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.2019.2903211, IEEE Internet of Things Journal

RAMTEL: Robust Acoustic Motion Tracking using Extreme Learning Machine for Smart Cities

Yang Liu, Wuxiong Zhang, Yang Yang, *Fellow, IEEE,* Weidong Fang, Fei Qin, and Xuewu Dai

***Abstract*—Motion tracking is attractive in what concerns a smart city environment, where citizens have to interact with Internet of Things infrastructures spread all around one particular city. Motion tracking is important for smart services and Location Based Services in smart cities, since it provides natural ways for users to interact with the IoT infrastructures, such as the ability to recognize of a wide range of hand motion in real-time. Compared with dedicated hardware devices, ubiquitous devices with reliable speakers and microphones can be developed to achieve cheap acoustic-based motion tracking, which is appropriate for low-power and low-cost IoT applications. However, for complex urban environments, it is very difficult for acoustic-based methods to achieve accurate motion tracking due to multipath fading and limited sampling rate at mobile devices. In this paper, a new parameter called Multipath Dispersion Vector is proposed to estimate and mitigate the impact of multipath fading on received signals using Extreme Learning Machine. Based on MDV, a Robust Acoustic Motion Tracking (RAMTEL) method is proposed to calculate the moving distance based on the phase change of acoustic signals, and track the corresponding motion in two-dimensional plane by using multiple speakers. The method is then proposed and implemented on standard Android smartphones. Experiment results show, without any specialized hardware, RAMTEL can achieve an impressive millimeter-level accuracy for localization and motion tracking applications in multipath fading environments. Specifically, the measurement errors are less than 2 mm and 4 mm in one-dimensional and two-dimensional scenarios, respectively.**

***Index Terms*—Motion Tracking; Multipath Fading; Internet of Things; Smart City**

Earlier version of this work was accepted by 2019 IEEE INFOCOM, Paris, France, Apr. 2019 [1]. This work is partially supported by the National Natural Science Foundation of China (No. 61571004, No. 61773111), the Shanghai Natural Science Foundation (No. 16ZR1435200), the Science and Technology Innovation Program of Shanghai (No. 17DZ1200302, No. 17DZ2292000, No. 16510711600), and the Scientific Instrument Developing Project of the Chinese Academy of Sciences with Grant (No. YJKYYQ20170074). *(Corresponding author: Wuxiong Zhang.)*

Yang Liu is with Shanghai Institute of Microsystem and Information Technology, and University of Chinese Academy of Sciences, Shanghai, China (e-mail: yang.liu@mail.sim.ac.cn).

Wuxiong Zhang and Weidong Fang are with Shanghai Institute of Microsystem and Information Technology, Shanghai, China (e-mail: wuziong.zhang@mail.sim.ac.cn; weidong.fang@mail.sim.ac.cn).

Yang Yang is with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China (yangyang@shanghaitech.edu.cn).

Qin Fei is with the School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China (email: fqin1982@ucas.ac.cn).

Xuewu Dai is with the State Key Laboratory of Synthetical Automation for Process Industry, Northeastern University, Shenyang, China (email: daixuewu@mail.neu.edu.cn).

I. Introduction

A. Motivation

MOTION tracking is attractive in what concerns a smart city environment, where citizens have to interact with Internet of Things (IoT) infrastructures spread all around one particular city. Motion tracking is important for smart services [2] and Location Based Services (LBS) [3] in smart cities, since it provides natural ways for users to interact with the IoT infrastructures, such as the ability to recognize of a wide range of hand motion in real-time [4]. For the interaction between user and IoT infrastructures, different kinds of motions caused by user's gesture or movement are required to tracked accurately in real-time, since slight different motions may represent different meanings at different times. Further, the motions in smart cities are usually slow or small, such as waving hand and walking, so the motion tracking technologies are required to track the small and slow movements in daily life. In addition, unlike rural areas, the urban environments of smart cities are usually complex, since there are many buildings, cars, and citizens, which bring interference on accurate motion tracking. Many efforts based on wireless signals have lately been made for standard mobile devices, such as smartphone, smartwatch, Tablet PC, which could gather information about the device movements according to the change of received wireless signals. Some works have been proposed based on different electromagnetic signals, such as Wi-Fi signal [5], visible light [6], and millimeter wave [7]. These methods usually sample the signals with high frequency to capture the subtle changes in the signals. The methods have been considered as a promising approach to locate smartphones, since they don't require additional hardware, beside wireless network card or camera on the smartphone. However, the computation overheads of them are huge, and it's challenging to track user's motion in real-time using commercial hardware devices. Further, they only achieve sub-meter-level accuracy, which is not good enough for tracking a user's gesture or posture. Compared with electromagnetic signals, acoustic signals have much lower frequencies and slower propagation speeds. Therefore, they are very suitable to be used for high-accurate and low-latency motion tracking applications, such as the interaction between citizens and IoT infrastructures in a smart city. Further, unlike dedicated hardware devices, the ubiquitous devices with reliable speakers and microphones can be developed to achieve cheap acoustic-based motion tracking, which is appropriate for low-power and low-cost IoT applications.

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.2019.2903211, IEEE Internet of Things Journal

B. Technical Challenges

For the smart services and LBS in smart cities, the first challenge is to achieve accurate motion tracking using acoustic signals. Existing acoustic based motion tracking system either use Doppler effect based methods [8] [9] [10], or Frequency Modulated Continuous Wave (FMCW) based methods [11] [12] [13]. Doppler effect based measurement can only provide the coarse-grained moving speed and direction, and cannot obtain specific position of the target. Traditional FMCW systems emit periodic pulses or chirps which bring in audible harsh noise on commercial speakers [14]. Both Doppler effect based and FMCW based methods assume that the target moves uniformly in a time window, so that they can use the frequency information in the time window to estimate target's moving distance, and thus tracking the target. However, in practical applications, such as tracking a waving hand, the target does not always move uniformly, which would bring error to their estimation. Thus, their works only achieve coarse-grained measurements.

The second challenge is to track small and slow movements of target devices. Specifically, the relative velocity between source and receiver could cause a frequency shift of received acoustic signals, and More obvious frequency shift could be achieved at higher relative velocity. Thus, traditional Doppler effect based systems require the target device to move quickly, such swinging the target vigorously, so that they can obtain obvious frequency changes of received signals caused by Doppler shift. However, the resolution of Doppler shift is limited by the low sampling rate on commercial mobile devices (*e.g.*, 48 KHz for typical mobile phones). The Doppler effect based methods can only coarsely estimate the target's moving speed and direction while the target moves fast, and cannot track slow movements due to limited resolution. In FMCW based methods, a source transmits acoustic signals with high bandwidth, and a receiver analyzes the information in frequency domain using Fast Fourier Transform (FFT) for example. The resolution of moving distance is limited by the bandwidth of the signals (7-8 KHz at most for typical smartphone without introduce notable noises). Thus, it is very difficult for those traditional methods to track small and slow movements of mobile devices.

The third challenge is to mitigate severe fading effects in complex urban environments. Normal mobile devices using traditional methods can hardly distinguish Line-of-Sight (LoS) signals from Non-LoS (NLoS) signals with slightly different delays from multiple propagation paths. Some typical studies about LoS and NLoS focus on the optimization of coverage probability, energy efficiency, transmissions and multislop path loss for cell networks [15] [16], which cannot be use in acoustic motion tracking directly. Due to the reflection of furniture and walls, the received acoustic signals are the superimposition of signals from different direct and reflected paths and each path has different delay and attenuation. So it's challenge to calculate the transmission time from the source to the receiver accurately, and track target's motion. There have been some works on improving the accuracy of acoustic motion tracking [11] [17] [18]. However, they either require specified hardware or have limited performance in multipath fading environments. For example, in SoundTrak [17], a dedicated smart watch tracks the motion of a micro-speaker using the inaudible acoustic signals, but it requires a specified microphone array placed on the smart watch to track the micro-speaker's motion. Vernier [18] proposes an active motion tracking approach by calculating the phase change of received signals and then estimating moving distance. However, the effects of multipath fading are not taken into consideration in Vernier because the target device is moved in a small area without the interference from multiple paths.

C. Proposed Approach

In this paper, a Robust Acoustic Motion Tracking (RAMTEL) method is developed to track the motion of typical mobile devices, *i.e.*, smartphone, smartwatch, Tablet PC, using Extreme Learning Machine (ELM). RAMTEL use the phase changes of received acoustic signals to estimate target device's moving distance relative to each source. Since the phase information can be derived in time domain, we can track the motion in real-time, and the target isn't required to move uniformly. Further, the resolution of moving speed and distance also is not limited by the sampling frequency and bandwidth of the signals as Doppler effect and FMCW based methods. Therefore, RAMTEL can track small and slow movements in real-time. In order to mitigate the effects of multipath fading, we leverage the fact that the multipath fading effects have different impacts on the phase-based measurement of different frequencies at the same time due to their different wavelengths and phases. Thus, RAMTEL could select the signal with the smallest impact of multipath fading, and provides fine-grained motion tracking for the device in typical urban environments, *i.e.* indoor environments.

Specifically, we use multiple sound sources transmit inaudible acoustic signals at different frequencies, the target mobile device receives these transmissions, and derives the distance change to each source based on phases of received signals. In order to achieve phase based motion tracking, we should address several major challenges. First, we need to calibrate the phase offsets between sender and receiver due to their asynchronous system clocks. Due to different system clocks used in receiver and sender, the phase difference between them increases over time. It is difficult to distinguish the phase change caused by the movement of receiver or the asynchronous system clocks. Thus, before phase based ranging, the phase offsets due to asynchronous clocks should be calibrated. The increases of phase offsets are slow and difficult to be measured by frequency analysis methods, *e.g.,* FFT. Second, we should accurately measure the phase of LoS signals in multipath fading environments. The received signal is a superposition of signals from different paths and it's difficult to distinguish LoS signals from NLoS signals on normal mobile devices. Thus, it's even harder to calculate the actual phase of LoS signals.

To address the challenges, we propose a fine-grained phase calibrating method, which could accurately calibrate the phase offsets at different frequencies between the target mobile

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.2019.2903211, IEEE Internet of Things Journal

device and all sources. Further, in order to obtain the accurate phases of LoS signals, which are vulnerable to multipath fading in indoor environments, a new parameter called Multipath Dispersion Vector (MDV) is proposed to estimate the impact of multipath fading on received signals at different frequencies. MDV is obtained from the scatter diagram of received quadrature and in-phase signals, which is extracted as a feature for the training set of ELM. Based on MDV, ELM can select the signal with minimal interference to track the motion, which can effectively mitigate multipath fading effects. Thus, RAMTEL can obtain the accurate phase change from the signals under different impact of the effects, and track corresponding motion even a small and slow one. Compared with our prior work [1], RAMTEL uses machine learning based methods to improve system performance in multipath fading environments, which could improve robustness in complex scenarios. In particular, we leverages a high dimension feature (MDV) of the physical waveform , as discussed in Section IV-B, and use Extreme Learning Machine to estimate the impact of fading effects based on MDV, as discussed in Section IV-C. Further, this work evaluates the performance of tracking small and slow movements by conducting experiments. In the experiments, RAMTEL monitors the breathing of six volunteers, as discussed in Section VII-A3. The results show that our system could track small and slow movements, which indicates that our method have potential to carry out new applications, such health care, alarm systems and intrusion detection. some complex scenarios are also evaluated using experiments, such as LoS and NLoS scenarios, scenarios under different types of noises, and the impact of movement around the direct path between sender and receiver.

The key contributions of this paper are summarized as follows:

- To address the technical challenges of accurate motion tracking using acoustic signals, we propose a fine-grained phase calibrating method to calibrate the phase offset caused by the asynchronous system clocks at sender and receiver, thus the movements can be accurately tracked by our system using the phase change of received signals.
- To address the technical challenges of motion tracking in a typical urban environment, a new parameter called Multipath Dispersion Vector (MDV) is proposed to estimate and mitigate the effects of multipath fading on received signals using Extreme Learning Machine, a single hidden layer feed-forward neural network.
- Based on MDV, a novel Robust Acoustic Motion Tracking (RAMTEL) method is developed to calculate the moving distance based on the phase change of acoustic signals, and track corresponding motion by using multiple speakers.
- A prototype system is implemented on a standard Android smart phone. Experiment results show our RAMTEL method can achieve an impressive millimeter-level accuracy for localization and motion tracking applications in multipath fading environments. Specifically, the measurement errors are less than 2 mm and 4 mm in one-dimensional (1-D) and two-dimensional (2-D) scenarios, respectively.

In summary, our work achieves: (1) feasible motion tracking with mm-level accuracy on mobile devices, (2) strong robustness in dense multipath fading environments.

The rest of this paper is organized as follows. Related work is reviewed in Section II. Section III provides the system design for phase calibration, phase change calculation, and phase based 1-D ranging. Detailed procedures of multipath mitigation are given in Section IV. The details of motion tracking in 2-D plane are given in Section V. Implementation details and experimental results are presented and discussed in Sections VI and VII, respectively. Some limitations and future work are analyzed in Section VIII. Finally, Section IX concludes this paper.

II. Related work

There are many existing works using acoustic signals to track the motion of mobile devices, which are clearly different from our work.

Doppler effect based acoustic motion tracking: Many schemes use Doppler effect to track devices' moving distance or trajectory [8] [9] [10]. They leverage the fact that the relative velocity between sound sender and receiver could cause a frequency shift of received acoustic signals, which can be measured by frequency domain analysis in a time window, such as Short Time Fourier Transform (STFT). Specifically, they assume that the devices move uniformly and quickly in each time windows, so that they can obtain obvious frequency shift of received signals due to Doppler shift, thus estimating the moving distance and direction according to the frequency shift in the time windows. However, the resolution of moving speed v_{res} is limited by the window size, and the assumption that requiring the devices move uniformly is noise-prone in practical systems. We have $v_{res} = c \cdot F_s/(N_{STFT} \cdot F_c)$ [8]. Where c is velocity of sound in air. F_s is the sampling rate. The maximum F_s is 48 KHz for typical smartphones. F_c is the original frequency of the signal. When F_c is 18 KHz and N_{STFT} is 4096, the resolution of moving speed is about 22 cm/s. Thus, Doppler based methods couldn't detect slow movement which is vital in mobile interaction. In contract, our system is able to track slow movements in real-time without limited by the moving window size.

FMCW based acoustic motion tracking: Some works track the motion of mobile devices using Frequency Modulated Continuous Wave (FMCW) signals [11] [12] [13]. They estimate the moving distance of the devices by calculating the frequency difference of the FMCW signals between receiver and sender. However, the resolution of moving distance d_{res} is limited by the bandwidth of transmitted signals, $d_{res} = c/B$ [11]. Where c is sound speed, and B is the bandwidth of transmitted signals. For example, when B is 7 KHz, the resolution is about 4.9 centimeter in indoor environments. Thus, FMCW based schemes are unable to track the motion of small movements. In our work, we calculate the distance change in time domain, instead of using frequency analysis, whose ranging resolution is not restricted by the bandwidth of signals. Our system is able to track the small movements without limited by bandwidth of signals.

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.2019.2903211, IEEE Internet of Things Journal

Phase based acoustic motion tracking: Recently, phase based acoustic motion tracking methods have been proposed [11] [17] [18]. The phase offsets between receiver and source due to asynchronous system clocks are approximately compensated as a fixed value at each frequency in [11] [18]. However, the rough approximations of the phase offsets limit accuracy and cause accumulating measuring error over time. Further, multipath effects aren't taken into consideration in typical active ranging approaches [17] [18], because the target devices in their works are moved in a small area without the interference from multiple paths, such as an area of about $10cm \times 10cm$ in [17]. For larger areas, the multiple effects can't be neglected and hinder the performance of motion tracking. Compared with our prior work [1], RAMTEL improves the robustness in different types of multipath fading environments since we use machine learning technologies with a high dimension feature (MDV) to improve the accuracy of estimating multipath fading effects.

Signature based acoustic motion tracking: Signatures can be obtained acoustic signals [19] [20] [21]. The position signature of mobile devices can be obtained from the characteristics of received signals, such as Received Signal Strength Indication (RSSI) and spectrum. They first collect the signal characteristics at a set of know position, and store this information as signatures in a database along with the known coordinates in an off-line phase. In order to obtain real-time localization, the signature based scheme needs to compare its measurement results with its huge database at all times. During the on-line tracking phase, users match received signals with those recorded characteristics, and choose the closest match as the estimated location. However, these signatures vary over time and environmental mobility, and need to be updated when the environment changes. Compared with signatures-based localization methods, we use a model-based method to track device's position, and reduce the dependence on the database collected in off-line phase.

III. System Design

In this section, the overview of our approach is firstly provided. Then, the system design is described, including phase calibration, multipath effect mitigation, and phase based ranging.

A. RAMTEL Overview

Because of the above limitations of existing approaches, we propose a phase-based ranging approach for acoustic signals using LoS signals, which has more robust and accuracy performance.

In order to make the sound inaudible, a static audio source, like a commercial speaker, continually transmits sinusoidal signals at single frequency in the range of 17-23 KHz. The sound in the range can be generated and received by many commercial devices without introducing audible noises.

The signals received by smartphone's microphone are composed of LoS signals and the NLoS signals reflected by static objects, such as wall and table. When the receiver moves close/away, the length of the LoS path would change and the reflection path remains stationary. When the receiver, like a smartphone, moves close/away, the phase of acquired signal would increase/decrease due to the length change of LoS path, as shown in Figure 1. As the phase of the signal increases by 2π, the path length would decrease by one wavelength of the sound wave. We can use phase changes $\Delta\theta$ to determine the movement direction and calculate the real-time relative moving distance Δl of the receiver.

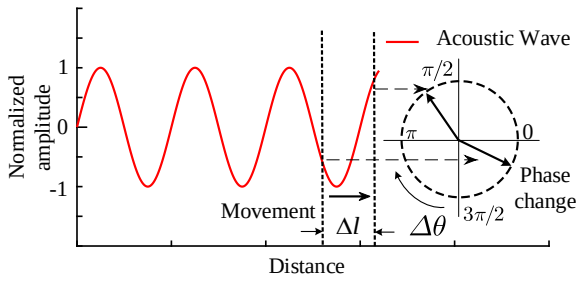


Fig. 1. Calculating moving distance using the phase based of acoustic signal.

We show the design of RAMTEL in Figure 2. The smartphone receives acoustic signals from its microphone. The signals first walk through multiple Band Pass Filters (BPFs). For each BPF, only the signals at the specific frequency could pass through the BPF, and the interference from the signals at other frequencies are both eliminated. After passing through the BPFs specified at different frequencies, the signals at different frequencies could be measured independently. Then, the transformation distance changes can be calculated based on the phase changes of signals. However, the phase change is very vulnerable to the interference of multipath effects in indoor environments, which could lead to incorrect measurement. In particularly, some transformation distance changes could be used to estimate the moving distance of the smartphone, while some of them are interfered by multipath fading effects, and could lead to incorrect estimation of the moving distance. In order reduce the impact of multiple paths, we propose a novel multipath mitigation algorithm based on machine learning technologies. The algorithm could select the signal with the smallest impact of multipath fading to estimate the moving distance in real-time, and thus improving the performance of the phase based measurement in practical multipath fading environments.

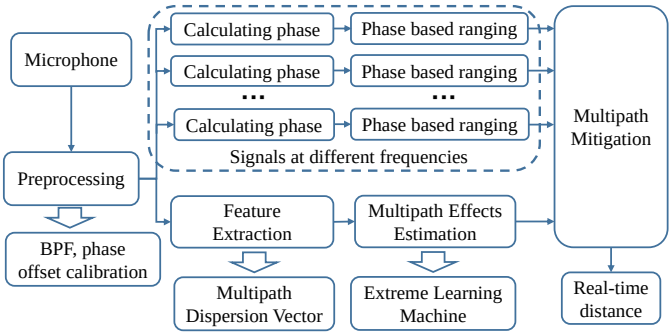


Fig. 2. The flowchart of RAMTEL

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.2019.2903211, IEEE Internet of Things Journal

B. Calculating Phase Change

Before phase based ranging, we need to calculate the phase change in the received signals.

We now give an overview of RAMTEL when a source transmits the signal at a single acoustic frequency, namely $A\cos(2\pi f_c t)$. A is the amplitude of sound and f_c is the frequency of the sound. The mobile device obtains acoustic signal $R_i(t)$ from its microphone, as the MIC block in Figure 3. In order to measure the phase change at each frequency independently, we use a BPF with narrow band which could pass the signals around center frequency f_c, and rejects signals at other frequencies, as shown in the figure. We delay the filtered signal R_α by a quarter of fundamental-wave period. The delayed signal R_β is orthogonal to R_α,

$$\begin{aligned} R_\alpha(t) &= A'\cos(2\pi f_c t - \tau), \\ R_\beta(t) &= A'\sin(2\pi f_c t - \tau). \end{aligned} \tag{1}$$

Where A' is the amplitude of the signal after transmission attenuation and filtering. τ is the phase delay due to sound's propagation from source to receiver at time t. R_α and R_β can be further used to estimate the impact of multipath effects in Section IV. We calculate the phase change using *Park*

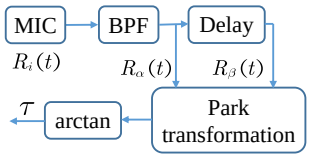


Fig. 3. Calculating phase change using *Park* transformation.

transformation [22]. After multiplying a transformation matrix $\mathbf{P}$ to R_α and R_β,

$$\begin{bmatrix} R_d(t) \\ R_q(t) \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} R_\alpha(t) \\ R_\beta(t) \end{bmatrix} \tag{2}$$

where $\mathbf{P}$ is the *Park* transformation matrix

$$\mathbf{P} = \begin{bmatrix} \cos(2\pi f_c t) & \sin(2\pi f_c t) \\ -\sin(2\pi f_c t) & \cos(2\pi f_c t) \end{bmatrix}.$$

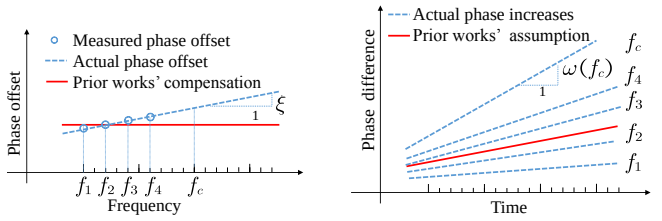
We can obtain two based band signals without the carrier frequency f_c component, corresponding to R_d and R_q:

$$\begin{aligned} R_d(t) &= A'\cos(\tau), \\ R_q(t) &= -A'\sin(\tau). \end{aligned} \tag{3}$$

Then, we calculate the phase delay τ at time t using inverse tangent transformation, and the phase change in each frame.

C. Calibrating Phase Offsets due to Asynchronous System Clocks

Due to different system clocks used in receiver and sender, the phase difference between them increases over time. It is difficult to distinguish the phase change caused by the movement of receiver or the asynchronous system clocks. Thus, before phase based ranging, the phase offsets due to asynchronous clocks should be calibrated. In prior works [11] [17] [18], they assume that the linear phase offsets can be seen as a fixed value at each frequency. The phase offsets at different frequencies are compensated by the phase offset at a certain frequency, as f_2 in Figure 4a. Thus, the phase difference at different frequencies are compensated as the same linear increase in time domain. However, according to our extensive experiments and the comparison with the impact of asynchronous system clocks on Orthogonal Frequency Division Multiplexing (OFDM) system [23], the asynchronous system clocks also lead to a linear phase increase in frequency domain, thus causing different phase increases at different frequencies in time domain, as shown in Figure 4b. The approximate compensations in prior works limit their accuracy and cause measuring error accumulated over time.



(a) The phase offset in frequency domain

(b) The phase increase in time domain

Fig. 4. Calibrating phase in time and frequency domain

In order to compensate the phase offset at each frequency, we first measure the phase offsets at several frequencies from a source, such as f_1,f_2,f_3, and f_4 in Figure 4b. Then, we use a linear fitting to estimate the slope and intercept of the linear phase shift in frequency domain. The signals from each source in our system are generated by a sound card and have the same clock, so that the phase of signals at frequency f_c from an arbitrary source can be compensated as

$$\begin{aligned} \tau^{adjusted}(f_c) &= \tau^{raw}(f_c) - \omega(f_c)t \\ \omega(f_c) &= \xi f_c + \mu. \end{aligned}$$

Where $\tau^{adjusted}(f_c)$ and $\tau^{raw}(f_c)$ are the adjusted and raw phase at frequency f_c, respectively. t is the time elapse from starting up. $\omega(f_c)$ is the phase offset at frequency f_c. ξ and μ are the slope and intercept of the linear phase shift in frequency domain.

D. Phase Based Ranging

After obtaining the phase changes at different frequencies, we use the phase change of the selected signal in each frame to determine the LoS path length change. In each frame, assume a speaker transmits signals at N_f frequencies. When the receiver moves close/away, the phase of acquired signals would increase/decrease. As the phase of the signal increases by 2π, the path length would decrease by one wavelength of the sound wave. The phase change of the signal at i-th frequency is denoted as $\Delta\theta_i + 2\pi k_i$, where $\Delta\theta_i$ denotes the wrapped phase change of the frequency relative to initial phase, $\Delta\theta_i \in [0,2\pi]$, and k_i is an integer which denotes the number of phase wraps during the movement, $i = 1,2,...,N_f$. When the phase varies from π to $-\pi$ / from $-\pi$ to π, k_i increases/decreases by 1. We can use the real-time distance change to calculate the speed, direction of the movement.

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.2019.2903211, IEEE Internet of Things Journal

Then, the transmission distance change Δl_i of the signal can be expressed as

$$\Delta l_i = -\frac{\Delta\theta_i + 2\pi k_i}{2\pi}\lambda_i. \tag{4}$$

Where λ_i represents wavelength of the signal *i*-th frequency. During the movement, RAMTEL calculates the transmission distance change at each frequency, which can be written as $\Delta \boldsymbol{l} = \left[\Delta l_1, \Delta l_2, ..., \Delta l_{N_f}\right]$. If there are no interference of multipath fading effects, the transmission distance changes at each frequency should be equal, *i.e.* $\Delta l_1 = \Delta l_2 =, ..., = \Delta l_{N_f}$, which is equal to the moving distance of receiver in the frame. However, due to multipath fading, the calculated transmission distance changes are under different interference. Thus, we propose an algorithm to select the signal with the smallest impact of multipath fading from the $N_f$signals to estimate the moving distance, and mitigate the impact of multipath fading. Detailed procedures of multipath mitigation are given in Section IV. Then, the moving distance l_{sum} of successive frames can be estimated as

$$l_{sum} = \sum l_{frame}. \tag{5}$$

Where l_{frame} is the calculated transmission distance change with the smallest impact of multipath fading between the N_f measurements in each frame. The phase of receiver changes as the movement of receiver in real-time, so that RAMTEL could detect small and slow movements which are unable to be measured by transitional Doppler based methods.

IV. Combating Multipath using Extreme Learning Machine

In practical system, sound signals propagate along a straight line to the receiver in free space. Due to the reflection of wall, floor, and other objectors, the received signal is a superimposition of the LoS signals and the reflected signals. Sometimes the receiver obtains different phase of the signal at the same frequency. It is difficult to distinguish the phase change caused by the length change of direct path. However, prior works [11] [18] neglect the influence of multipath effects in active ranging system. According our extensive experiments, the multipath effects should be taken into consideration in the active tracking systems. The multipath effects would lead to periodic attenuation of the received signals, which not only brings error in the estimation of moving distance, but also incorrect moving direction. Some new studies use novel features to evaluate wireless channel model, such as using the fractal coverage characteristic to evaluate the performance of the small-cell networks [24]. However, few studies have investigated to mitigate the effect of multipath fading for phase based acoustic measurements.

A. Multipath Fading Effect

Suppose that the receiver signals are the superimposition of signals from N_p paths and each path has different delay and attenuation. In the paths, the i-th signal $R_i(t)$ has delay τ_i and amplitude a_i. Then, the receiver signal $R(t)$ is

$$R(t) = \sum_{i=1}^{N_p} R_i(t). \tag{6}$$

where $R_i(t) = a_i \cos(2\pi f_c t - \tau_i)$. Thus, it is difficult to obtain the actual phase change from the superimposition of signals traveled from different paths.

To address this issue, we propose an algorithm which could estimate the impact of multipath fading effects on the phase-based ranging and motion tracking. We use a speaker to transmit the signal at N_f different frequencies. In each frame, the receiver could measure the phase change at each frequency independently using the BPF, and estimate the transmission distance changes based on the phase-based ranging proposed in Section III-D, *i.e.* $S = \left\{\Delta l_1, \Delta l_2, ..., \Delta l_{N_f}\right\}$, where Δ_i is the distance estimated by the signal at *i*-th frequency. The signals at different frequencies have different wavelengths and are transmitted through the same multiple paths to the receiver. So, the phase changes of the same multipath paths are different under different frequencies. We leverage the fact that the multipath fading effects have different impact on the phase-based measurement of different frequencies at the same time due to their different wavelengths and phases. Thus, in each frame, we can use the phase change with the smallest effects to estimate the moving distance, thus reducing the impact of multipath fading effects, *i.e.*

$$l_{frame} = \arg\min_{\Delta l \in S} \mathrm{f}(\Delta l). \tag{7}$$

Where $f(.)$ is the function of calculating the impact of multipath fading effects. l_{frame} is the distance estimated by the signal with the smallest impact of the effects in the frame.

So how to find the $f(.)$ and determine the impact of multipath fading effects on the signals? Previous work [25] uses linear regression to remove abnormal estimation at certain frequencies and calculates the distance change using the remaining frequencies. However, this algorithm needs high bandwidth to ensure good regression results, and couldn't been used in our system due to limited bandwidth in each speaker. In this paper, we propose a novel algorithm to evaluate the impact of multipath fading effects on the signals at different frequencies using a feed-forward neural network with single hidden layer, *i.e.* Extreme Learning Machine (ELM). We call this multipath mitigation algorithm as Combating Acoustic Multipath using ELM (CAME). We first explain the features we consider, and then provide implementation details of the ELM techniques.

B. Feature Selection and Extraction

The foundation of *CAME* is the accurate feature selection and extraction of signals under different multipath fading effects. ELM is carried out to estimate the impact of multipath fading based on these feature vectors. For the signal at each frequency, we obtain R_α and R_β in the preprocessing step.

In the off-line phase, firstly, we use a speaker to transmit acoustic signals at different frequencies. Then a smartphone receives the signal and moves uniformly in a straight line. We calculate received signals' normalized scatter diagram at each frequency in a frame between R_α and R_β, which are obtained in the preprocessing step, as shown in Figure 5c and Figure 5d. We use the feature of the normalized scatter diagram 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.2019.2903211, IEEE Internet of Things Journal

define whether the phase-based measurement is affected by the multipath fading effects in the frame.

Lemma 1. *When the impact of multipath fading is small, all the points in the normalized scatter diagram at each frequency are located on the edge of a unit circle.*

Proof: It is simpler to treat this case in complex form

$$\begin{aligned} R(t) &= \sum_{i=1}^{N_p} R_i(t) = \mathrm{Re}\left(\sum_{i=1}^{N_p} a_i e^{j(2\pi \mathrm{f}_c t - \tau_i)}\right) \\ &= \mathrm{Re}\left(e^{j2\pi f_c t}\left(a_1 e^{-j\tau_1} + a_2 e^{-j\tau_2} ... + a_N e^{-j\tau_{N_p}}\right)\right) \\ a_0 e^{-j\tau_0} &= \left(a_1 e^{-j\tau_1} + a_2 e^{-j\tau_2} ... + a_N e^{-j\tau_{N_p}}\right) \\ R(t) &= \mathrm{Re}\left(a_0 e^{j(2\pi f_c t - r_0)}\right) = a_0 \cos\left(2\pi f_c t - r_0\right). \end{aligned} \tag{8}$$

The superimposition of signals from multipath paths is a cosine signal. a_0 is the amplitude and τ_0 is the phase of signal. a_0 and τ_0 are related to a_i, and τ_i, and independence with time and carrier frequency. The large scale fading, such as path-loss and shadowing, only introduces the smooth changes of received signals, while small scale multipath fading changes the amplitudes of signals rapidly. The radius of the points a_0 is a function of a_i, and τ_i. The length of multiple paths would change during the movement, which leads to the change of a_i and τ_i. When the impact of multipath fading is small, a_i and τ_i changes slowly in each frame due to the smooth propagation loss of the path. Thus, a_0 has almost the same value in the frame, as shown in Figure 5c. As a result, $R_\alpha^2 + R_\beta^2 = a_0^2 \cos^2(2\pi f_c t - r_0) + a_0^2 \sin^2(2\pi f_c t - r_0) = a_0^2$, all the points have fixed radius a_0 for each frequency and are located on the edge of a unit circle in the normalized scatter diagram. ■

When the impact increases, the small scale multipath fading causes a rapid and irregular change of a_0 in the frame, and all the points in the normalized scatter diagram at each frequency are located on a circular ring, as shown in Figure 5d. The distribution of points on the ring can be used to estimate the impact of multipath fading effects. The more serious multiple fading effects, the more irregular distribution of the points. Thus, we can extract features from the distribution of the points to estimate the impact of multiple paths.

Secondly, based on these observations, we propose to quantify the impact of multipath fading effects based on the features extracted from the normalized scatter diagram. For the signal at each frequency, assume there are N_s samples in a frame, and the Euclidean distance vector between these N_s points and circle center is $\boldsymbol{D}_{ED}$:

$$\boldsymbol{D}_{ED} = [D_1, ..., D_i, ..., D_{N_s}], \tag{9}$$

where D_i is the Euclidean distance between i-th point and the circle center, i = 1,2,...,N_s. $\boldsymbol{D}_{ED}$ can't be used as the feature for training because N_s is usually large (more than 500) which could bring large computation and latency.

In order to reduce the dimension of features and computation overhead, we extract a new feature from the normalized scatter diagram. We divide the normalized scatter diagram into N_c evenly spaced concentric rings ($N_c \ll N_s$), as shown in Figure 5d. Then, we calculate the number of points between adjacent concentric circles, as shown in Figure 5d, and convert the numbers into a vector $\boldsymbol{x}$:

$$\boldsymbol{x} = [F_1, ..., F_j, ...F_{N_c}], j = 1,2,...,N_c. \tag{10}$$

where F_j is the number of points that satisfy

$$\left\{ D_i \left| \frac{j-1}{N_c} < D_i \leqslant \frac{j}{N_c} \right. \right\}, i = 1,2,...,N_s. \tag{11}$$

N_c (we set 30 in our experiments) is much smaller than the length of $\boldsymbol{D}_{ED}$ (we set 512 in our experiments). Thus, the $\boldsymbol{x}$ is selected as a feature vector for training set without introducing large computation overhead, and we call this feature as Multipath Dispersion Vector (MDV).

Thirdly, the application of machine learning to new problems requires labeled training data. So, each input features vector of ELM should be labeled with a training target value which could describe the impact of multipath mathematically. In each frame, we collect the estimated distance $\widehat{P}$, in conjunction with a measure of certainty P. Then, we use the Ranging Error Rate (*RER*) (denoted by t)as the training target value corresponding to the feature vector in the frame

$$t = \frac{\widehat{P} - P}{P}. \tag{12}$$

The *RER* indicates the impact of multipath fading effects to the measurement. The small t means the estimated distance is close to the measure of certainty, and indicates small multipath fading interferes on the measurement.

In the off-line phase, we collect M training samples in multipath fading indoor environments. Specifically, for the i-th training sample, we calculate the MDV as the input feature vector $\boldsymbol{x}_i$, and measure the error rate t of the signal as the output training target value t_i. The training sample can be denoted as $(\boldsymbol{x}_i, t_i)$, $i = 1, 2, ..., M$. Then, ELM use the training samples to evaluate the impact of multipath fading at different frequencies.

C. ELM for Multipath Mitigation

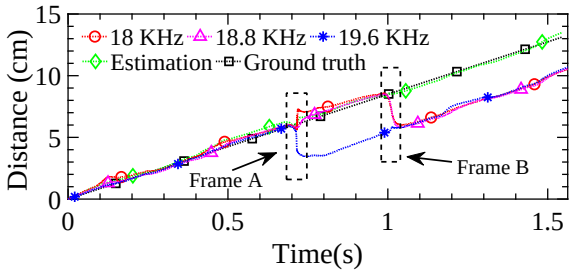
ELM is a Single Hidden Layer Feed-forward Neural Network (SLFN) [26], which is adopted to evaluate the impact of multipath fading in our experiments.

Assume the training set has M samples, *i.e.* $(\boldsymbol{x}_i, \boldsymbol{t}_i)$, where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in R^n$, $\boldsymbol{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in R^m$, and $j = 1, 2, ..., M$. n is the dimension of each feature vector $\boldsymbol{x}_i$, and m is the length of output vector $\boldsymbol{t}$. Standard SLFNs with L hidden nodes are mathematically modeled as

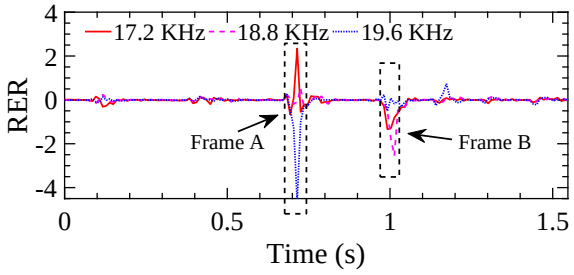
$$\begin{aligned} \sum_{i=1}^{L} \boldsymbol{\beta_i} g_i(x_j) = \sum_{i=1}^{L} \boldsymbol{\beta_i} g_i(\boldsymbol{\omega_i} \cdot x_j + b_i) = t_i \\ j \in 1,2,...,M. \end{aligned} \tag{13}$$

Where the hidden node $g(\cdot)$ is a nonlinear activation function. $\boldsymbol{\omega}_i = [\omega_{i1}, \omega_{i2}, ..., \omega_{im}]^T$ is the weight connecting the i-th hidden node and the input nodes. $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ is the weight connecting the i-th hidden node and the output nodes. b_i is threshold of the i-th hidden nodes. L is the number of hidden nodes.

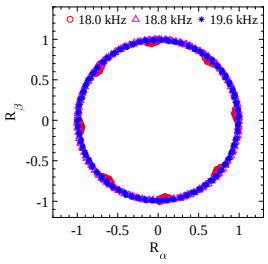
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2903211, IEEE Internet of Things Journal



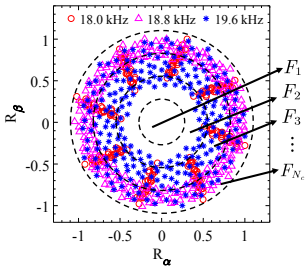
(a) The distance measured by signals at different frequecies and the estimated distance using ELM



(b) The *RER* of the signal at different frequecies are estimted by ELM, which are used to evaluate the impact of multipath effects in real-time.



(c) The points in the normalized scatter diagram are located on the edge of a unit circle, which indicates small effects of multipath fading.



(d) In Frame A, the points in the normalized scatter diagram are scattered in a circle due to strong multipath fading effects.

Fig. 5. In order to mitigate multipath effects, we propose a new feature (MDV) to determine the impact of multiple paths on the signal at different frequencies using ELM.

In ELM, the parameters $(\boldsymbol{\omega}_i, b_i)$ of the hidden nodes are randomly assigned without any iterative tuning, which makes it faster than traditional SLFNs and shows good generalization performance in real-world application [27]. Since we already know the feature vector $\boldsymbol{x}_i$ and $(\boldsymbol{\omega_i}, b_i)$ is randomly assigned, the hidden layer output matrix $\boldsymbol{H}$ can be easily calculated

$$\begin{aligned}&\boldsymbol{H}\left(\boldsymbol{\omega}_1,...,\boldsymbol{\omega}_L,b_1,...,b_L,\boldsymbol{x}_1,...,\boldsymbol{x}_M\right)\\&=\begin{bmatrix}g\left(\boldsymbol{\omega}_1\cdot\boldsymbol{x}_1+b_1\right)&\cdots&g\left(\boldsymbol{\omega}_L\cdot\boldsymbol{x}_1+b_L\right)\\\vdots&\ddots&\vdots\\g\left(\boldsymbol{\omega}_1\cdot\boldsymbol{x}_M+b_1\right)&\cdots&g\left(\boldsymbol{\omega}_L\cdot\boldsymbol{x}_M+b_L\right)\end{bmatrix}_{M\times L}\\&\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_1^T\\\vdots\\\boldsymbol{\beta}_L^T\end{bmatrix}_{L\times m},\boldsymbol{T}=\begin{bmatrix}\boldsymbol{t}_1^T\\\vdots\\\boldsymbol{t}_M^T\end{bmatrix}_{M\times m}\end{aligned}$$

Then, Equation 13 can be written compactly as

$$\boldsymbol{H\beta}=\boldsymbol{T} \tag{14}$$

Refer to the analysis of constrained-optimization-based ELM in [28], the output weight matrix $\boldsymbol{\beta}$ is established as

$$\boldsymbol{\beta}=\left(\frac{\boldsymbol{I}}{C}+\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{\boldsymbol{T}}\boldsymbol{T}. \tag{15}$$

Where C is a user-specified parameter (regularization factor), which makes the solution more stable but introducing bias. So, the parameter $1/C$ should also be a very small value while still maintaining model stability. In our system, we have found that better results are usually achieved at small parameter $1/C$. For simplicity, C is specified as 2^{15}.

In our experiments, we adopt a $sigmoid$ function as the nonlinear activation function $g(\cdot)$. The number of hidden node is set as $L = 50$. All the hidden-node parameters $(\boldsymbol{\omega}_i, b_i)$ are randomly generated with the uniform distribution. The dimension of feature vector $\boldsymbol{x}_i$ is set as $n = 20$. Since we only use the *RER* as the training target value to label a feature vector, the length of output training target vector $\boldsymbol{t}$ is set as $m = 1$.

In the off-line phase, we collect 1200 training samples in multipath fading indoor environments, and calculate the output weight matrix $\boldsymbol{\beta}$ using Equation 15. In the on-line phase, we use the speaker to transmit signals at N_f frequencies. In each frame, we calculate input feature vector $\boldsymbol{x}(i)'$ at the i-th frequency. The predicted *RER* (denoted by $t'(i)$) corresponding to $\boldsymbol{x}(i)'$ can be denoted as

$$t'(i)=\boldsymbol{x}(i)'\boldsymbol{\beta}. \tag{16}$$

The predicted ranging error rate vector, *i.e.* $t' = [t'(1), t'(2), ..., t'(N_f)]$ evaluates the impact of multipath fading to the phase-based ranging at different frequencies. In each frame, we select the signal with the smallest *RER* to estimate the movement distance in the frame. The pseudocode of CAME algorithm is in Algorithm 1. Note that, compared with signatures based methods which collect features in off-line phase, and estimate the position of the devices using machine learning technologies, our method leverages the features to select a signal with the smallest impact of multipath fading, and then use a model-based method to tracking devices' motion. Thus, our work is less influenced by the database collected in off-line phase.

An example to illustrate the migration of multipath fading effects is shown in Figure 5. A speaker continuously transmits acoustic signals at 18 KHz, 18.8 KHz and 19.6 KHz at the same time. Then, we move an Android smartphone away uniformly from the speaker at the distance of 1 meter. The smartphone obtains the signals frame by frame, and each frame has 512 sampling points with 48 KHz sampling rate. we calculate the *RER* in each frame, as shown in Figure 5b. During the movement, we measure the moving distance using the method in Section III-D at each frequency, and the moving distance at each frequency should be the same in each frame if there are no multipath fading effects. However, due to multipath fading effects, the measured distances are obviously different at 18 KHz and 19.6 KHz in Frame A, and at 18 KHz and 18.8 KHz in Frame B, as shown in Figure 5a. In Frame A, the *RER* of the signals at 18.8 KHz are obviously smaller than 18 KHz and 19.6 KHz. The larger *RER* indicates larger ranger

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.2019.2903211, IEEE Internet of Things Journal

error caused by multipath fading, which is consistent with the measurements in Frame A, as show in Figure 5a. Thus, the signals at 18.8 KHz can be selected to estimate the moving distance due to the smallest *RER* in the frame. In Frame B, the *RER* of the signals at 19.6 KHz is obviously smaller than 18 KHz and 18.8 KHz, so that the signal at 19.6 KHz is chosen to estimate the distance. We compare our estimation with the actual distance, and the result indicates that CAME could evaluate and mitigate the impact of multipath fading effects.

Algorithm 1: CAME algorithm

Input: Distance change estimated by the phase changes of signals at different frquenceis from a speaker;
Output: Moving distance relative to the speaker;
foreach *frame* **do**
  foreach *frequency* **do**
    Calculate the distance to the center of each point in the nomailized scatter diagram using R_α and R_β;
    Calculate the number of points between adjacent concentric circles;
    Extract MDV as the input feature vector of ELM;
    Use ELM to estimate *RER* at each frequency which indicates impact of multipath fading on the signal;
  end
  Calculate the moving distance using the signal with the smallest *RER*;
end
Sum the calculated the moving distance in each frame, and obtain moving distance relative to the speaker;

V. Motion Tracking Algorithm

In this section, we first use a calibration scheme to obtain a reference position. Then, we combine the initial position with the fine-grained distance change to enable 2-D motion tracking.

A. Estimating Reference Position

The phase based algorithm in Section III-D only measures the relative distance change, which is not sufficient for motion tracking. We cannot determine actual position only using the relative distance change due to the lack of the initial position. One way is to measure using some additional tools, such as ruler, hand-hold distance finder. However, those methods are cumbersome and error-prone. In this subsection, we propose a calibration method to estimate reference position. We choose two of the speakers assigned with different frequencies to estimate the initial position. Without loss of generality, we assume that two speakers A and B are placed along an x-axis. The coordinates of A and B are given, as shown in Figure 6.

We let a user move a smartphone parallel to the x-axis with unknown distance a. As the mobile is moving close/away each speaker, the distance between the smartphone and the speakers reduces/increases. When smartphone moves to the C/D position, the relative distance between A/B and the smartphone has minimum value. So we can determine the position of C and D on the moving path when the distances become minimum. We could calculate the difference d_{ac} of the relative distance between the smartphone and speaker A at C and D. Due to the property of a rectangular triangle,

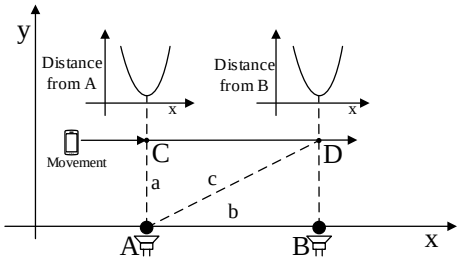


Fig. 6. Estimating reference position by moving a smartphone parallel to the x-axis.

$a^2+b^2=c^2$ and $c=d_{ac}+a$, the distance a can be calculated by

$$a=\frac{b^2-d_{ac}^2}{2d_{ac}} \tag{17}$$

Where b denotes the absolute distance between A and B, and c denotes the absolute distance between A and D. Thus, the coordinates of C and D can be obtained, and can be used as the reference position. To improve the accuracy, we can sweep the smartphone along the moving path multiple times, and use the mean positions as the estimation for C and D.

B. Tracking Motion by Computing Real-time Position

In order to track the device's motion, we should obtain the real-time position of the devices.

The range measurement between the smartphone and the *i*-th (i = 1,2,3,...,N) speaker is denoted as $\hat{d}_i$, where N is the number of speakers. Let $[x,y]^T$ be the unknown coordinate of the smartphone, and let $[x_i,y_i]^T$ be the known coordinate *i*-th speaker. The known coordinate of the reference position introduced in the previous subsection is $[x_R,y_R]^T$.

The error-free distance change between the smartphone and *i*-th speaker relative to the distance at reference position is calculated as

$$d_i=\sqrt{(x-x_i)^2+(y-y_i)^2}-R_i \tag{18}$$

where $R_i=\sqrt{(x_R-x_i)^2+(y_R-y_i)^2}$. The measurements of the range differences are modeled by

$$\hat{d}_i=d_i+\varepsilon_i, \qquad i=1,2,3,...,N \tag{19}$$

where ε_i is the measurement error of $\hat{d}_i$.

The LS error function is then defined as the difference between the measured and true values

$$\mathbf{e}=\hat{\mathbf{d}}-\mathbf{d} \tag{20}$$

where $\hat{\mathbf{d}}=\left[\hat{d}_1,\hat{d}_2,...,\hat{d}_N\right]^T$ and $\mathbf{d}=[d_1,d_2,...,d_N]^T$.

We define vector $\mathbf{\Lambda}\triangleq\left[x \;\; y \;\; x^2+y^2\right]^T$. Then, we can rewrite Equation 20 in matrix form as

$$\mathbf{e}=\mathbf{A}\mathbf{\Lambda}-\mathbf{b} \tag{21}$$

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.2019.2903211, IEEE Internet of Things Journal

where

$$\mathbf{A} = \begin{bmatrix} -2x_1 & -2y_1 & 1 \\ -2x_2 & -2y_2 & 1 \\ \vdots & \vdots & \vdots \\ -2x_N & -2y_N & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \left(\hat{d}_1 + R_1\right)^2 - x_1^2 - y_1^2 \\ \left(\hat{d}_2 + R_2\right)^2 - x_2^2 - y_2^2 \\ \vdots \\ \left(\hat{d}_N + R_N\right)^2 - x_N^2 - y_N^2 \end{bmatrix}$$

Finding the LS solution based on the LS criterion of Λ is a linear minimization problem, can be as

$$\min_{\mathbf{\Lambda}} \left[\mathbf{A\Lambda} - \mathbf{b}\right]^T \left[\mathbf{A\Lambda} - \mathbf{b}\right]. \tag{22}$$

According to LS algorithm, the solution minimizing is given by

$$\mathbf{\Lambda} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{23}$$

The unknown coordinate of the smartphone can be expressed as

$$[x, y]^T = \left[\mathbf{\Lambda}\left(1\right), \mathbf{\Lambda}\left(2\right)\right]^T. \tag{24}$$

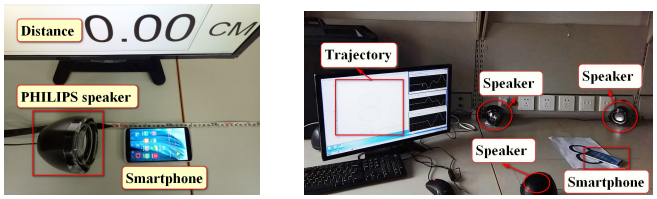
Note that the smartphone's coordinate is updated in each sampling time, and we can track the smartphone's motion using the real-time coordinate. The pseudocode of our motion tracking algorithm is in Algorithm 2.

Algorithm 2: RAMTEL's motion tracking algorithm

```
Input: An audio signal segmentation acquired by smartphone's
       bottom microphone;
Output: Updated smartphone position;
if uninitialized then
    Calibrate phase offsets due to asynchronous system clocks
      between sender and receiver;
    Find a reference point;
end
foreach speaker do
    foreach frequency do
        Apply BPF to measure each frequency independently;
        Use Park transformation to calculate the real-time phase of
          received signals;
        Calculate the phase change relative to the phase at
          reference position; Estimate the 1-D distance change
          based on the phase change at different frequencies;
    end
    Use CAME to mitigate multipath fading effects based on the
      estimated 1-D distance at different frequencies;
    Update the relative distance to the reference point using the
      distance change;
    Update the absolute distance to each speaker using the relative
      distance and the position of reference point;
end
Apply LS algorithm to calculate the localization and update the
  smartphone position;
```

VI. IMPLEMENTATION

We develop RAMTEL on two platforms. The first platform consists of a speaker connected with a GIGABYTE desktop with Intel I7 CPU and 8 GB RAM, and is used to evaluate the accuracy of 1-D ranging and the performance of CAME. In order to have larger degree of freedom and achieve 2-D tracking, the second platform adds multiple speakers on the basis of the first platform, and is used to demonstrate the feasibility of motion tracking in real-time. Specifically, the smartphone is a ZTE B2015 mobile phone with Android 5.1 Operating System. The desktop has a VIA sound card which could support at most 8 speakers. Some PHILIPS PA311/93 speakers ($8 each) connect to the desktop and transmit inaudible signals at different frequencies, each at certain frequency. We use the bottom microphone of the smartphone to receive the acoustic signals with the sampling rate of 48 KHz, which is supported by most mobile devices. Then, we use the bottom microphone of the smartphone to record the sound wave with the sampling rate of 48 KHz, which is supported by most smartphones. Inspired by the task scheduling in fog networks [29] [30], we offload the computation of ranging and tracking to the desktop in order to extend batter lifetime.



(a) Testbed setup for 1-D ranging

(b) Testbed setup for the performance evaluation of motion tracking.

Fig. 7. Experimental Setup

VII. EVALUATION

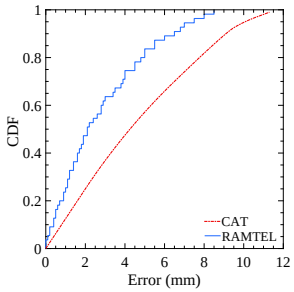
For 1-D ranging, we move the smartphone close/away the audio source, and record the distance change during the movement. Since the distance between them changes over time, we use a rule along the moving path that collects the ground truth data, as shown in Figure 7a. We use the platform to evaluate the error of 1-D ranging, the multipath effects mitigated by CAME. Then, we compare the performance of multipath mitigation using the linear regression algorithm and CAME, and evaluate the impact of different MDV lengths on the measurements. In order to evaluate RAMTEL's performance of detecting small and slow movement, we use the platform to monitor the respiratory rate of some volunteers at different distances. Then, we evaluate impact of typical noises in indoor environments on the measurements, such as ambient noise, discussion noise, and music noise, and the performance of 1-D ringing in NLoS scenarios. Finally, we also move obstacles' movement near the receiver to estimate the impact of dynamic reflection signals on the measurements.

For 2-D tracking, we use a three-speaker system as shown in Figure 7b. The separation between adjacent speakers is 70 centimeters. Each speaker is allocated 0.8 KHz bandwidth with 200 Hz interval. The three-speaker system occupies 17.2-19.4 KHz, which are virtually inaudible to most people. We first evaluate the performance of motion tracking by evaluating the median error between the trajectories tracked by RAMTEL versus the ground truth trajectories recorded by a camera. Then, we compare our method with Doppler based and FMCW based methods. Finally, we evaluate the impact of training set size on performance of motion tracking. The ranging, tracking, and visualization, are both done on-line in real-time.

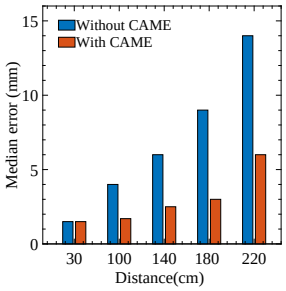
A. Experimental Results

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.2019.2903211, IEEE Internet of Things Journal

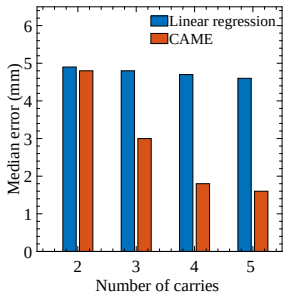
1) The accuracy of 1-D ranging: RAMTEL achieves an average movement distance error of 2 mm when the smartphone moves for 40 cm at a distance of 1 m. In our experiments, we use RAMTEL to measure the distance changes between a smartphone and a speaker, and use a rule to collect ground truth data. The smartphone's initial position is 1 m away from the speaker, and moves away from the speaker for a distance of 40 cm. Figure 8a plots the Cumulative distribution function (CDF) of the 1-D relative distance measurement error for 100 measurements. The median error is 2 mm, and 90-th percentile error is 6 mm. We also compare our approach with CAT, a moving distance measurement approach based on FMCW [11]. Results show that our approach outperforms CAT in terms of distance measurement accuracy by 50% on average.



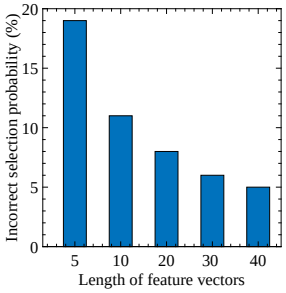
(a) CDF for 1-D ranging



(b) The effects of multipath fading reduced by CAME



(c) Median error of using different number of carriers



(d) The probability of incorrect selection using feature vectors with different length

Fig. 8. The accuracy of 1-D ranging, and the performance of multipath fading mitigation

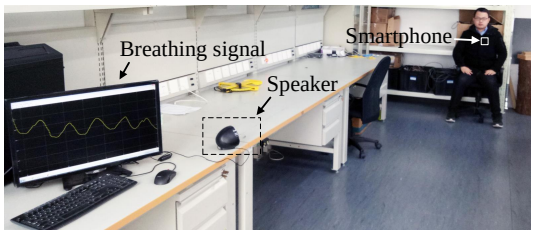
2) Performance of multipath mitigation: In this experiment, we examine the performance of multipath mitigation using *CAME*. Particularly, we conduct following experiments: we use the speaker to play the sinusoidal audio signals at four frequencies. We first estimate the moving distance using the signal at one of the frequencies without using *CAME*, and using *CAME* with four carriers, respectively. Then, we calculate the measurement error at different distances. The comparisons of the results from the estimation without/with CAME are shown in Figure 8b. We collect the data from 50 measurements for single-carrier and multi-carrier respectively. We observe no significant difference between two cases at a distance of 30 cm, and the median error is about 1.5 mm, which indicates that multipath fading effects have limited impact at the distance. The median error increases as the distance increases because the effects of multipath fading are strengthened as the transmission distance increases. The median error of using *CAME* is significantly lower than without mitigating multipath fading effects. When the distance is 220 cm, *CAME* reduces median error from 14 mm to 6 mm. The results indicate *CAME* can improve the ranging accuracy and the robustness against effects of multipath fading.

We compare the performance of *CAME* with the linear regression based algorithm proposed in [25], as discussed in Section IV-A. In our experiments, a speaker transmits signals at different number of carriers, and we calculate the median error of ranging at a distance of 1 meter. We repeat each measurement for 50 times, and results are shown in Figure 8c. When 2 carriers are used, CAME, and linear regression algorithm have similar performance which is close to with using any multipath mitigation algorithm, because both the algorithms require multiple carriers to estimate the accuracy moving distance under effects of multipath fading, and 2 carriers cannot satisfy the requirement. We can see that the median error of using *CAME* reduces as the number increases. When four carriers are used, the error reduces to 2 mm, while linear regression based algorithm cannot reduce the median error. This is because linear regression based algorithm requires wider bandwidth to ensure good regression results, while CAME has more stable and robust performance using limited bandwidth. As we would expect, the results indicate RAMTEL can achieve good ranging performance with a limited bandwidth of acoustic signals (*e.g.*, 0.8 KHz), which ensures great robustness for practical usage on mobile devices.

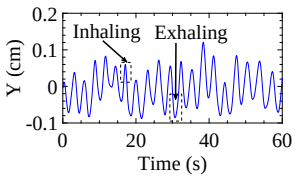
We also examine the impact of the length (N_c) of MDV used in ELM (Equation 10 in Section IV-A). The aim of *CAME* is to select the signal with the smallest impact of multipath fading effects from the signals at different frequencies. So, in our experiments, we use the probability of selecting incorrect signal (not the smallest one) to indicate the impact of N_c to our system. We use the same experiment setup in Section VII-A1. We collect training samples using different length of MDVs. In on-line phase, ELM estimates the *RER* using Equation 16 in Section IV-A at each frequency using the same length of MDV as its training samples. Meanwhile, we collect the actual values of *RER* (Equation 12 in Section IV-A) corresponding to the estimated values. In each frame, the smallest actual value and estimated value at the same frequency indicates that *CAME* selects the correct signal to estimate the moving distance. Figure 8d plots the probability of incorrect selection as we vary the length of MDV used in the on-line and off-line phase of the *CAME*. When 5 features are used in training and estimating *RER*, as the N_c increases, the probability reduces significantly. With 30 features in a MDV (the default value in our evaluation), the probability of selecting an incorrect signal reduces to 6%. We calculate the probability by repeating 100 measurements for each length of feature vectors, and the results indicate that the MDV would be extracted as the feature for ELM and reduces computation overhead effectively.

3) The detection of small and slow movement: RAMTEL could detect small and slow movements in real-time, because it uses the phase change of received signals to estimate moving movement. We use the accuracy of respiratory rate monitoring to evaluate RAMTEL's performance of detecting small and slow movement. We recruit six participants (3 females and 3

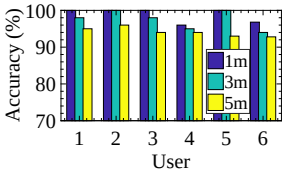
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2903211, IEEE Internet of Things Journal



(a) Experimental setup for respiratory monitoring

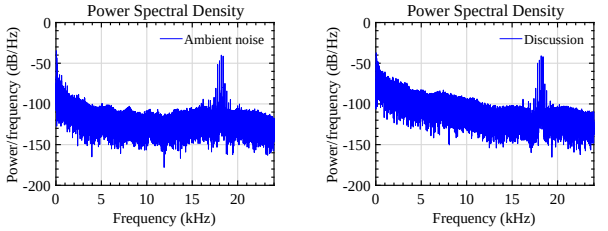


(b) Chest's small and slow movement measured by RAMTEL

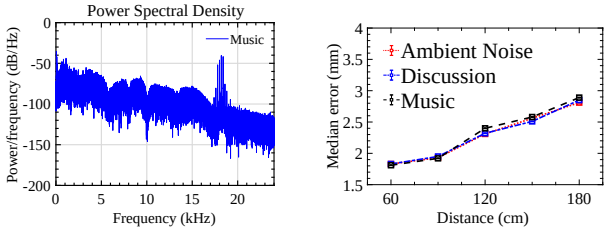


(c) Monitoring accuracy

Fig. 9. Small and slow movement monitoring.



(a) Ambient noise (b) Discussion

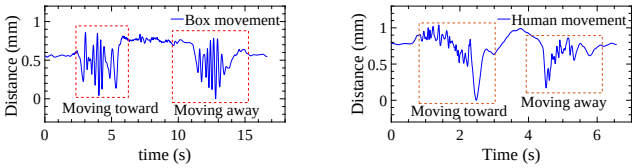


(c) Music (d) Difference cases

Fig. 10. Power spectral density of three noise cases measured at 1 meter and 1-D ranging performance of the cases.

males) between the ages of 24-28. The participants are asked to wear a cloth coat with a phase-compensated smartphone placed in its chest pocket, and sit at a distance of 1, 3, and 5 meters to a speaker, as shown in Figure 9a. We calculate participants' respiratory rate according to the ups and downs of their chest which lead micro movement to the smartphone. The micro movement direction and amplitude can be measured by RAMTEL, as shown in Figure 9b. We also ask the participants to take a stopwatch to record the times of inhalation or exhalation which could calculate participants' respiratory rate and can be seen as an actual value. We collect the data from 5 minutes' measurement for each user, and compare RAMTEL's measurement with the actual value, the average monitoring accuracies for each user are shown in Figure 9c. The average respiratory rate monitoring accuracies for different users at the distance of 1 meters are in the range of 96%-100%. When the distance increases to 5 meters, the monitoring accuracies are in the range of 92.8%-96%, with an average accuracy of 94% over all users. The results indicate that RAMTEL could detect slow and sight movements accurately.

4) Performance under noisy conditions: In our experiments, we evaluate the performance of RAMTEL under different noisy conditions: ambient noise in office area, speaking, self-interference. First, we measure ranging error in the laboratory with ambient noise. The ambient noise in the laboratory is mainly caused by the outlets of the central air conditioner, computer fans, and some other electrical devices. Second, we recruit a couple of students to have a group discussion at a distance of 30 cm from the smartphone. Finally, we use the smartphone to play different kinds of music (*e.g.*, Pop, Classic, and Rock) together to examine the impact of self-interference. Figure 10 plots the Power Spectral Density (PSD) measured by a smartphone at a distance of 1 meter from a speaker in different noisy conditions. The results show that human discussion significantly improves the PSD from -100 dB/Hz to -50 dB/Hz on average at the frequencies less than 1 KHz. Music played by the smartphone improves the PSD in a wider range of frequencies, from 100 Hz to 16 KHz. However, even though the ambition noise and human has significant impact on the PSD, the PSD of audio signal for relative distance measurement couldn't be affected and still has -50 dB/Hz on average, so that the signal can be easily detected. We measure the median error of ranging in different conditions, and repeat the experiment 50 times for each case. Figure 10d shows that the performance for all three cases at different distances are similar. These experimental results indicate that RAMTEL is robust against common indoor noises and not affects the normal use of the smartphone.

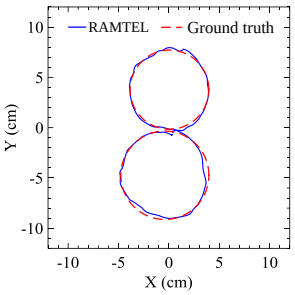


(a) Box movment (b) Human movement

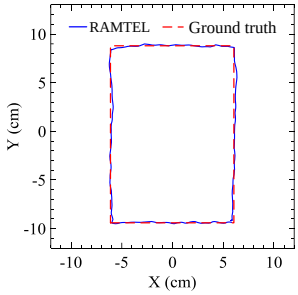
Fig. 11. The movements of box and human near the receiver have limited impact on our measurements.

5) Impact of movement in the surrounding: In our experiments, we examine the impact of movement in the surrounding by measuring the accuracy of 1-D measurement while we move different objects near a RAMTEL-enabled smartphone. Figure 11a shows the sample measurement when we move a paper box $(20cm \times 15 \times cm \times 5cm)$ toward/away from the smartphone perpendicular to the direction of audio propagation. We can observe that the distance remains static when the box is not moving and varies like sinusoids when the box moves towards/away from the smartphone. The human movement at 20 cm from the smartphone also brings fluctuation to the distance measurement, a sample measurement as shown in Figure 11b. We also move our hands at the same distance, and observe the similar fluctuation as that caused by

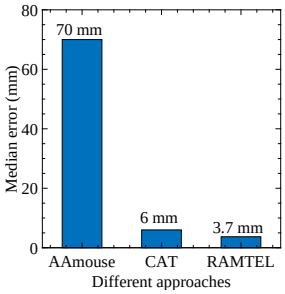
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2903211, IEEE Internet of Things Journal



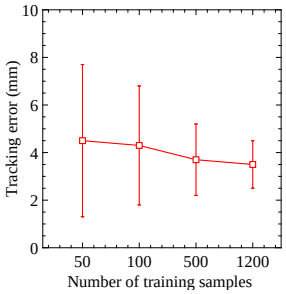
(a) Trajectory of double-circle



(b) Trajectory of rectangle

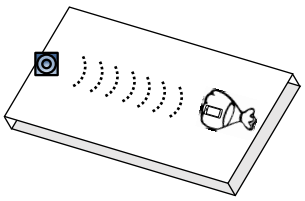


(c) Median error for motion tracking using different algorithms

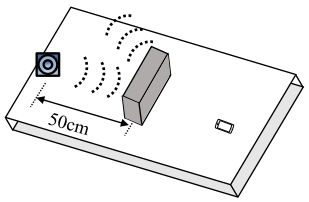


(d) Tracking error using different number of training samples

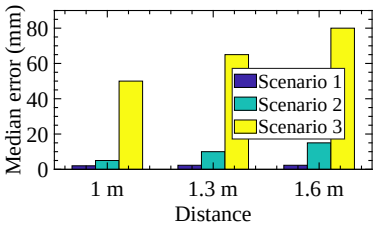
Fig. 12. Motion tracking performance.



(a) A smartphone placed in a cloth bag



(b) A paper box placed between a speaker and a smartphone



(c) The median ranging error in the scenarios

Fig. 13. 1-D Ranging performance of in the NLoS scenarios.

human movement. We repeat this experiment 50 times, and find that the movements near the receiver has slight impact on our measurement (less than 0.5 mm on average). This is because that the fluctuation is caused by the change of reflected paths received by the smartphone, and we use direct-path to measure the distance. Further, the fluctuation doesn't affect the measurement permanently, and the ranging error will be eliminated when the object moves away from the smartphone.

6) Performance of 1-D ranging in NLoS scenarios: RAMTEL leverages the phase change of LoS signals to estimate moving distance. However, it doesn't mean that the LoS signals can't be blocked, and it comes to whether the LoS signals are stronger than NLoS signals. In order to evaluate the performance of 1-D ranging in NLoS scenarios, we conduct our experiments in three scenarios. Scenario 1: we measure the median ranging error at different distances without objects that block the LoS signals as ground truth data. Scenario 2: we put the smartphone in a cloth bag as a receiver which imitates a smartphone placed in a pocket, then we move the bag with the smartphone to evaluate the ranging performance as Scenario 1. Scenario 3: we put a paper box $(30cm \times 21 \times cm \times 5cm)$ at a distance of 50 cm to the speaker so that the LoS signals are blocked while NLoS signals can be received, and conduct the same experiments as Scenario 1. Figure 13 shows the experimental setup and results of experiments in the scenarios at different distance, specifically 1 m, 1.3 m, and 1.6 m. We repeat the experiment at each distance for 50 times in each scenario and the results show that the median errors increase in Scenario 2 and Scenario 3 compared with Scenario 1. The median errors in Scenario 2 are obviously less than that in Scenario 3, and are allowable for practical applications, such as real-time gesture recognition. In Scenario 3 , the median ranging errors increase significantly, and grow to 80 mm at the distance of 1.6m. This is because that the LoS signals and NLoS signals have similar fading through the cloth bag in Scenario 2 and the phase change of LoS signals can be approximately calculated, while the phase change of LoS signals can't be obtained in Scenario 3 due to the obstruction of the paper box. In order to overcome the challenge of large median errors in Scenario 3, we consider to use extra-speakers as source to make sure that the smartphone can obtain LoS signals from at least three speakers. Then, the smartphone determines which signal is LoS and choose all the LoS signals to track its motion, and we leave this in future works.

7) The accuracy of motion tracking: We use a RAMTEL-enabled smartphone to draw a double-circle, and a rectangle in 2-D space. We use a camera to collect the ground-truth trajectories of the smartphone, and the experimental setup is shown in Figure 7b. We compare the measured trajectories versus the original trajectories to quantify the accuracy. Then, we evaluate the tracking error by measuring the average least perpendicular distance of each point in the trajectories estimated by RAMTEL with the closest point in the ground-truth trajectories. We repeat the experiment 50 times for each pattern, and compute the average error in each grid by averaging across the experiments. Figure 12a and 12b show the sample results of the estimated trajectories and ground truth trajectories. We compare the median error of tracking with CAT and a Doppler effect based method [9] (denoted by AAMouse), as shown in Figure 12c. The median error of RAMTEL is 3.7 mm, while that of CAT and AAmouse are 6 mm and 7 cm respectably. The results show that the tracking accuracy of RAMTEL out-performs CAT and AAMouse significantly.

8) Impact of training set size: In order to estimate the impact of training set size, we evaluate the tracking performance using different number of training samples. In the off-line phase, we collect 1200 training samples (default value in our evaluations), and choose different number of them to generate output weight matrix using Equation 15 in Section IV-C. In the on-line phase, RAMTEL tracks the smartphone

in 2-D plane using the output weight matrices generated by different number of training samples. Figure 12d plots the median error of motion tracking and the vertical lines denote standard deviation corresponding to the median error. The results indicate that RAMTEL is not evidently depend on the size of training set, because only 50 training samples could provide a fine-grained motion tracking with about 4.5 mm median error. Further, the results also show that the increase of training samples could improve the performance of motion tracking significantly since it reduces the deviation of tracking error and provides more stable results.

VIII. LIMITATIONS

In this section, we discuss some limitations of our current system and potential directions for future work. First, the total bandwidth of sound is limited on smartphone. In RAMTEL, each speaker is allocated 0.8 KHz bandwidth separated by 200 Hz guard band to avoid carrier interference. The guard band and the band-pass filter used in our system limit the maximum movement speed (about 30-40 cm/s) of devices that could be accuracy tracked. In future work, we plan to reduce the limitation of speed by improving bandwidth. For example, we plan to use a smartphone as source and some microphones as receivers, and, so that different degrees of freedom can share the total bandwidth.

Second, RAMTEL selects the signals with the smallest *RER* from a speaker to estimates the moving distance between the speaker and mobile devices. However, there are a few cases where multipath fading effects are strong, and the signals at all the frequencies are interfered by the effects, so RAMTEL cannot obtain accurate moving distance. This can occur when the receiver is far from the speaker or direct path signals are blocked by impenetrable obstacles, such as human body and wall. Note that we can easily obtain accurate moving distance when the receiver moves at a distance of 2 meters as shown in Figure 8b. It's easy to satisfy in most IoT devices in smart city environments, since the IoT devices that required to interact with users are usually placed in a place that is easy to reach.

IX. CONCLUSIONS

This paper presents the system design of RAMTEL, which provides a fine-grained mobile interaction solution for commercial mobile devices. We propose a phase calibration method that can compensate the phase offset between devices accurately, and calculate the moving distance based on the phase change of acoustic signals. Based on Extreme Learning Machine, we propose a novel algorithm to mitigate multipath effects, and obtain the accurate phase change of LoS signals in multipath fading environments. In this way, we implement a prototype of RAMTEL using a commercial smartphone and some speakers connected with a desktop. The prototype could track the motion of the smartphone with mm-level accuracy, even if the smartphone moves slowly and slightly. We conduct systematic evaluation based on the prototype. Experiment results validated our idea as well as the system design.

REFERENCES

[1] Y. Liu, W. Zhang, Y. Yang, W. Fang, F. Qin, and X. Dai, "PAMT: phase-based acoustic motion tracking in multipath fading environments," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications (INFOCOM 2019)*, Paris, France, Apr. 2019.

[2] A.-V. Anttiroiko, P. Valkama, and S. J. Bailey, "Smart cities in the new service economy: building platforms for smart services," *AI & SOCIETY*, vol. 29, no. 3, pp. 323–334, Aug 2014. [Online]. Available: https://doi.org/10.1007/s00146-013-0464-0

[3] N. Komninos, M. Pallot, and H. Schaffers, "Special issue on smart cities and the future internet in europe," *Journal of the Knowledge Economy*, vol. 4, no. 2, pp. 119–134, Jun 2013. [Online]. Available: https://doi.org/10.1007/s13132-012-0083-x

[4] M. Kranz, P. Holleis, and A. Schmidt, "Embedded interaction: Interacting with the internet of things," *IEEE Internet Computing*, vol. 14, no. 2, pp. 46–53, March 2010.

[5] K. Joshi, S. Hong, and S. Katti, "Pinpoint: Localizing interfering radios," in *Proceedings of the 10th USENIX Conference on Networked Systems Design and Implementation*, ser. nsdi'13. Berkeley, CA, USA: USENIX Association, 2013, pp. 241–254. [Online]. Available: http://dl.acm.org/citation.cfm?id=2482626.2482651

[6] S. Zhu and X. Zhang, "Enabling high-precision visible light localization in today's buildings," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '17. New York, NY, USA: ACM, 2017, pp. 96–108. [Online]. Available: http://doi.acm.org/10.1145/3081333.3081335

[7] O. Abari, H. Hassanieh, M. Rodreguiz, and D. Katabi, "Poster: A millimeter wave software defined radio platform with phased arrays," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16. New York, NY, USA: ACM, 2016, pp. 419–420. [Online]. Available: http://doi.acm.org/10.1145/2973750.2985258

[8] W. Huang, Y. Xiong, X.-Y. Li, H. Lin, X. Mao, P. Yang, and Y. Liu, "Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*. IEEE, apr 2014.

[9] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '15. New York, NY, USA: ACM, 2015, pp. 15–29. [Online]. Available: http://doi.acm.org/10.1145/2742647.2742662

[10] K.-Y. Chen, D. Ashbrook, M. Goel, S.-H. Lee, and S. Patel, "Airlink: Sharing files between multiple devices using in-air gestures," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14. New York, NY, USA: ACM, 2014, pp. 565–569. [Online]. Available: http://doi.acm.org/10.1145/2632048.2632090

[11] W. Mao, J. He, and L. Qiu, "Cat: High-precision acoustic motion tracking," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16. New York, NY, USA: ACM, pp. 69–81.

[12] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun, "Indoor follow me drone," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '17. New York, NY, USA: ACM, 2017, pp. 345–358. [Online]. Available: http://doi.acm.org/10.1145/3081333.3081362

[13] R. Gierlich, J. Huttner, A. Dabek, and M. Huemer, "Performance analysis of fmcw synchronization techniques for indoor radiolocation," in *2007 European Conference on Wireless Technologies*, Oct 2007, pp. 24–27.

[14] C. Peng, G. Shen, and Y. Zhang, "Beepbeep: A high-accuracy acoustic-based system for ranging and localization using cots devices," *ACM Trans. Embed. Comput. Syst.*, vol. 11, no. 1, pp. 4:1–4:29, Apr. 2012. [Online]. Available: http://doi.acm.org/10.1145/2146417.2146421

[15] B. Yang, G. Mao, M. Ding, X. Ge, and X. Tao, "Dense small cell networks: From noise-limited to dense interference-limited," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4262–4277, May 2018.

[16] B. Yang, G. Mao, X. Ge, M. Ding, and X. Yang, "On the energy-efficient deployment for ultra-dense heterogeneous networks with nlos and los transmissions," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 369–384, June 2018.

[17] C. Zhang, Q. Xue, A. Waghmare, S. Jain, Y. Pu, S. Hersek, K. Lyons, K. A. Cunefare, O. T. Inan, and G. D. Abowd, "Soundtrak: Continuous 3d tracking of a finger using active acoustics," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 30:1–30:25, Jun. 2017. [Online]. Available: http://doi.acm.org/10.1145/3090095

[18] Y. Zhang, J. Wang, W. Wang, Z. Wang, and Y. Liu, "Vernier: Accurate and fast acoustic motion tracking using mobile devices," in *Proceedings of the IEEE International Conference on Computer Communications*, 2018.

[19] D. Li and Y. H. Hu, "Energy-based collaborative source localization using acoustic microsensor array," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 4, p. 985029, Mar 2003. [Online]. Available: https://doi.org/10.1155/S1110865703212075

[20] S. Chung and I. Rhee, "vtrack: Virtual trackpad interface using mm-level sound source localization for mobile interaction," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ser. UbiComp '16. New York, NY, USA: ACM, 2016, pp. 41–44. [Online]. Available: http://doi.acm.org/10.1145/2968219.2971450

[21] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '11. New York, NY, USA: ACM, 2011, pp. 155–168. [Online]. Available: http://doi.acm.org/10.1145/1999995.2000011

[22] A. Pigazo, V. M. Moreno, and E. J. Estbanez, "A recursive park transformation to improve the performance of synchronous reference frame controllers in shunt active power filters," *IEEE Trans. Power Electron.*, vol. 24, no. 9, pp. 2065–2075, Sept 2009.

[23] M. Speth, S. A. Fechtel, G. Fock, and H. Meyr, "Optimum receiver design for wireless broad-band systems using ofdm. i," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1668–1677, Nov 1999.

[24] X. Ge, X. Tian, Y. Qiu, G. Mao, and T. Han, "Small-cell networks with fractal coverage characteristics," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5457–5469, Nov 2018.

[25] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16. New York, NY, USA: ACM, 2016, pp. 82–94. [Online]. Available: http://doi.acm.org/10.1145/2973750.2973764

[26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489 – 501, 2006, neural Networks. [Online]. Available: https://doi.org/10.1016/j.sigpro.2014.06.031

[27] M. S. Hossain and G. Muhammad, "Audio-visual emotion recognition using multi-directional regression and ridgelet transform," *Journal on Multimodal User Interfaces*, vol. 10, no. 4, pp. 325–333, Dec 2016. [Online]. Available: https://doi.org/10.1007/s12193-015-0207-2

[28] G. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, April 2012.

[29] Y. Yang, K. Wang, G. Zhang, X. Chen, X. Luo, and M. Zhou, "Meets: Maximal energy efficient task scheduling in homogeneous fog networks," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4076–4087, Oct 2018.

[30] Y. Yang, "Multi-tier computing networks for intelligent IoT," *Nature Electronics*, vol. 2, no. 1, pp. 4–5, jan 2019.