multipath problem.

We have evaluated the performance of MMUSIC. We simulate three coherent echoes in the mixed signals and apply MMUSIC to check the results. Fig. 11 depicts the pseudo-spectrum before and after MMUSIC. Apparently, three peaks are identified and sharpened by MMUSIC decomposition. The comparison results clearly show the effectiveness of MMUSIC algorithm.

## V. Performance Evaluation

We implement a proof-of-concept system prototype on Android platforms. We write an android program to record and stream audio signals to a remote PC via UDP. The audio signals are processed by a MATLAB script in real-time. To achieve high computation efficiency, we program the UDP communication block and other signal processing modules using the C language. Finally, we realize a real-time system on Android platforms and Raspberry Pi to evaluate its performance.

The hardware we used in our system includes Samsung Galaxy S5, HUAWEI MATE7, a custom-built acoustic board with Raspberry Pi 3B, and a PC. The Samsung Galaxy S5 equips with powerful quad-core CPUs and 2G RAM, running Android 6.0.1 and the HUAWEI MATE7 has comparable hardware performance. The acoustic board is tailored for our ranging purpose and is controlled by the Pi. It has a INMP411 microphone and two TDA2822-based audio power amplifiers. The codec on the acoustic board is WM8731. The PC has Quad-core i3 and 4G RAM, running 64bit windows 10. Our equipment also includes an $AD8002$-based classic-AB audio power amplifier [2]. Since it is hard to obtain the ground truth, we have customized a platform for performance evaluation as depicted in Fig. 12a. It consists of a rail with a length of 1.6 meters, a stepper motor controlled by Arduino MEGA 2560 [5], and a Thinkpad T440s laptop. The measurement device can be fastened up on the rail and move back and forth. We write a MATLAB GUI program to control the rail at a precision of 0.5mm. This platform gives us the ultra-precise ground truth for performance evaluation. The LFM signal in our system prototype spans the frequency range from 18 kHz to 22 kHz, occupying 4 kHz bandwidth. The duration and the "silent period" both last for 0.04s. At each distance estimation cycle, it retrieves 1920 samples. The interval between each cycle is $2T = 0.08$s.

It is worth noting that directly transmitting the LFM signals introduces audible noise albeit they are in the inaudible band. Such a phenomenon, also known as frequency leakage [45], is mainly caused by the speaker diaphragm inertia [1] which has been demonstrated in [38], [45]. The frequency leakage appears when the signals have abrupt amplitude or phase changes, which can also serve as guidelines to address the problem. Several waveform reshaping techniques have been put forward in literary research. Adding a raised cosine window to reshape the waveform, fitting fade-in and fade-out signals to force phase consistency [19], or just increasing and decreasing the amplitude of the first and last few samples [56], are common waveform reshaping strategies to mitigate the audible noises. Channel estimation [38], a common strategy in communication systems to equalize the channel response, is also a feasible approach to ease the problem. Being effective, these methods still have their shortcomings. For instance, the channel estimation method needs expensive advanced equipments. Moreover, adding a triangle window on the samples [56] leads to non-flat frequency response. During our experiments, we observe that this waveform reshaping method can result in the instability of the frequency estimation results. After extensive trials, we found that linearly increasing and decreasing the amplitude of the first and last as few as 10 samples are enough to mitigate the audible noise. In our system, we choose to manipulate the first and last 100 samples.

Non-flat frequency response [20], also known as frequency selectivity, can also affect the ranging performance. It has been demonstrated in [27] that this non-flat frequency response can affect the sharpness of the peaks in MUSIC spectrum. Our experiments reveal that it can also lead to large estimation variances. In our system, the non-flat frequency response problem is addressed by convolving a compensation filter which has a reciprocal frequency response of the received signals. After the above preparations, we then evaluate our system prototype in various environment settings. We have also explored the impact of different parameter settings on the ranging performance.

### A. SFO estimation performance

In this section, we present our OFDM-based SFO estimation results. We split the available 24 kHz bandwidth into 240 subcarriers and deploy our calibration signal at the $181-200$th ones, occupying the bandwidth from 18 kHz to 20 kHz. The data on the unused subcarriers is set to zero. To compare the performance with the method in [26], we inject an FMCW signal at the lower frequency band. This FMCW signal spans the bandwidth from 12 kHz to 16 kHz with a duration of 0.04 s. Following the above setting, we use a MATLAB script to generate a high-quality audio file in wav format. The auxiliary device triggers the calibration signal by playing the audio file.

At the receiver end, we use a bandpass filter to separate the FMCW and OFDM signal and feed these samples into separate estimation strategies. In the OFDM-based method, we first divide the received audio samples into different segments,

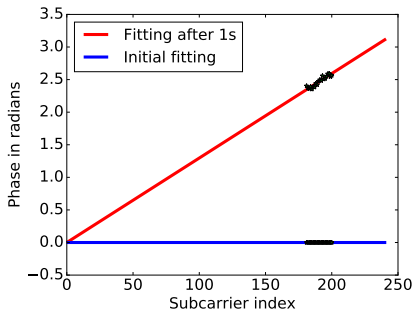



Fig. 13: The OFDM-based SFO estimation method.

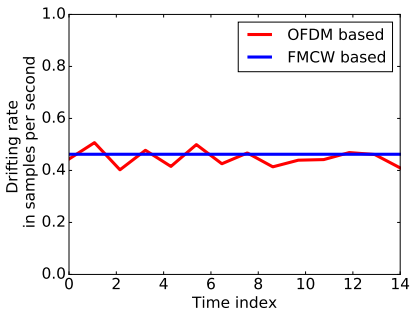



Fig. 14: The SFO estimation results between OFDM-based and FMCW-based.

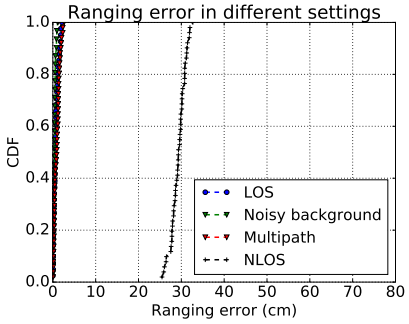



Fig. 15: The ranging error in different scenarios.

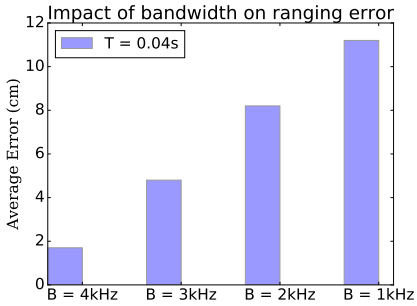



Fig. 16: The impact of bandwidth on ranging performance.

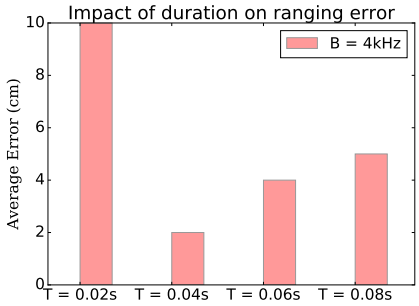



Fig. 17: The impact of duration on ranging performance.

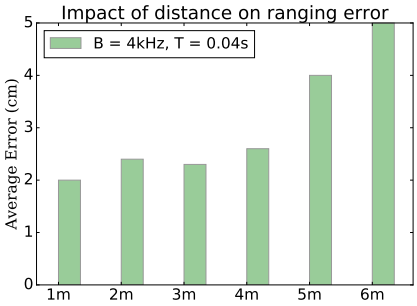



Fig. 18: The impact of distance on ranging performance.

each of which contains $480$ samples. Then we perform FFT on each segment and extract the $181-200$th subcarriers for use. The subcarriers in each segment are normalized by the first one. We interpolate the phase of the $181-200$th subcarriers and extract the slope to estimate the drifting rate as depicted in Fig. 13. In the FMCW-based method, we track the shifted frequency peaks of the mixed signal for $20$ s and interpolate the peaks with respect to time to obtain the drifting rate. Fig. 14 shows the comparison results between our proposed OFDM-based method and the FMCW-based one. The estimation results between two methods are comparable. However, our estimation results have finer granularity. We can update the drifting rate every second while FMCW-based method consumes orders of much more time. Moreover, we can estimate sub-sample drift as depicted in Figure. 13, which, however, is infeasible for FMCW-based method.

### *B. Ranging performance*

We first test the ranging performance under different environmental settings including multipath rich, LOS, None-Line-of-Sight (NLOS), and noise background scenarios. To create a multipath rich environment, the auxiliary device and measurement device are placed in a 2 meter-wide corridor. We connect the audio power amplifier at the output of the auxiliary device to level up the signal strength, enhancing the multipath effect. We then block the LOS path, imitating an NLOS case. Afterwards, we place a mobile phone parallel to the LOS path which plays music at its maximum volume to emulate a noisy background. The sampling rate on the auxiliary device and the measurement device are both $48$ kHz. The measurement device is fastened on our custom-built rail.

Fig. 15 depicts the ranging error under different environmental settings. The average distance estimation error is within $1.7$ cm in LOS, noisy background, and multipath rich environment. The performance becomes worse in NLOS case. Evidently, our acoustic ruler is quite robust to the multipath effect, as the ranging performance is comparable in a multipath rich environment with that in LOS cases.

We then explore the impact of the bandwidth and duration on the ranging performance. Fig. 16 shows that the ranging error decreases with the increase of bandwidth. Such an result is expected as the estimated distance is given by $d = \frac{fcT}{B}$ and when $B$ increases, the resolution becomes higher, leading to a much smaller ranging error. Likewise, when $T$ increases, the resolution decreases thus the ranging error becomes worse as depicted in Fig. 17. However, when $T = 0.02$s, the error becomes large on the contrary, which is up to $10$cm. The large error at $T = 0.02$ s is mainly caused by the inadequate samples. When $T = 0.02$s, there are only $T \cdot f_s = 960$ valid samples. The frequency estimation strategy cannot achieve reasonable accuracy with these limited measurements [49]. Therefore, $T$ has an optimal value and we find that $T = 0.04$ s gives reasonable performance.

We then evaluate the system performance under different distances and orientations. Specifically, we test the ranging performance at distances up to $6$ m and at various incident angles including $-45°$, $0°$, and $45°$ as depicted in Fig. 20. Fig. 18 shows that the ranging errors only slightly increase when the

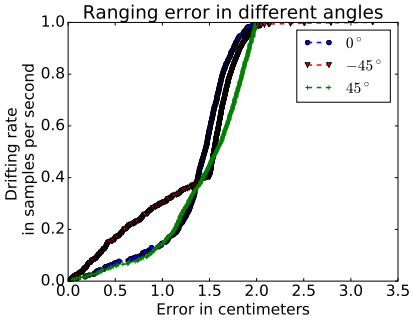



Fig. 19: Ranging error under different angles.

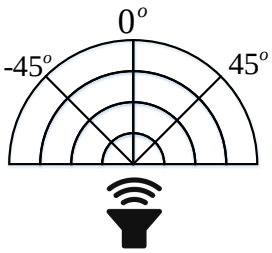



Fig. 20: The angle relative to the speaker.

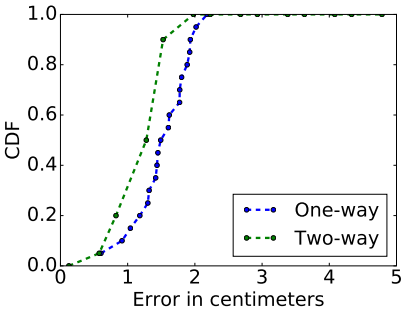



Fig. 21: Comparison between one-way and two-way ranging

distance becomes large. The maximum ranging error is 5 cm at $d = 6$ m. Such an error can be mitigated by reinforcing the signal strength at the transmitter side. It should be noted that we use mobile device for ranging, the output capability of which is not powerful, limiting the maximum operational range. One can employ more powerful smart devices such as smart TV, Google Home, or Amazon Echo as the auxiliary device or alternatively employ lower band signals to boost the range. Last but not least, our proposed acoustic ruler is also robust to different incident angles as Fig. 19 evidently depicts that the ranging errors across different angles are comparable. The resistance to different incident angles enables precise multiple dimension measurement simultaneously.

We finally compare our one-way ranging method with the two-way ranging approach [35]. The two approaches utilize the same chirp signals with $f_{min} = 18$ kHz, $B = 4$ kHz, and $T = 0.04$ s. The experiment is accomplished using smartphones. The results are depicted in Fig. 21. We can observe that the two methods achieve comparable performance. However, the two-way approach is subject to large errors. In addition, the two-way ranging approach is based on cross-correlation which is subject to $2-3$ samples errors [34] under low SNR. This error is equivalent to $\frac{2}{48000} \times 340 \approx 1.4$ cm. Furthermore, in practical usage, two-way ranging approach suffers from microphone-speaker distance ambiguity since it is really non-trivial to measure the ground truth of the microphone-speaker distance. Such a distance ambiguity, if not carefully calibrated, can easily result in decimeter-level errors. In comparison, our one-way approach does not have these shortcomings.

### C. Computation cost

We have tried to deploy our system on different platforms including a PC, a Samsung S5 smartphone, and a Raspberry Pi board. The corresponding processing time for different platforms are depicted in Table I. On the PC, we compile the sluggish MATLAB code into more efficient c version for acceleration. The total processing time decreases from 91 ms down to 15 ms. On the Android platform, we utilize the Android NDK APIs to speed up code execution. The total processing time is about 66 ms. On the Raspberry Pi, we do not implement MMUSIC algorithm since it causes too much latency. Therefore, we use naive FFT analysis. The naive FFT analysis cannot handle the multipath problem but usually, a measurement takes place under good LOS conditions. Therefore, such an implementation is still suitable for most cases. We utilize the GPU [7] and neon [6] structure for acceleration, making the total processing time be within 30 ms. Nevertheless, in our implementation, the fastest update rate is $2T = 80$ ms, which is the duration of a chirp transmission. Therefore, our algorithm can run in real-time across diverse platforms. It should be noted that for resource-constrained devices, they can serve as the auxiliary devices, which only need to play already-generated audio files. Such a lightweight operation can be accomplished by most IoT devices.

| Platforms | PC | Android smartphone | Raspberry Pi |
|---|---|---|---|
| Time | 15 ms | 66 ms | 30 ms |

TABLE I: Time costs on different platforms

## VI. Conclusions

This paper presents the design and implementation of a purely software-based acoustic ruler on COTS mobile devices. It exhibits salient features compared with literary work. Our approach does not need any tight synchronization, involvement of the central coordinator, cumbersome kernel modifications, or extra hardware. This acoustic ruler adopts one-way sensing mechanism which enables multi-dimensional measurements simultaneously. In realizing such a system, we have put forward several proposals to address the design challenges. We propose an OFDM-based method to cope with the sampling frequency offset, which is able to precisely perform sub-sample drift estimation. We adopt an adaptive sampling strategy and MMUSIC algorithm to enhance the robustness of our acoustic ruler. The system prototype demonstrates a median ranging error of 1.7cm.

## References

[1] Loud speaker diaphragm. In *US Patent 2,905,260*, Sep 1959.
[2] Classic-AB audio power amplifier. http://www.shenzhensum.com/products/datasheet/8002(2.0W).pdf, 2017.
[3] iPin. http://www.ipinlaser.com/, 2017.
[4] The growth of mobile phone users. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/, 2017.

[5] Arduino MEGA 2560. https://store.arduino.cc/arduino-mega-2560-rev3, 2018.

[6] ARM neon. https://developer.arm.com/technologies/neon, 2018.

[7] GPU FFT. http://www.aholme.co.uk/GPU_FFT/Main.htm, 2018.

[8] I. Bisio, A. Delfino, A. Grattarola, F. Lavagetto, and A. Sciarrone. Ultrasounds-based context sensing method and applications over the internet of things. *IEEE Internet of Things Journal*, 2018.

[9] H. Bolcskei. Principles of mimo-ofdm wireless systems, 2004.

[10] M. Chen and Y. Hao. Task offloading for mobile edge computing in software defined ultra-dense network. *IEEE Journal on Selected Areas in Communications*, 36(3):587–597, 2018.

[11] M. Chen, Y. Miao, Y. Hao, and K. Hwang. Narrow band internet of things. *IEEE Access*, 5:20557–20577, 2017.

[12] M. Chen, J. Yang, X. Zhu, X. Wang, M. Liu, and J. Song. Smart home 2.0: Innovative smart home system powered by botanical iot and emotion detection. *Mobile Networks and Applications*, 22(6):1159–1169, 2017.

[13] S. Counts and E. Fellheimer. Supporting social presence through lightweight photo sharing on and off the desktop. In *In Proceedings of the ACM SIGCHI*, 2004.

[14] D. Frohlich, A. Kuchinsky, C. Pering, A. Don, and S. Ariss. Requirements for photoware. In *In Proceedings of the ACM Computer Supported Cooperative Work*, 2002.

[15] J. Heiskala and J. Terry, Ph.D. *OFDM Wireless LANs: A Theoretical and Practical Guide*. 2001.

[16] K. Kinoshita and T. Nakatani. Microphone-location dependent mask estimation for bss using spatially distributed asynchronous microphones. In *2013 International Symposium on Intelligent Signal Processing and Communication Systems*, pages 326–331, Nov 2013.

[17] D. Kundu. Modified music algorithm for estimation doa of signals. *Signal Process.*, 48(1):85–90, Jan. 1996.

[18] P. Lazik, N. Rajagopal, O. Shih, B. Sinopoli, and A. Rowe. Alps: A bluetooth and ultrasound platform for mapping and localization. In *In Proceedings of ACM SenSys*, 2015.

[19] P. Lazik and A. Rowe. Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In *In Proceedings of ACM SenSys*, 2012.

[20] H. Lee, T. H. Kim, J. W. Choi, and S. Choi. Chirp signal-based aerial acoustic communication for smart devices. In *In Proceedings of IEEE INFOCOM*, 2015.

[21] H. Li, K. Ota, and M. Dong. Learning iot in edge: Deep learning for the internet of things with edge computing. *IEEE Network*, 32(1):96–101, Jan 2018.

[22] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. *IEEE Internet of Things Journal*, 4(5):1125–1142, Oct 2017.

[23] C. Liu, L. Zhang, Z. Liu, K. Liu, X. Li, and Y. Liu. Lasagna: Towards deep hierarchical understanding and searching over mobile sensing data. In *In Proceedings of ACM MobiCom*, 2016.

[24] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye. Push the limit of wifi based localization for smartphones. In *In Proceedings of ACM MobiCom*, 2012.

[25] K. Liu, X. Liu, and X. Li. Guoguo: Enabling fine-grained indoor localization via smartphone. In *In Proceedings of ACM MobiSys*, 2013.

[26] W. Mao, J. He, H. Zheng, Z. Zhang, and L. Qiu. High-precision acoustic motion tracking: Demo. In *In Proceedings of ACM MobiCom*, 2016.

[27] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun. Indoor follow me drone. In *In Proceedings of ACM MobiSys*, 2017.

[28] H. Meyr, M. Moeneclaey, and S. Fechtel. *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. 1997.

[29] S. Miyabe, N. Ono, and S. Makino. Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in stft domain. In *In Proceedings of IEEE ICASSP*, 2013.

[30] S. Miyabe, N. Ono, and S. Makino. Optimizing frame analysis with non-integrer shift for sampling mismatch compensation of long recording. In *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 1–4, Oct 2013.

[31] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. S. Oh. Semisupervised deep reinforcement learning in support of iot and smart city services. *IEEE Internet of Things Journal*, 5(2):624–635, April 2018.

[32] T. Nadeem and L. Ji. Location-aware ieee 802.11 for spatial reuse enhancement. *IEEE Transactions on Mobile Computing*, 6(10):1171–1184, Oct 2007.

[33] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan. Dhwani: Secure peer-to-peer acoustic nfc. In *In Proceedings of ACM SIGCOMM*, 2013.

[34] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *In Proceedings of ACM CHI*, 2016.

[35] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *In Proceedings of the ACM SenSys*, 2007.

[36] G. Premsankar, M. D. Francesco, and T. Taleb. Edge computing for the internet of things: A case study. *IEEE Internet of Things Journal*, 5(2):1275–1284, April 2018.

[37] P. Robertson and S. Kaiser. Analysis of the effects of phase-noise in orthogonal frequency division multiplex (ofdm) systems. In *In Proceedings of the IEEE ICC*, 1995.

[38] N. Roy, H. Hassanieh, and R. Roy Choudhury. Backdoor: Making microphones hear inaudible sounds. In *In Proceedings of ACM MobiSys*, 2017.

[39] R. O. Schmidt. A signal subspace approach to multiple emitter location and spectral estimation. *Thesis Ph.D. Stanford University 1982*, 1982.

[40] M. Speth, S. Fechtel, G. Fock, and H. Meyr. Optimum receiver design for ofdm-based broadband transmission .ii. a case study. *IEEE Transactions on Communications*, 49(4):571–578, Apr 2001.

[41] S. Stefania, T. Issam, and B. Matthew. *LTE - The UMTS Long Term Evolution: From Theory to Practice, 2nd Edition*. 2011.

[42] F. Student. Acoustic Ruler Pro. https://itunes.apple.com/us/app/acoustic-ruler-pro/id475081963?mt=8, 2017.

[43] S. Sur, T. Wei, and X. Zhang. Autodirective audio capturing through a synchronized smartphone array. In *In Proceedings of ACM MobiSys*, 2014.

[44] J. Tang, D. Sun, S. Liu, and J. L. Gaudiot. Enabling deep learning on iot devices. *Computer*, 50(10):92–96, 2017.

[45] Y.-C. Tung and K. G. Shin. Expansion of human-phone interface by sensing structure-borne sound propagation. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '16, 2016.

[46] M. Uddin and T. Nadeem. Rf-beep: A light ranging scheme for smart devices. In *In Proceedings of IEEE PerCom*, 2013.

[47] C. Wang, H. Lin, and H. Jiang. Cans: Towards congestion-adaptive and small stretch emergency navigation with wireless sensor networks. *IEEE Transactions on Mobile Computing*, 15(5):1077–1089, 2016.

[48] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu. Understanding and modeling of wifi signal based human activity recognition. In *In Proceedings of ACM MobiCom*, 2015.

[49] W. Wang, A. X. Liu, and K. Sun. Device-free gesture tracking using acoustic signals. In *In Proceedings of ACM MobiCom*, 2016.

[50] Y. Wang, J. Li, R. Zheng, and D. Zhao. ARABIS: an asynchronous acoustic indoor positioning system for mobile devices. *CoRR*, abs/1705.07511, 2017.

[51] J. Xiong and K. Jamieson. Arraytrack: A fine-grained indoor location system. In *In Proceedings of USENIX NSDI*, 2013.

[52] Y. Yang, J. Nie, and J. Luo. Reflexcode: Coding with superposed reflection light for led-camera communication. In *In Proceedings of ACM MobiCom*, 2017.

[53] E. Yuce, S. Minaei, and S. Tokat. Root-mean-square measurement of distinct voltage signals. *IEEE Transactions on Instrumentation and Measurement*, 56(6):2782–2787, Dec 2007.

[54] S. Yun, Y.-C. Chen, and L. Qiu. Turning a mobile device into a mouse in the air. In *In Proceedings of ACM Mobysis*, 2015.

[55] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao. Strata: Fine-grained acoustic-based device-free tracking. In *In Proceedings of ACM MobiSys*, 2017.

[56] B. Zhou, M. Elbadry, R. Gao, and F. Ye. Batmapper: Acoustic sensing based indoor floor plan construction using smartphones. In *In Proceedings of ACM MobiSys*, 2017.