

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/345004508>

AcuTe: Acoustic Thermometer Empowered by a Single Smartphone

Conference Paper · November 2020

DOI: 10.1145/3384419.3430714

CITATIONS

2

READS

32

6 authors, including:



Henglin Pu

University of Michigan

4 PUBLICATIONS 12 CITATIONS

[SEE PROFILE](#)



Jun Luo

Nanyang Technological University

142 PUBLICATIONS 5,795 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Contact-free Sensing Driven by Deep Learning [View project](#)



Localization and Tracking with Pervasive Sensing [View project](#)

AcuTe: Acoustic Thermometer Empowered by a Single Smartphone

Chao Cai¹, Zhe Chen¹, Henglin Pu², Liyuan Ye², Menglan Hu², and Jun Luo¹

¹School of Computer Science and Engineering, Nanyang Technological University, Singapore

²School of Electronic Information and Engineering, Huazhong University of Science and Technology, China

Email: ¹{chris.cai, chen.zhe, junluo}@ntu.edu.sg

ABSTRACT

Though measuring ambient temperature is often deemed as an easy job, collecting large-scale temperature readings in real-time is still a formidable task. The recent boom of network-ready (mobile) devices and the subsequent mobile crowdsourcing applications do offer an opportunity to accomplish this task, yet equipping commodity devices with ambient temperature sensing capability is highly non-trivial and hence has never been achieved. In this paper, we propose **Acoustic Thermometer (AcuTe)** as the first ambient temperature sensor empowered by a single commodity smartphone. AcuTe utilizes on-board dual microphones to estimate **air-borne** sound propagation speed, thereby deriving ambient temperature. To accurately estimate sound propagation speed, we **leverage the phase of chirp signals to circumvent the low sample rate on commodity hardware**. In addition, we propose to use **both structure-borne and air-borne propagations to address the multipath problem**. Furthermore, to prevent disruptive audible transmissions, we convert chirp signals into white noises and propose a pipeline of signal processing algorithms to denoise received samples. As a mobile, economical, highly accurate sensor, AcuTe may potentially enable many relevant applications, in particular large-scale indoor/outdoor temperature monitoring in real-time. We conduct extensive experiments on AcuTe; the results demonstrate a robust performance, a median accuracy **of 0.3°C** even at a varying humidity level, and the ability to conduct distributed temperature sensing in real-time.

CCS CONCEPTS

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

KEYWORDS

Acoustic sensing, mobile computing, phase, multipath problem.

ACM Reference Format:

C. Cai, Z. Chen, H. Pu, L. Ye, M. Hu, and J. Luo. 2020. AcuTe: Acoustic Thermometer Empowered by a Single Smartphone. In *The 18th ACM Conference on Embedded Networked Sensor Systems (SenSys '20)*, November 16–19, 2020, Virtual Event, Japan. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3384419.3430714>

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

SenSys '20, November 16–19, 2020, Virtual Event, Japan

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7590-0/20/11...\$15.00

<https://doi.org/10.1145/3384419.3430714>

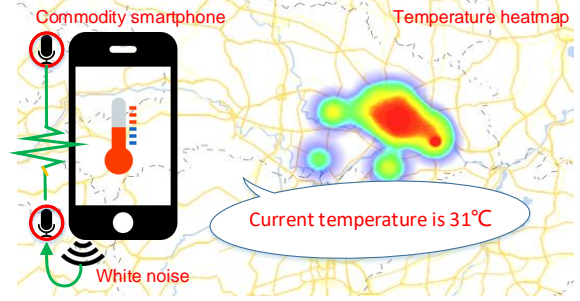


Figure 1: AcuTe turns a commodity smartphone into a thermometer, exploiting on-board acoustic sensors. It can support large-scale temperature monitoring, leveraging the pervasive availability of smartphones.

1 INTRODUCTION

Temperature sensing plays a pivotal role in a number of applications, including weather forecasting [9, 66], precision agriculture [37, 42], anomaly detection [11, 39], and building management [27, 64]. In particular, maintaining a satisfactory level of *thermal comfort* [22, 47] with minimal energy consumption has become increasingly important for smart buildings [4, 16, 25, 44], where the key enabling technology is distributed ambient temperature sensing in real-time [13]. Although temperature readings can be readily obtained via conventional thermometers, these devices are mostly not network-ready and ubiquitously available, impeding the deployment of real-time sensing at a very large scale. Ideally, distributed real-time temperature sensing could be enabled by the pervasively available mobile devices, but the on-board temperature sensors equipped by these devices cannot be used for this purpose, as they are interfered by self-generated heat, unless special hardware designs (e.g., heat isolation or proper device layout [1, 51]) are adopted or coarse inferences can be tolerated [25]. Since these stringent design requirements contradict the compactness of mobile devices, it is imperative to seek for an alternative solution.

In recent years, mobile devices including smartphones enjoy an explosive growth. In 2020, worldwide mobile users are projected to reach 4.78 billion, among which 3.5 billion are smartphones [3]. Driven by the fast development of mobile computing, smartphones now have rich on-board sensors and powerful processors, enabling ubiquitous computing and sensing [5, 26, 54, 65]. The richness in sensing capability has led to a booming trend in innovative utilization of on-board sensors beyond their primary use. These include, among others, using camera for visible light positioning [57] or communication [60, 63], inertial measurement unit for navigation [46],

WiFi signals for inertial sensing [56], vibration motor for acoustic sensing [43], and acoustic transducers for pressure detection [52]. Nevertheless, leveraging the on-board sensing abilities to enable ambient temperature sensing has never been realized.

In this paper, we design the first **Acoustic Thermometer (AcuTe)** with commodity devices using acoustic sensing. AcuTe empowers a single smartphone to perform high-accuracy ambient air temperature sensing. Basically, AcuTe **utilizes on-board dual microphones to estimate sound propagation speed, thereby deriving ambient air temperature readings according to a roughly linear relationship between temperature and sound speed.** To accurately estimate sound propagation speed, we leverage the phase of chirp signals, circumventing the low sampling rate on commodity hardware. In addition, we propose to use both structure-borne and air-borne propagation to address the multipath problem, improving system robustness. Furthermore, to prevent disruptive audible transmissions, we convert chirp signals into white noises and propose a signal denoising pipeline. AcuTe is portable, economical, and of high-accuracy, making it competitive with commercial thermometers built on dedicated hardware. Incorporating recent advances in mobile crowdsensing [21, 23, 24, 62], AcuTe has the potential to enable large-scale temperature sensing in real-time, thus revolutionizing current practice in distributed temperature acquisition.

However, implementing AcuTe on a commodity smartphone is very challenging. First of all, to obtain ambient temperature from sound speed, a micro-second timing resolution is required. Assuming the distance between two microphones is 0.2m, the sound speed change caused by a 1°C temperature variation only yields a $1.09\mu\text{s}$ difference in travelling time. To detect this minor difference, the sampling rate should be at least 914kHz, but the maximum sampling rate on most smartphones is 48kHz, only $\frac{1}{19}$ of the minimum requirement and thus far from being sufficient. Second, using two microphones to record self-transmitted signals would encounter multipath effect due to special acoustic transducer layout. The multipath problem, if not handled properly, could severely affect the sensing results. Last but not least, most audio signal transmissions, though deployed at common inaudible frequency range (normally above 18kHz [50]), could still be perceived by infants and animals and hence hinder their wide adoptions. Although phase whitening [53] can be a feasible approach, it is difficult to recover clean chirp signals due to severe channel distortion.

We design several strategies to address the above challenges. First, to circumvent the insufficient sampling rate of commodity smartphones, we utilize the phase of chirp signals. Because phase is a continuous variable and independent of the sampling rate, it could theoretically achieve an infinite resolution, thereby yielding highly accurate temperature resolution. Second, to deal with the multipath effect, we leverage air-borne and structure-borne propagation together with physical constraints imposed by microphone-speaker settings; this enables us to extract useful information from corrupted signals. Last but not least, we directly **apply phase whitening to randomize the chirp signals before transmissions**, in order to mitigate the disruptive audibility issue. However, we design a novel signal denoising pipeline to overcome the impact of channel distortions on chirp recovery. To summarize, this paper makes the following contributions:

- We design the first single smartphone based acoustic thermometer that achieves a sub-centigrade resolution.
- We propose to exploit the **phase of chirp signals** to achieve an accurate sound speed estimation, under a largely deficient sampling rate offered by smartphones.
- We design novel algorithms to deal with the audibility issue, signal denoising, and most importantly, multipath effect.
- We implement AcuTe on commodity smartphones [7] and conduct extensive experiments to showcase the robustness of AcuTe under various environmental conditions.
- We conduct both campus-scale and wide-scale temperature monitoring, demonstrating the promising potential of AcuTe to accomplish tasks not made possible before.

Combined with other on-board sensors such as GPS, AcuTe may enable many novel applications, such as acquiring in-building temperature distribution and even nation-wide temperature graph. The rest of our paper is organized as follows: Section 2 discusses background and literature. Section 3 presents system design in details. In Section 4, we report extensive experiment results. We finally conclude our paper in Section 5.

2 BACKGROUND AND RELATED WORK

We first discuss the literature on general temperature sensing in this section, then we also briefly study recent proposals on other sensing functions enabled by smartphones. Basically, there are two well-known approaches for temperature sensing, namely contact and non-contact methods [40]. Unfortunately, smartphones' contact sensing ability cannot support ambient temperature sensing, forcing us to resort to non-contact sensing.

Contact sensing methods, as its name refers, require a physical contact between a sensor and a test subject, which leads to quite a few limitations. First, before retrieving a temperature reading, contact sensing methods have to wait until a heat equilibrium between the sensor and subject is reached. Consequently, these methods always incur a long latency. Second, these methods can only obtain the temperature of a specific point where the sensor touches. Third, the physical contact may significantly affect the temperature of the subject due to heat dissipation. Finally, devices dedicated to contact sensing often require special hardware designs. As recommended by chip vendors [1, 51], obtaining ambient temperature accurately needs special hardware designs including proper device layout, heat isolation, etc. While the first three limitations are at odds with the requirement of ubiquitous sensing, the last limitation largely hinders the integration of temperature sensors into compact smartphones. Currently, only coarse temperature inference can be achieved by on-board temperature sensors of smartphones [25].

Non-contact approaches are born to overcome the shortcomings of contact methods. This line of work is far more responsive but is also more sensitive to environmental noises especially heat radiations [2] in proximity. One of the most widely known non-contact methods is the infrared-based [31]. It measures the temperature of a subject by gauging the nearby radiated heat from the subject. However, it fails to work for measuring ambient air temperature, as it requires a subject to radiate heat. Other non-contact sensing techniques based on, for example laser [29, 33] or radio frequency (RF) [18, 45], obtain temperature by a sophisticated analysis on

the reflected signal properties. They are common in remote atmospheric temperature sensing at high elevations but are too costly for daily use. In meteorology, sound has been exploited to estimate air temperature and wind speed [28, 48], given a roughly linear relationship between temperature and sound speed. Nonetheless, these solutions often require either high sampling rates or ultrasonic signals, making them infeasible to be deployed on commodity mobile devices. In contrast, ACuTe is the first acoustic-enable ambient temperature sensing on a single smartphone.

ACuTe is built upon recent developments in acoustic sensing [12, 34, 50, 52, 61] that have been used to enable localization, tracking, and gesture recognition. For instance, Wang et al. [49, 50] exploit the phase of acoustic reflections to perform high-precision gesture tracking and recognition. Mao et al. [34, 36] borrow the techniques from RF radar and utilize its acoustic implementation to achieve sub-centimeter level tracking. ForcePhone [52] models a smartphone as a forced and damped mass-spring system, where the relation between the applied force and the reduced vibration amplitude can be analytically represented, thus enabling touch pressure detection without any dedicated sensors. Acoustic sensing has been further extended to respiration rate monitoring [38, 53, 61] and imaging [35] on commodity smartphones. In fact, there is a booming trend for innovative utilization of on-board sensors beyond their primary use, such as using camera for visual light communication [63], vibration motor for acoustic sensing [43], and acoustic transducers for electromagnetic sensing [32]. These exciting developments all serve as our inspirations.

3 ACuTe

The core of ACuTe is to obtain highly accurate sound speed, thereby deriving temperature. In this section, we first outline the idea and challenges behind ACuTe. Then we present our approach for sound speed estimation, and we also address two technical challenges, namely **multipath effect and audible noise**. We finally discuss **parameter selection** at the end.

3.1 Idea and Challenges

ACuTe enables a single smartphone to measure ambient air temperature, relying on a roughly linear relationship between temperature and sound speed. Essentially, ACuTe tracks the sound propagation time from a smartphone speaker to its two microphones so as to obtain sound propagation speed, thereby deriving temperature. As shown in Figure 2, ACuTe **mixes the chirp signals** transmitted from its speaker and received by the two microphone channels. It then performs a **DFT analysis** and obtains the acoustic propagation delay via a multipath handling pipeline. Incorporating this delay with a calibrated Speaker-to-Two-Microphone Distance Difference (STMDD), ACuTe deduces the sound speed, and in turn estimates the temperature by the following relation [28]:

$$c^2 = \alpha T_{\text{emp}} (1 + \beta h), \quad (1)$$

where T_{emp} denotes the **temperature in Kelvin**, c represents the sound speed (m/s) in air, h is the ratio between **vapor pressure of water in air and absolute atmospheric pressure** (affected by humidity), α and β are two constants. Theoretically, the term h in Equation (1) can be eliminated only in dry air, but earlier proposal [48] and our

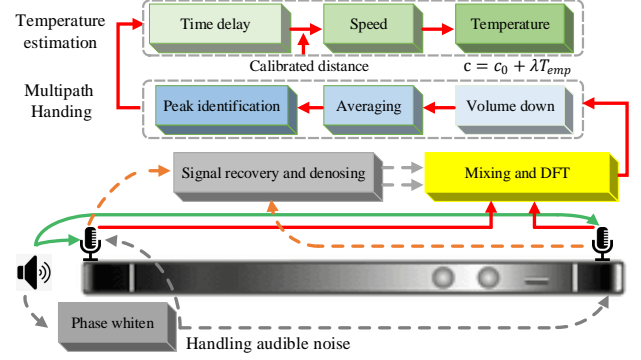


Figure 2: ACuTe system architecture.

empirical results reveal that this term introduces only a negligible impact on the final results, hence we set it to 0. This allows us to further simplify Equation (1) as:

$$c^2 = \alpha T_{\text{emp}}. \quad (2)$$

We can further simplify Equation (2) via Taylor expansion but omit high order terms to obtain

$$c = c_0 + \lambda T_{\text{emp}}, \quad (3)$$

where $c_0 = 331$ and $\lambda = 0.606$ are two constants [6].

According to Equation (3), temperature can be obtained if sound propagation speed c is known, which however, is non-trivial to obtain. First, as we mentioned earlier, commodity smartphones have insufficient sampling rate, making it challenging to achieve high accurate speed estimation. Second, multipath problem may seriously affect the accurate detection of signal arrival and thus hurt speed estimation accuracy. Finally, the audible noise during measurements may prevent ACuTe from practical adoptions. We address these challenges in the following.

3.2 Sound Speed Estimation

ACuTe utilizes the propagation time of acoustic signals from a smartphone speaker to its two microphones to estimate sound speed. Ideally, the propagation time to either microphone should suffice, but **uncertain system delays** would “pollute” this measurement. We take the advantage of common smartphone layout arranging the two microphones at different distances from the speaker (the STMDD feature explained earlier).¹ Basically, the two microphones on a smartphone share the same clock and are hence in-phase. Therefore, samples from one channel can be deemed as a delayed version of another; the induced time delay between them is free of uncertain delays and can be readily obtained given a sufficient sampling rate. As STMDD remains constant for a given smartphone, any sound speed variations caused by temperature changes can be reflected in this time delay between the two channels.

Let us consider the following transmitted chirp signal:

$$s(t) = e^{-j(2\pi f_{\text{min}} t + \pi \frac{B}{T} t^2)}, 0 < t < T, \quad (4)$$

where, f_{min} is the initial frequency, B denotes the bandwidth, and T represents the duration (given a zero initial phase). Let $s^s(t) =$

¹Normally, the speaker and one microphone are at the bottom of a smartphone, while another microphone is at the top.

$e^{-j(2\pi f_{\min}t + \pi \frac{B}{T} t^2)}$ denote the chirp signal from one (short) channel, then that of the other (long) channel delayed by Δt become $s^l(t) = e^{-j(2\pi f_{\min}(t-\Delta t) + \pi \frac{B}{T} (t-\Delta t)^2)}$. To calculate the time delay Δt , using cross-correlation is a common method but that would require an extremely high sampling rate f_s to achieve an adequate time resolution. Consider an STMDD of $d = 0.2\text{m}$, in order to detect the change in sound speed caused by a 1°C temperature variation, we would need $f_s > \frac{1}{\Delta t} \approx \frac{1}{\frac{d}{c_0 - c_1}} = 914\text{kHz}$, where $c_1 = c_0 + 0.606\text{m/s}$

according to Equation (3). In reality, the maximum sampling rate on commodity mobile devices is often around 48kHz , only $\frac{1}{19}$ of the minimum requirement and thus far from being sufficient.

To circumvent this insufficiency in sampling rate, we leverage phase information. We first mix (i.e., multiply) the signals from two channels to obtain:

$$s_{\text{mix}}(t) = e^{-j(2\pi \Delta t \frac{B}{T} t + 2\pi f_{\min} \Delta t - \pi \frac{B}{T} \Delta t^2)} + (\text{high freq. terms}), \quad (5)$$

then we remove the high frequency terms by a low pass filter, and we take the derivative of the phase component with respect to t in Equation (5), obtaining the remaining frequency component f_d as:

$$f_d = B\Delta t/T. \quad (6)$$

Now that f_d and Δt has a linear relationship as B and T are deemed as two constants for now, so we can get Δt if f_d is known, which could be achieved by a DFT analysis. To obtain finer results, we interpolate the filtered signals with a factor of 128, before performing DFT analysis on s_{mix} . Ideally, we can inspect a peak in the spectrum, whose frequency f_p and the phase $\angle f_p$ are used to estimate f_d :

$$f_d = f_p + \angle f_p / 2\pi. \quad (7)$$

With the obtained Δt , the sound speed c can be calculated as $c = \frac{d}{\Delta t}$, where d is the STMDD. The speed c finally allows us to derive the temperature according to Equation (2). Unfortunately, the multipath effect could complicate the mixing results, leading to multiple peaks in the spectrum and preventing us from correctly identifying the useful peak that indicates the time delay Δt .

3.3 Handling Multipath Effect

The multipath effect may jeopardize the identification of the correct peak frequency for sound speed estimation, hence resulting in inaccurate temperature readings. In a multipath channel, the received signal is a summation of the transmitted signal and its several attenuated and delayed copies. Therefore, we can characterize a multipath profile of one microphone channel as:

$$\sum_{i=1}^l \rho_i e^{-j(2\pi f_{\min}(t-\tau_i) + \pi \frac{B}{T} (t-\tau_i)^2)}, \quad (8)$$

where l is the number of multipath reverberations, ρ_i and τ_i are the path loss and time delay of the i -th path, respectively. Assuming there are m multipath reverberations in another channel (represented in the same way), so if we mix the signals of two channels, we can get up to $l \times m$ peaks in the DFT spectrum. In addition, there may exist several coherent multipath reverberations that add up constructively and result in even higher energy than the useful peak from direct Line-of-Sight (LoS) transmissions. Consequently, the multipath problem is much more severe than common systems that only use a single microphone channel. To mitigate the multipath problem, we take the following three steps.

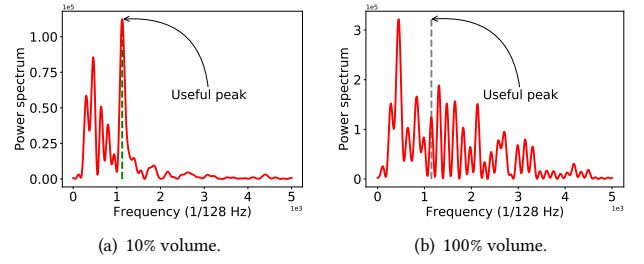


Figure 3: Tuning down the transmission volume is effective in mitigating the multipath problem. The mixed signal spectrum has its number of peaks under 10% transmission volume (a) far less than that under 100% volume (b). Also, the useful peak for temperature estimation becomes dominant when the volume is tuned down. These measurements are taken on a HUAWEI MATE9 Pro.

3.3.1 Volume Down. Since acoustic sensors in our system are in close proximity (usually $< 25\text{cm}$), we thus tune down the volume during signal transmissions, leading to less multipath reverberations. Figure 3 demonstrates that tuning down the transmission volume is effective in mitigating the multipath problem, as the number of peaks in the spectrum is greatly reduced when the volume is turned down. As this could only reduce multipath reverberations bouncing off distant objects but could not handle nearby interference, more processing techniques are needed.

3.3.2 Averaging. We sum several adjacent mixed spectrum and take its average, thus attenuating instable and weak multipath distortions. During the measurement, we can slightly rotate or move the smartphone so as to randomize the multipath reverberations and corrupt their coherence, further improving averaging results. The underlying reason is that direct Line-of-Sight (LOS) signal is intuitively more stable and powerful than multipath reverberations experiencing severer path loss. Therefore, if we average several frames, the peak by LOS signal remains unchanged, while others by the multipath reverberations would be significantly attenuated as their locations in the spectrum vary from time to time.

To help understand the above intuition, suppose at certain frame, there are w candidate peaks whose magnitude response are a_i , $i \in [1, w]$, each located at frequency bin f_i^w in the mixed spectrum. In the next frame, some of the peaks a_i may almost remain the same, especially the one generated by the LOS signal while others are highly probable to change their locations f_i^w . So when summing several adjacent spectra, the peak a_k introduced by LOS signal become $\frac{\sum_{i=1}^{w-w} a_k}{w} = a_k$, devoid of any attenuations. On the contrary, other peaks, say a_q incurred by a multipath reverberation, after summation becomes $\frac{\sum_{i=1}^{w-w} a_q \{f_i^w = f_q^w\}}{w} = \frac{a_q}{w}$, apparently w times attenuated and thus mitigating the multipath effect. Figure 4 depicts the effects of averaging. It is observable that averaging could effectively mitigate multipath reverberations as the number of peaks in the spectrum is largely reduced. However, there may still exist several peaks with a comparable strength as the LOS one, and they are likely to be very close to the LOS one in the spectrum. As a result, one last but the most important processing step is still needed.

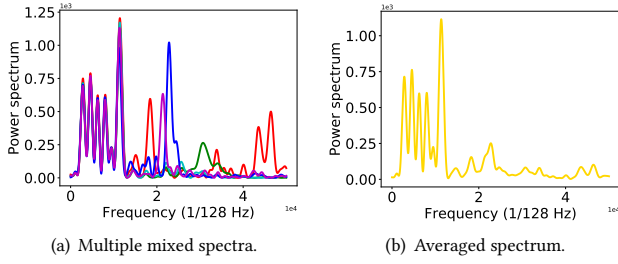


Figure 4: Averaging could effectively attenuate instable multipath reverberations. The multiple mixed spectra without averaging in (a) show many instable peaks above 150 Hz. In contrast, after averaging as shown in (b), the spectrum is significantly cleaner. These measurements are taken on HUAWEI MATE9 Pro at a corner.

3.3.3 Peak Identification. We now leverage both structure-borne and air-borne sound propagation to pinpoint the authentic peak frequency generated by LOS signals for temperature estimation. To start with, we first build a signal propagation model. As shown in Figure 5, there are two types of paths for signals to travel from bottom speaker to two microphones directly. One type goes through the device body, known as structure-borne propagation, where the travel speed could reach over 2000m/s [52]. The other type is the air channel, a more common path known as air-borne propagation, where the travel speed is temperature-dependent and is around 340m/s under a temperature of 20°C. The structure-borne signals are often orders weaker than air-borne signals, but may still be perceivable [50, 52].

Based on the above propagation model, the multipath profiles of the two microphone channels (short and long) can be extended as:

$$\begin{aligned}
 s^s(t) &= \rho_s^s s(t + \Delta t_1) + \rho_a^s s(t + \Delta t_4) + \sum_{i=1}^{l-2} \rho_i^s s(t + \Delta t_i^s), \\
 s^l(t) &= \rho_s^l s(t + \Delta t_2) + \rho_a^l s(t + \Delta t_3) + \sum_{i=1}^{l-2} \rho_i^l s(t + \Delta t_i^l),
 \end{aligned} \quad (9)$$

where $s(t)$ is the transmitted signal shown in Equation (4). $\rho_s^s, \rho_a^s, \rho_i^s$ are the path loss for structure-borne propagation, air-borne propagation, and the i -th multipath air-borne propagation of the short channel, $\rho_s^l, \rho_a^l, \rho_i^l$ are those for the long channel, and $\Delta t_i^s, \Delta t_i^l$ are the propagation delays of the i -th path. Since the travel speed of structure-borne sound is faster, the time spent for signals to travel from the speaker to a microphone can be negligible hence

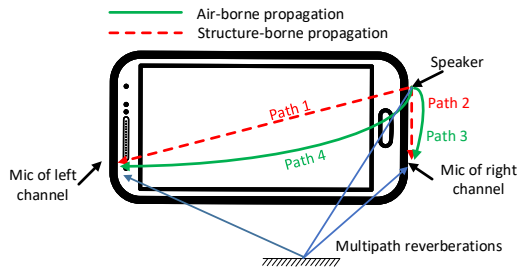


Figure 5: Signal propagation model.

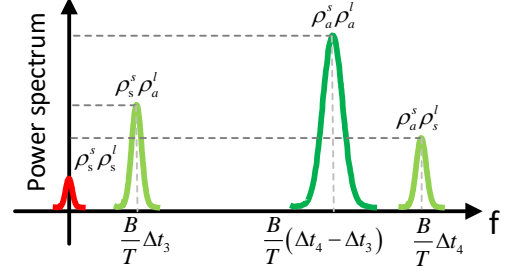


Figure 6: Power spectrum without multipath.

$\Delta t_1 = \Delta t_2 \approx 0$. If we mix $s^s(t)$ and $s^l(t)$, the first two terms in $s^s(t)$ and $s^l(t)$ would generate four peaks located at frequency bins: $\frac{B}{T}|\Delta t_1 - \Delta t_2| = 0$, $\frac{B}{T}|\Delta t_1 - \Delta t_3| = \frac{B}{T}\Delta t_3$, $\frac{B}{T}|\Delta t_2 - \Delta t_4| = \frac{B}{T}\Delta t_4$, and $\frac{B}{T}|\Delta t_3 - \Delta t_4| = \frac{B}{T}\Delta t$. The peak located at $\frac{B}{T}\Delta t$ is the one used for temperature sensing, while other peaks separated from it at fixed distances can in turn be used to correctly identify the useful peak. The mixed frequency response without multipath is shown in Figure 6. The multipath signals (plentiful but not plotted in Figure 6) would randomly locate at any frequency bins, and their magnitude could also vary drastically.

The peaks, incurred by path 1 to path 4 depicted in Figure 5, would always show up in the mixed spectrum regardless of the device position. Their magnitude may exhibit variations due to channel dynamics but their relatively positions in the spectrum are fixed hence can be used to identify the peak at $\frac{B}{T}\Delta t$ for temperature sensing. The two peaks located at DC and $\frac{B}{T}\Delta t_3$ are prone to low frequency noise and are hence not utilized. Both the magnitudes and the relative positions of remaining peaks located at $\frac{B}{T}\Delta t$ and $\frac{B}{T}\Delta t_4$ are used to resolve the multipath effect. For the convenience of description, we denote the peak at $\frac{B}{T}\Delta t$ as *main peak* and the other one that helps identify the main peak as *brother peak*. Figure 7 shows some real measurements from different smartphones.

Now that the problem is converted into finding the main and brother peaks that are $\frac{B}{T}\Delta t_3$ apart from each other in the mixed spectrum. To solve the problem, we adopt a **maximum a posterior probability (MAP) estimator** to exploit the fixed “distance” $\frac{B}{T}\Delta t_3$. In particular, the estimation takes the following steps. **First, we pick up several candidate peaks whose magnitudes are above 50% of the maximum.** Suppose there are w peaks, only $a_j^m > \frac{1}{2} \max(a_i^m)$, $i, j \in [1, w]$ are chosen. **Second, for each main peak candidate, say the j -th candidate a_j^m located at f_j^m , we choose another local maximum peak, say a_j^b at f_j^b close to $\frac{B}{T}\Delta t_3 + f_j^m$, as the brother peak.** We model the probability of a_j^m being the correct main peak by the following equation:

$$p_j = \frac{(a_j^m / a_j^b) e^{-\frac{(f_j^b - f_j^m - f_\Delta)^2}{2\sigma^2}}}{\sum_{i=1}^{i=w} (a_i^m / a_i^b) e^{-\frac{(f_i^b - f_i^m - f_\Delta)^2}{2\sigma^2}}}, \quad (10)$$

where $f_\Delta = \frac{B}{T}\Delta t_3$ and σ is a standard deviation. We empirically set σ to a small value, so as to penalize the probability if the frequency difference deviates too much from f_Δ . In our system, $\sigma = 10 \frac{f_s}{N}$, where

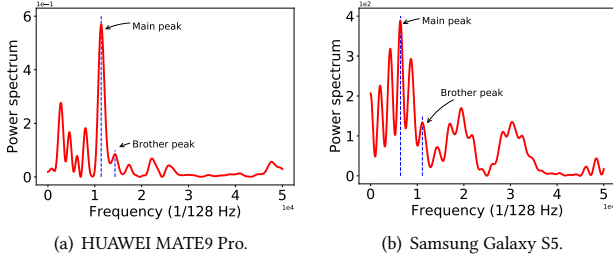


Figure 7: Mixed power spectrum of real measurements.

N is the number of points for DFT analysis. The ratio of $\frac{a_j^m}{a_j^l} = \frac{\rho_q^l}{\rho_s^l}$ rewards the probability if the main peak is higher than the brother peak, because received signals via structure-borne propagation is significantly weaker than those via air-borne propagation. Finally, after computing the probability for all the main peak candidates, we choose the one with the maximum probability for temperature sensing. The algorithm for the above process is shown in Algorithm 1. Figure 7 shows the obtained main peaks and brother peaks from different smartphones. As is shown in the figure, it is relatively straightforward to identify the main and brother peaks in the mixed spectrum from HUAWEI MATE9 Pro, whereas there are many comparable sidelobes in that from Samsung Galaxy S5. Fortunately, the proposed algorithm can always correctly pinpoint them. After the above process, we use both the frequency and phase of the main peak to get f_d , further deriving $c = \frac{dB}{f_d T}$ based on Equation (6).

Algorithm 1: Multipath handling algorithm.

Data: Mixed power spectrum s .

Result: Main peak a_q located at f_q .

- 1 Find several local peaks around $\frac{dB}{cT}$;
 - 2 Retrieve local maximum peaks $a_j^m, j \in [1, w]$ located at f_j^m that are above $\frac{1}{2} \max(a_j^m)$;
 - 3 $\Delta f = \frac{B}{T} t_3$;
 - 4 **while** $i < w$ **do**
 - 5 Find a brother peak for a_i^b whose frequency satisfies $f_i^b = \arg \min_{f_i^b} |f_i^b - f_i^m - \Delta f|, i = i + 1$;
 - 6 **while** $j < w$ **do**
 - 7 Calculate the probability p_j for the j -th peak by Equation (10), $j = j + 1$;
 - 8 $a_q = \arg \min_{f_i^m, a_i^m} p_i$
-

3.4 Handling Audible Noise

To handle the audible noise during the measurement process, we adopt a phase whiten technique [53] but propose an effective denoising pipeline for signal recovery. Note that this noise handling process is not compulsory if a user can tolerate the audible noise.

3.4.1 White Noise Transmission. In this section, we present details on how to transform chirp signals into white noises, thus reducing discomfort noises. As we have stated before, chirp transmissions,

even deployed in inaudible frequency range (above 18kHz) [50], could still be perceived by infants and animals. To resolve this issue, we randomize the phase of chirp signals [53], making them sound virtually equivalent to white noise (but at a very low volume) and thus much less perceivable.

Specifically, this phase randomization process is performed in frequency domain, and detailed steps are as follows. We perform DFT on time-domain chirp signal and get its frequency representation by $\sum_{f_i=0}^{f_i=f_s T} e^{-j(2\pi f_i + \theta_i)}$, where f_i is the i -th frequency bin in the spectrum, θ_i is its original phase, f_s denotes the sampling rate, and $f_s T$ represents the number of frequency bins (or the number of points in a DFT analysis). We remove the original phase θ_i and add a random phase $\phi_i, i \in [1, f_s T]$ to each frequency bin $f_i, i \in [1, f_s T]$, resulting in $\sum_{f_i=0}^{f_i=f_s T} e^{-j(2\pi f_i + \phi_i)}$. The $[\theta_i, \phi_i]_i$ pairs are kept in the system for signal recovery. Each phase ϕ_i is produced by a Mersenne Twister pseudo-random generator and is distributed within $[0, \pi]$. Taking the results from above process, we perform IDFT and convert the signal back to time-domain, which is now ready for transmission. Note that pseudo-random phases added to adjacent chirp transmissions are different so that the delayed reflections of preceding chirps impose less impacts on the subsequent ones. The results of the above process on the original chirp signal are shown in Figure 8.

3.4.2 Signal Recovery and Denoising. Given the known $[\theta_i, \phi_i]_i$ pairs, it might seem straightforward to recover the original signal, as, for each microphone channel, $\theta_i - \phi_i$ can be added to each frequency bin of the DFT results of received signals, reverting white noises back to chirp signals. Unfortunately, due to severe channel loss, the recovered chirp signals come with significant noises, as illustrated by the spectrogram $\mathcal{F}(t, f)$ shown in Figure 9(b). Although we can visually identify a linear time-frequency feature in the spectrogram, recovering the original signal is still very challenging as the noise bandwidth exactly overlaps the signal bandwidth. Moreover, the energy of noise, of comparable strength to that of the desired signals, spreads equally within the entire bandwidth, preventing us from applying advanced filters such as wavelet.

To this end, we use Smoothed Pseudo Wigner–Ville Distribution (SPWVD) [14, 41] to compute its major term, and the result are shown in Figure 9(c). The SPWVD results help preserve the principle components in $\mathcal{F}(t, f)$ and thus filter out noises, making the original chirp signals more identifiable. Specifically, our denoising algorithm proceeds as follows. Assuming r_n is the received

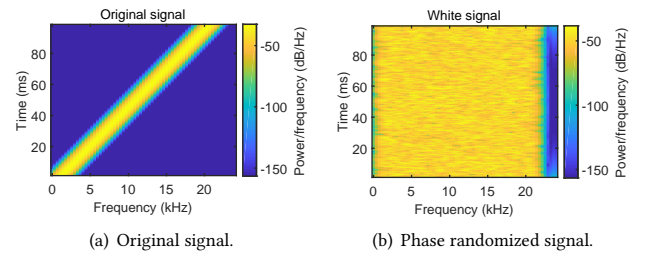


Figure 8: Resolving the audibility issue by randomizing chirp phases, turning audible chirps into white noise. We represent both of them in time-frequency spectrograms.

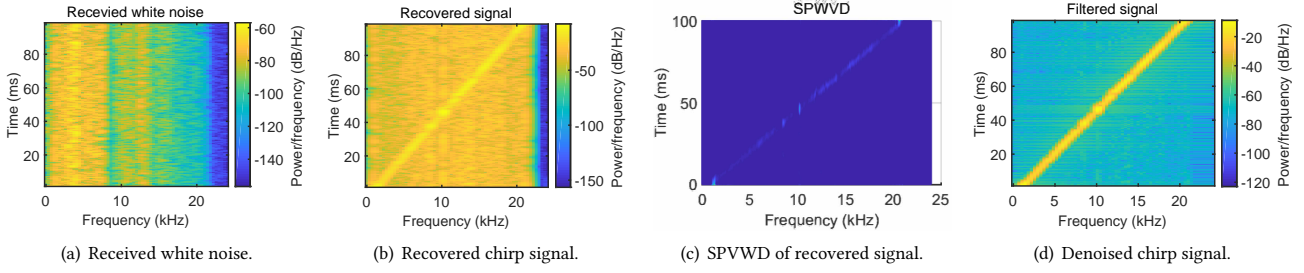


Figure 9: Process of reverting white noise back to chirp signal and denoising. (a) Time-frequency spectrogram of received signals. (b) Time-frequency spectrogram of recovered chirp signals, which shows significant noise. (c) Smoothed pseudo Wigner–Ville distribution of noisy chirp signals. (d) Denoised chirp signals.

noisy signal, we first compute its Short Time Fourier Transform (STFT), obtaining the time-frequency spectrogram $\mathcal{F}(t, f)$. Then we calculate SPWVD of r_n , resulting in its principle components in the spectrogram. Taking the SPWVD results, we generate a mask matrix M . The mask matrix M can be deemed as a binary time-frequency spectrogram where its principle components are assigned for “1” and other regions are “0”. To prevent possible loss due to non-flat frequency response, we interpolate M using RANSAC fitting algorithms, getting a more smoothed mask matrix M' . Following that, we calculate the Inverse Short Time Fourier Transform (ISTFT) of $M' \times \mathcal{F}(t, f)$ and getting a denoised chirp signal, $r = \mathcal{F}^{-1}(M' \times \mathcal{F}(t, f))$, as shown in Figure 9(d), which is significant cleaner. The pseudocodes of the signal recovery and denoising algorithm is presented in Algorithm 2. After denoising, the signals are ready for temperature estimation.

Algorithm 2: Signal recovery and denoising.

Data: Received noisy signal r_n .

Result: Denoised signal r .

- 1 Perform FFT on r_n and compensate the non-flat frequency response by adding a weight w_f to each frequency bin[36]; Perform IFFT to time domain and computing its STFT $\mathcal{F}(t, f)$;
 - 2 Calculate SPWVD to get the principal components;
 - 3 Covert SPWVD into binary values to get a mask matrix M ;
 - 4 Smooth M via RANSAC fitting, obtaining a better mask matrix M' ;
 - 5 Get denoised signal by performing ISTFT on $M' \times \mathcal{F}(t, f)$;
 - 6 $r = \mathcal{F}^{-1}(M' \times \mathcal{F}(t, f))$
-

3.5 Parameter Selection

In order to choose optimal parameters for temperature sensing, we plug Equation (6) into Equation (3) to get $T_{\text{emp}} = \frac{Bd}{\lambda f_d T} - \frac{c_0}{\lambda}$. We further approximate the above equation using Taylor expansion:

$$T_{\text{emp}} = \frac{k}{-\left(1 - \frac{f_d}{B}\right) + 1} - \frac{c_0}{\lambda} \approx k + k \left(1 - \frac{f_d}{B}\right) - \frac{c_0}{\lambda} + o\left(1 - \frac{f_d}{B}\right)^2$$

$$\approx -\frac{d}{\lambda BT} f_d + 2\frac{d}{\lambda T} - \frac{c_0}{\lambda} + o\left(1 - \frac{f_d}{B}\right)^2, \quad (11)$$

where $k = \frac{d}{\lambda T}$. The above equation shows that, f_d and T_{emp} have an **approximately negative linear relationship**. Therefore, decreasing the coefficient $\frac{d}{\lambda BT}$ could improve accuracy. Since d and λ cannot be changed, **the only choice is to increase the product $B \times T$** , which characterizes the *processing gain*. The higher the processing gain, the better the *signal-to-noise ratio* (SNR). This implies that B and T should be set at their maximum configurable values. However, if the duration T of a signal becomes too long, the received signal is more prone to multipath effect. In addition, as we have mentioned in **Algorithm 1**, the separation between the main peak and brother peak is $\frac{B}{T} \Delta t_3$, so a smaller T could improve the resolution in differentiating multiple reverberations. Finally, though the available bandwidth is 24kHz, signals at a frequency below 100Hz or above 22kHz suffer from severe attenuation and should be avoided.

The above analysis leads to a choice of $B = 20$ kHz, ranging from 1 kHz to 21 kHz, and $T = 0.1$ s, which strikes a good balance. Given the above settings, the frequency of the main peak should be $\frac{dB}{cT} \approx 82.3$ Hz, assuming $c \approx 340$ m/s and $d \approx 0.15$ m/s. Since we have interpolated the samples by a factor of 128, the actual location of the main peak would be $128 \times 82.3 = 10534.4$. Considering that the temperature could widely change within a range from -30°C to 40°C [10], we can further narrow down the scope to search the actual main peak. Given Equation (6), we can empirically known that, 1 Hz frequency change leads to

$\Delta T = \left| \frac{20000 \times 0.15}{0.6 \times 0.1 \times 82.3} - \frac{20000 \times 0.15}{0.6 \times 0.1 \times (82.3 - \frac{1}{128})} \right| \approx 0.0576^\circ\text{C}$. This minor change, indicating measurement resolution, could be further improved by involving phase hence accurate enough for most daily use. Moreover, such a setting yields a higher resolution to differentiate different signal components compared with common methods. For instance, the distance between a main peak and a brother peak is around $\frac{20000}{0.1} \times \frac{0.04}{340} \approx 3012$ (or 0.04m, roughly the distance from the speaker to the bottom microphone) on HUAWEI MATE9 Pro, far enough to recognize them.

For long term temperature monitoring, we apply **Kalman filter to smooth the estimation results**. Figure 10 depicts the obtained frequency f_d with and without Kalman filter. A peak 0.13Hz frequency change, equivalent to approximately 0.96°C temperate deviation, is observed without Kalman filtering. Otherwise the frequency change narrows down to about 0.06Hz, equivalent to 0.44°C temperature change that is closer to ground truth of only 0.3°C variation.

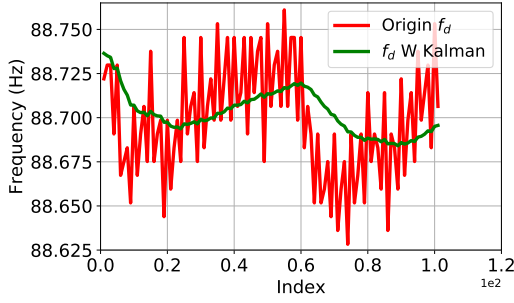


Figure 10: Kalman filtering results.

4 PERFORMANCE EVALUATION

We have implemented a AcuTe prototype on popular Android platform and evaluated the performance in representative environments. The minimum required OS version is Android 4.4 KitKat, and no root or NDK APIs are required. The prototype can operate in real-time on most smartphones; it leaves negligible energy footprint hence produce little self-generated heat. The smartphones we used include Samsung Galaxy S5, HUAWEI MATE9 Pro, One Plus 5T, etc. The bandwidth B for the chirp signal is 20kHz, ranging from 1kHz to 21kHz, with a duration of $T = 0.1$ s. During the experiment, the speaker volume is set to only 5% of the maximum. The safe temperature range to let smartphones (hence AcuTe) to function properly is 5 to 38°C, which is important to avoid problems such as sudden shutoff [25]. The detailed settings are shown in Table 1. To measure ground truth temperature, we use a dedicated sensor named GY-213V-SI7021 with an accuracy of 0.05°C in measuring temperature and an accuracy of 2% in relative humidity.

4.1 Micro Benchmark

In this section, we present micro benchmark evaluations including audibility issues and model verification.

4.1.1 Audibility Issue. We hereby present field tests to check whether AcuTe incurs perceivable chirp sound during its operation. We operate AcuTe on several smartphones and ask a group of people whether they can recognize the “chirp sound” during white noise transmissions. This group has 35 person aged from 10 to 56. To prevent any hearing impairment, the experiments are conducted under safe control. The maximum volume is kept under 50% and the minimum distance between experimenters and smartphones are over 0.5 m. The experiments are conducted in a very quiet place and the SoundMeter [20], an application to measure sound intensity, reports that the background noise is around 35dB, with a maximum of 48dB. We first train the experiment participants to differentiate chirp sound and pure white noise. After that, we use

Table 1: Settings for the transmitted signals

Parameters	Value
Sampling rate	48 kHz
Chirp bandwidth	20 kHz
Chirp duration	0.1 s
Initial frequency	1 kHz
Volume	5%

different phones to run AcuTe and hence to trigger white noise transmissions generated by randomizing the phase of chirp signals. As expected, no one report that they can perceive any chirp sound.

4.1.2 Model Verification. Now we present experiment results to verify our estimation model. The experiments are conducted in an indoor room, and the setup is shown in Figure 11. The electric heater in the figure is used to warm up air, and we move the distance between a smartphone to the heater so as to equivalently adjusting ambient air temperature. We put the dedicated temperature sensor beside a smartphone and probe its readings via Arduino UNO. We connect the UNO with a laptop that are synchronized with the smartphone via local WLAN. Upon each white noise transmission, the smartphone uploads its local timestamps to the laptop, which in turn triggers one temperature and humidity sampling. During the experiments, the relative humidity varies within a limited range from 47.6% to 50.2%.

We first record a set of frequencies of main peaks and their corresponding temperatures and then compare the measured frequency-temperature relationship with theoretical one explained in Equation 2. Figure 12(a) and (b) show the results. It is observable that the frequency of main peak changes at different temperature points as shown in Figure 12(a). And Figure 12(b) reveals that the frequency-temperature relationship indicated by measured data are highly consistent with theoretical basis, demonstrating the effectiveness of our proposed model. Figure 12(a) also showcases that the measured frequency of main peak is quite stable. A maximum deviation of only 0.08Hz is observed when the temperature is 24.48°C, which is equivalent to less than 0.05°C temperature variations, showcasing the robustness of our system.

4.2 Impact Analysis

In this section, we explore the impacts of various factors, including humidity, different parameter settings, device diversity, and background noise, on the sensing performance.

4.2.1 Humidity Impact. We hereby present statistical results to demonstrate that humidity poses negligible impacts on temperature sensing results. In this experiment, we adopt the same setup as that shown in Figure 11. We first trigger a smartphone for constant temperature monitoring, then we apply a water spray to release vapor towards the smartphone. The water has been heated so as to minimize the impacts of vapor on temperature. Figure 12(c) depicts the measured frequency of the main peak, as well as estimated and ground truth temperatures during this drastic humidity change. Note that all the measurements are recorded without filtering.

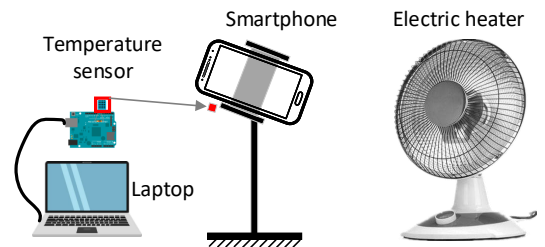


Figure 11: Experiment setup.

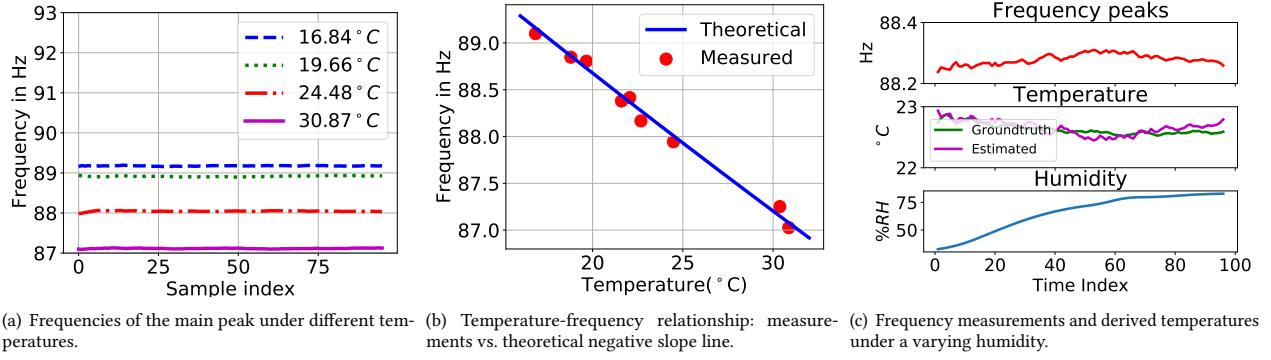


Figure 12: Benchmark results: the proposed model in Equation (2) are verified in (a) and (b), while the negligible impact on temperature sensing imposed by humidity is demonstrated in (c).

From Figure 12(c) we can see that, the relative humidity varies from a minimal of 64.5% to a maximum of 85.7%. During this period, the temperature has a maximum variation of 0.19°C, ranging from 24.62°C to 24.81°C. This 0.19°C temperature variation can also appear even when there is no such humidity changes. To further quantify the temperature and humidity relationship, we use Pearson Product-Moment Correlation Coefficient (PPMCC), a well-known statistical approach to quantify the correlation between frequencies of main peak and humidity. The PPMCC [17, 19] is defined by

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (12)$$

where μ_X , μ_Y and σ_X , σ_Y are mean and variance for data X , Y , respectively, and $\text{cov}()$ denotes the covariance. The PPMCC is within $[-1, 1]$ and if the absolute PPMCC is close to zero, X and Y are supposed to have loose correlation, otherwise strong correlation. Our measurements reveal $\rho_{X,Y} = -0.03$, which implies a rather loose correlation. This suggests that temperature variations caused by humidity are under noise floor and impacts from humidity should be less than 0.19°C as part of variations should be resulted from inherent noise. Meanwhile, this minor 0.19°C temperature change cannot impose any significant impacts on daily life. Consequently, it is reasonable to conclude that humidity impose negligible impact on temperature sensing.

4.2.2 Different Volumes. Here we study the impact of sound volume on the sensing results. Volume can impact the sensing results in two aspects. First, as we mentioned earlier, the volume, determining the signal intensity of transmitted signals, can affect the multipath profiles of the two microphone channels, thus resulting in different mixed spectrum. Specifically, large volume can lead to more severe multipath reverberations, which may occasionally add to the main peak either destructively or constructively. In either case, it can cause the frequency of the main peak shift hence causing instable temperature readings. Second, large volume can cause large displacement of speaker diaphragm, causing the calibrated STMDD d to vary over time, thereby instable temperature readings again. To understand the later aspect, we here briefly introduce the physics behind speakers.

Figure 13 depicts the structure of a speaker. It consists of several major components including magnet, voice coil, pole piece, and

speaker diaphragm, etc. The electric acoustic signals, acting as alternate current, drive the voice coil move back and forth along the pole piece. This in turn causes the speaker diaphragm to collapse and expand, generating acoustic waves. The intensity of generated acoustic waves is proportional to the displacement of speaker diaphragm. Therefore, the distance d_1 , d_2 for acoustic waves to propagate from speaker to microphone A and B depicted in Figure 13 is time-variant. Consequently, the STMDD $d = d_2 - d_1$ varies in time, equivalently resulting in variations in temperature readings. This may partially explain the frequency variations depicted in Figure 10. However, the displacement of speaker diaphragm on smartphones (usually < 2 W) is at sub-millimeter level, imposing limited impacts on the final temperature readings. To further reduce its impact, one can use high-frequency signals as the displacement is inversely proportional to frequency.

We have conducted extensive experiments under different volumes, and Figure 14 depicts the results. As expected, in Figure 14(a), when the volume increases from 5% to 40%, the frequency variance of main peak increases drastically from 0.015 Hz to 3.7 Hz. Meanwhile, Figure 14(c) reveals that the measurement accuracy would be affected when increasing the volume, which is particularly serious at 40% volume. This variance could be significantly damped by Kalman filter when the volume is below 20%, as shown in Figure 14(b). However, the situation could not be improved when the volume reaches certain extent, say around and beyond 40%. Figure 14(a) and (b) also show that Kalman filtering could lead to slight shifts in the frequency of the main peak. These minor shifts

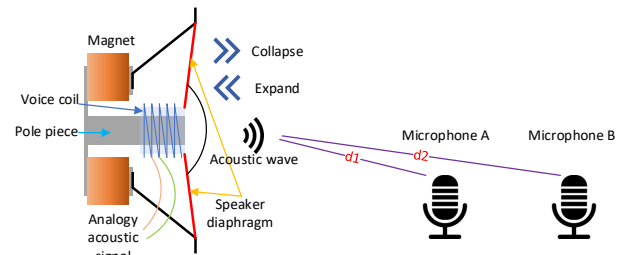


Figure 13: Speaker structure. The vibration of diaphragm generates acoustic signals.

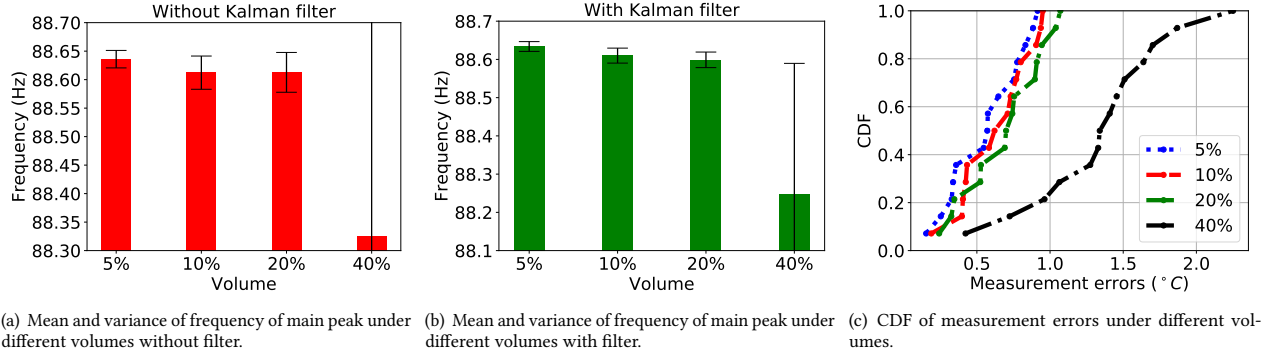


Figure 14: Impact of volume: Results in (a) and (b) show the mean and variance of frequency of main peak under different volumes with and without Kalman filter, respectively. They reveal that the variance increases drastically when increasing the volume to a certain extent (40%). Results in (c) show the CDF of measurement errors under different volumes. It is clear that tuning down the volume can effectively improve measurement accuracy.

can be deemed as constant bias and hence can be compensated to leave no impact on the measurement accuracy.

4.2.3 Mutipath Effect. To explore the performance of AcuTe under multipath effect, we have conducted experiments at different environments. Specifically, we let a smartphone (type A, see the next section for details) perform sensing at four different scenarios: an open space at the top of a building, a corner of an indoor room, a narrow balcony, and a small kitchen. These four scenarios are termed Env1 to Env4 in Figure 15(a). The results in Figure 15(a) reveals that AcuTe exhibits almost identical performance at different multipath scenarios, demonstrating its robustness.

4.2.4 Different Smartphones. We have conducted evaluations using different smartphones. Specifically, we choose smartphones with different acoustic sensor layouts for experimentation. We first make a survey of popular smartphones to inspect typical sensor layout and Figure 16 shows the results. Among over 50 surveyed smartphones, type A and B shown in Figure 16(a) are the most commonly used, where two microphones are placed at top and bottom, and the speaker is close to the bottom microphone (usually >2cm apart). Over 95% smartphones adopt these two layouts; they

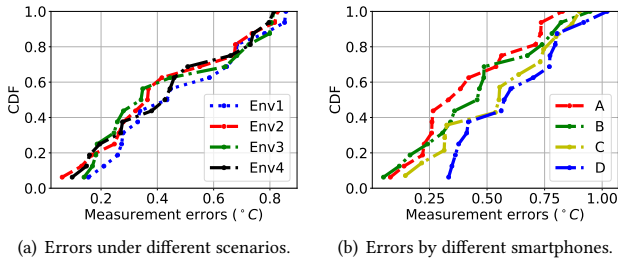


Figure 15: Results in (a) validate that AcuTe is robust to different multipath scenarios, as the CDF of measurement errors under four multipath settings exhibit no difference. (b) presents that smartphones with distinctive sensor layouts exhibit slightly different performance.

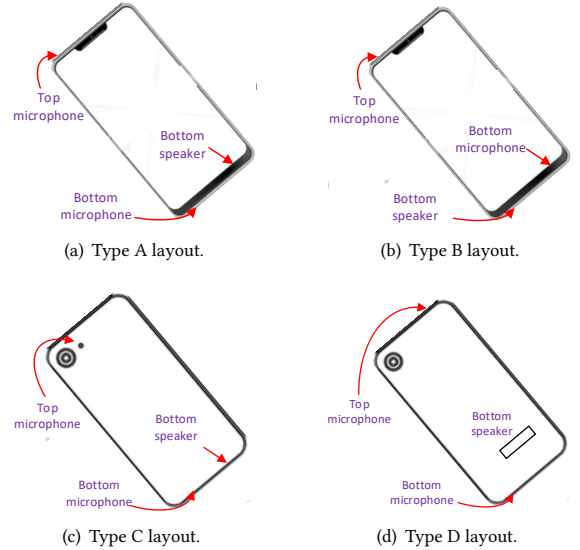


Figure 16: Acoustic transducer layouts of different phone models: front views in (a) & (b), and back views in (c) & (d).

typical include HUAWEI MATE30, HUAWEI Honor Play, OPPO A59, and Xiao MI8 SE. Type C and D are less common and occupy less than 5% of surveyed smartphones: One Plus 5T is type C while Samsung Galaxy S5 and HUAWEI MATE9 are type D. In each type, we select one smartphone as representative to evaluate the system performance and Figure 15(b) shows the results.

Figure 15(b) reveals that type A and B exhibit slightly better performance than type C and D. Using type A and B, AcuTe achieves a median accuracy of around 0.4°C, but that for type C and type D is only 0.6°C. This coincides with an observation that using smartphones of type A and B, the mixed spectrum have less multipath effect, i.e., less peaks than that of type C and D. A possible reason for this may be that speakers have more “leaked” acoustic signals inside the phone bodies, thus resulting in more severe multipath distortions hence worse measurement accuracy.

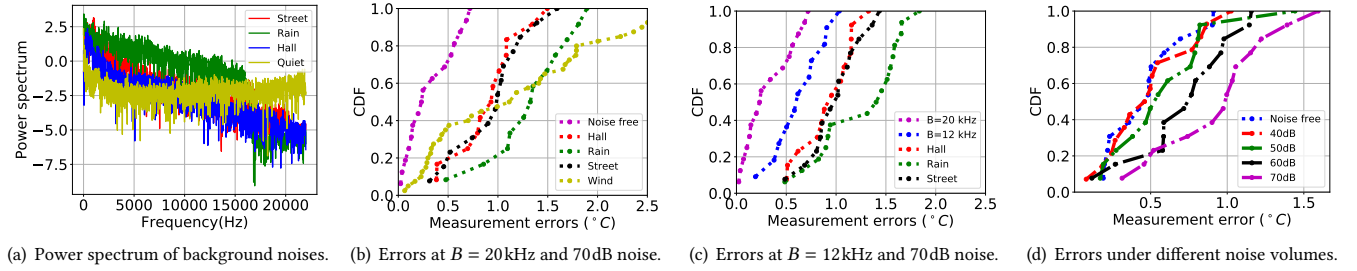


Figure 17: Impact of background noise and wind. (a) shows the power spectrum of different noises. (b) presents the CDF of measurement errors at 70dB noise or under wind (speed $\leq 1\text{m/s}$), compared with that without noise and wind. It is clear that the noises and wind can affect system performance. (c) gives the measurement performance at a reduced signal bandwidth $B = 12\text{kHz}$, compared with that without noise. It is observable that curtailing bandwidth does no help to improve the system performance. (d) demonstrates that lowering noise volume is effective in improving system performance with $B = 20\text{kHz}$.

4.2.5 Background Noise and Wind. Finally, we evaluate the performance under noisy and windy environments. We use another phone to play pre-recorded audio clips at 70dB, emulating background interference during the experiments. Here we choose 70dB as it is a threshold when people become distracted and cannot concentrate their attention, but most background noises from common places are likely to be below 70dB. Specifically, we choose to record the audio clips at three natural scenarios: under a roof while it rains heavily, inside a concert hall where there are crowds talking, on a busy street where there are speedy cars. The power spectra of these noises are shown in Figure 17(a). And we use a electric fan to generate wind with different speeds during experiments.

The CDF of measurement results are shown in Figure 17(b). It reveals that the performance is indeed affected by the 70dB strong interference: while the median error is below 0.3°C without noise, that under 70 dB interference of rainy sounds may reach 1.3°C . This 0.3°C median error of AcuTe significantly outperforms that reported in [25], where the average error reaches up to 1.25°C . Figure 17(b) also reveals a median accuracy of 1.2°C under a wind speed below 1m/s , but increasing the wind speed to $2\text{--}3\text{m/s}$ could heavily affect the performance (errors may reach up to 10°C and hence not presented). In fact, thermometers used by weather stations normally conduct measurements in a shaded box. Therefore, we recommend to operate AcuTe under hand-cover like using a lighter in the presence of wind.

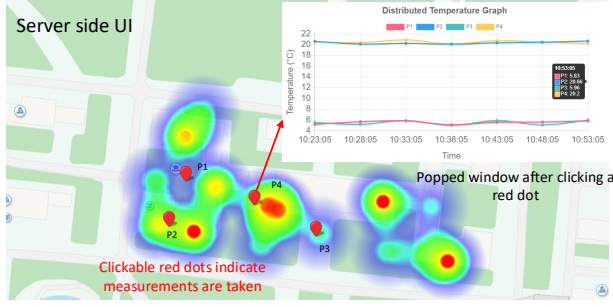
One might think that the performance degradation may come from the overlapped signal bandwidth between employed chirp signals and background noises. To check this, we curtail the signal bandwidth to 12kHz , ranging from 10kHz to 22kHz , so as to minimize the impact of background interference. The performance under the above configurations are shown in Figure 17(c), which appears to exhibit no noticeable improvements but degrades a lot even under the noise-free cases. We believe that sensing at reduced bandwidth resulting in no improvements may be caused by the following reasons. On one hand, curtailing signal bandwidth can lose resolution as explained in Equation (11). This may certainly undermine the performance even without noise. On the other hand, the processing gain $B \times T$ (thus SNR) is reduced due to the less signal bandwidth, making it more vulnerable to interference, hence worse performance when the strong 70dB interference presents.

Fortunately, if we reduce the background noise, say below 50dB as shown in Figure 17(d), AcuTe suffers from no noticeable performance degradation. The perceivable performance degradation starts to appear only when the background noise reach over 60dB, yet the median error is still well below 1°C . Since the background noise are below 70dB under most circumstances, we conclude that AcuTe is robust to common noises. Also, as we are now fully aware of the impact of the background noise and wind-force, we are able to provide useful guidelines for our users to properly operate AcuTe under favourable conditions.

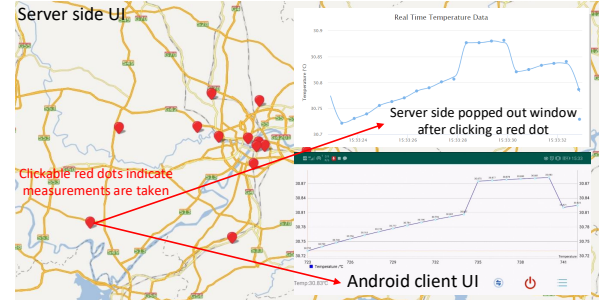
4.3 Distributed Temperature Monitoring

As one of the most important motivation of AcuTe, distributed temperature monitoring becomes possible with our aforementioned extensive evaluations on individual devices. In this section, we first report our scale-limited experiments on distributed monitoring, then we present an extension to wide-scale applications.

4.3.1 Small-scale Temperature Distribution. It is a common phenomenon that, even when room temperature is already cool enough from the thermal comforting point of view, the air conditioner still keeps working, wasting significant amount of energy. This is partially due to the fact that the air conditioner measures room temperature at a single point close to it. Therefore, it either stops working when the measured temperature reaches a certain threshold but leaves much of the room uncooled, or it keeps working so as to reach a certain level of heat equilibrium that may not ever be reachable (e.g., a large room). The situation could be improved if there are multiple temperature sensors distributed in the room, reporting an instantaneous temperature distribution rather than just single-point measurements. However, deploying a sensing infrastructure for this purpose can be impractical, yet with AcuTe, this problem becomes readily solvable. When people present in a room, their smartphones, possibly distributed in different places, can perform real-time temperature sensing. The results can then be geo-tagged based on certain indoor localization systems [30, 58, 59, 67]. Now what we have distributed temperature readings collected in real-time, allowing for a cheaper, faster, and more precise room temperature monitoring and thus efficient actuation including air conditioner control.



(a) Small-scale temperature distribution.



(b) Wide-scale surface temperature mapping.

Figure 18: Interfaces of distributed temperature monitoring application, with a campus-scale view shown in (a) and a nation-wide view shown in (b). The red dots on the map show positions where temperature measurements are taken; clicking any of them invokes a pop-up window to visualize real-time temperature changing on the server side.

To demonstrate the feasibility of monitoring scale-limited temperature distribution, we implement an system prototype using Node.js (server side application) and Android (client side application). We deploy this distributed monitoring system in our school: the android devices report their instantaneous temperature readings to a remote server restoring and plotting data at different time scales. Also, the average results over certain time window are visualized over a map, and Figure 18(a) shows the five-minute scale results. The results clearly showcase the feasibility of distributed temperature monitoring of our systems.

4.3.2 Wide-Scale Surface Temperature Mapping. Obtaining a precision geo-temperature is of vital significance for its potential uses in finer weather forecasting, precision agriculture, etc. Specially, distributed near-surface air temperature is of vital importance in hydrological modeling in meteorology [68] due to its impacts on earth surface system. Currently, air temperature monitoring is obtained through dedicated air stations. Due to their heavy cost, they are sparsely spaced hence low-spatial-density temperature data. With our technology, we can utilize ubiquitous smartphones worldwide to perform geo-location dependent air temperature sensing. Combined with a crowdsourcing platform, we can build a ever precise temporal and spatial nation-wide temperature graph, an almost infeasible task for current approaches.

We have deployed our prototype on a cloud service and ask several users distributed in a state province to upload their temperature readings. Unfortunately, our distributed experiments were hit by the COVID-19 pandemic and hence failed to reach their full scale. However, it is indeed the networked crowdsensing ability of AcuTe (piggybacking on pervasively available devices) that still allows us to report the following results even under this extremely difficult time. Figure 18(b) shows the snapshots of user interface for both the server and the Android application. The plotted temperature line (spacing in two seconds) shown on the server side and target side are highly consistent, demonstrating the feasibility of large-scale and real-time temperature monitoring.

4.4 Discussions and Limitations

AcuTe requires a calibration to obtain the STMDD d (see Section 3.1) for accurate sound speed estimation. Although this only needs to

be done once for all, it can still be a difficult task for common users. Therefore, we recommend that phone manufactures to accomplish this task and make it an accessible parameter to any applications. In addition, as AcuTe utilizes sound speed to estimate temperature, it is hence vulnerable to wind-blowing: our measurements reveals up to 10°C temperature variation if wind speed reaches 2-3 m/s. Consequently, the best practice is to shield a smartphone from wind when conducting temperature sensing.

It is worth noting that AcuTe should be performed outside bag or pocket as it requires LoS signal transmissions. However, holding a smartphone during a measurement does not matter, as far as the microphones and speaker are not covered. Also, cross-phone interference is nearly impossible due to low volume and high signal attenuation, yet such interference (caused by phone placing too close) can be handled by letting them collaboratively estimate temperature [8]. We also note that AcuTe aims at measuring ambient air temperature rather than that of an object. A feasible solution for measuring object temperature may need to exploit reflected signals potentially affected by the heat dissipation from that object. Whereas extending AcuTe for this purpose can be a feasible, an alternative solution may explore radio frequency signals that have been adopted for sensing vibration [55] and activities [15].

5 CONCLUSION

In this paper, we present the first acoustic thermometer empowered by a single smartphone. AcuTe utilizes on-board dual microphones to estimate sound speed, thereby deriving ambient air temperature. In AcuTe, we propose a pipeline of signal processing blocks to denoise received signals and a multipath handling algorithm to account for multiple reverberations, enhancing system robustness. We extensively evaluate AcuTe at representative environments and the results demonstrate a median accuracy of 0.3°C . We believe AcuTe has potential to reshape the status of current distribute temperature monitoring.

6 ACKNOWLEDGEMENTS

We are grateful to anonymous shepherd and reviewers for their valuable comments. This research is supported in part by AcRF Tier 2 Grant MOE2016-T2-2-022 and AcRF Tier 1 Grant RG17/19.

REFERENCES

- [1] ANALOG DEVICES (ADI). Temperature Measurement Theory and Practical Techniques. https://www.analog.com/media/en/technical-documentation/application-notes/an_892.pdf, 2020.
- [2] ANDERSON, M., NORMAN, J., DIAK, G., KUSTAS, W., AND MECIKALSKI, J. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote Sensing of Environment* 60, 2 (1997), 195–216.
- [3] BANKMYCELL. How Many Smartphones Are in the World? <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world>, 2020.
- [4] BARBATO, A., BORSANI, L., CAPONE, A., AND MELZI, S. Home Energy Saving through a User Profiling System Based on Wireless Sensors. In *Proc. of ACM Workshop on BuildSys* (2009), pp. 49–54.
- [5] BI, C., AND XING, G. Real-Time Attitude and Motion Tracking for Mobile Device in Moving Vehicle. In *Proc. of the 16th ACM Sensys* (2018), pp. 357–358.
- [6] BIES, D., HANSEN, C., AND HOWARD, C. *Engineering Noise Control, Fifth Edition*. CRC Press, 11 2017.
- [7] CAI, C. Acoustic Thermometer Empowered by a Single Smartphone. https://caichao.github.io/proj_dirs/acute.html, 2020.
- [8] CAI, C., PU, H., HU, M., ZHENG, R., AND LUO, J. SST: Software Sonic Thermometer on Acoustic-enabled IoT Devices. *IEEE Transactions on Mobile Computing* (2020), 1–14.
- [9] CAMPBELL, S. D., AND DIEBOLD, F. X. Weather Forecasting for Weather Derivatives. *Journal of the American Statistical Association* 100, 469 (2005), 6–16.
- [10] CASEY, J. World Climate Maps. <https://www.climate-charts.com/World-Climate-Maps.html>, 2020.
- [11] CELIK, M., DADASER-CELIK, F., AND DOKUZ, A. S. Anomaly Detection in Temperature Data using DBSCAN Algorithm. In *2011 International Symposium on Innovations in Intelligent Systems and Applications* (2011), pp. 91–95.
- [12] CHAUHAN, J., HU, Y., SENEVIRATNE, S., MISRA, A., SENEVIRATNE, A., AND LEE, Y. BreathPrint: Breathing Acoustics-based User Authentication. In *Proc. of the 15th ACM MobiSys* (2017), pp. 278–291.
- [13] CHEN, J., TAN, R., WANG, Y., XING, G., WANG, X., WANG, X., PUNCH, B., AND COLBRY, D. A Sensor System for High-Fidelity Temperature Distribution Forecasting in Data Centers. *ACM Trans. Sens. Netw.* 11, 2 (Dec. 2014), 30:1–25.
- [14] CLAASEN, T., AND MECKLENBRÄUKER, W. The Wigner Distribution—A tool for Time–Frequency Signal Analysis—Part II: Discrete Time Signals. *Philips Research* 35 (Jan 1980), 276–300.
- [15] DING, S., CHEN, Z., ZHENG, T., AND LUO, J. RF-Net: A Unified Meta-Learning Framework for RF-enabled One-Shot Human Activity Recognition. In *Proc. of the 18th ACM SenSys* (2020), pp. 1–14. <https://doi.org/10.1145/3384419.3430735>.
- [16] EL-SAYED, N., STEFANOVICI, I. A., AMVROSLADIS, G., HWANG, A. A., AND SCHROEDER, B. Temperature Management in Data Centers: Why Some (Might) like It Hot. In *Proc. of the 12th ACM SIGMETRICS* (2012), pp. 163–174.
- [17] FISHER, R. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. In *Biometrika* (1915), vol. 10, pp. 507–521.
- [18] GAO, H., MATTERS-KAMRNER, M. K., HARPE, P., MILOSEVIC, D., JOHANNSEN, U., VAN ROERMUND, A., AND BALTUS, P. A 71GHz RF Energy Harvesting Tag with 8% Efficiency for Wireless Temperature Sensors in 65nm CMOS. In *Proc. of 2013 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (June 2013), pp. 403–406.
- [19] GAYEN, A. The Frequency Distribution of the Product Moment Correlation Coefficient in Random Samples of Any Size Draw from Non-Normal Universes. In *Biometrika* (1951), vol. 38, pp. 219–247.
- [20] GOOGLE PLAY. Sound Meter. https://play.google.com/store/apps/details?id=com.gamebasic.decibel&hl=en_SG, 2020.
- [21] GUO, B., WANG, Z., YU, Z., WANG, Y., YEN, N. Y., HUANG, R., AND ZHOU, X. Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm. *ACM Comput. Surv.* 48, 1 (2015).
- [22] GUPTA, V., MITTAL, S., BHAMIK, S., AND ROY, R. Assisting Humans to Achieve Optimal Sleep by Changing Ambient Temperature. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2016), pp. 841–845.
- [23] HAN, K., ZHANG, C., AND LUO, J. Taming the Uncertainty: Budget Limited Robust Crowdsensing through Online Learning. *IEEE/ACM Trans. on Networking* 24, 3 (2016), 1462–1475.
- [24] HAN, K., ZHANG, C., LUO, J., HU, M., AND VEERAVALLI, B. Truthful Scheduling Mechanisms for Powering Mobile Crowdsensing. *IEEE Trans. on Computers* 65, 1 (2016), 294–307.
- [25] HE, L., LEE, Y., AND SHIN, K. G. Mobile Device Batteries as Thermometers. In *Proc. of ACM Ubicomp* (2020), pp. 1–21.
- [26] HE, Y., LIANG, J., AND LIU, Y. Pervasive Floorplan Generation Based on Only Inertial Sensing: Feasibility, Design, and Implementation. *IEEE Journal on Selected Areas in Communications* 35, 5 (2017), 1132–1140.
- [27] JAIN, M., SINGH, A., AND CHANDAN, V. Portable+: A Ubiquitous And Smart Way Towards Comfortable Energy Savings. *Proc. of ACM UbiComp* (2017).
- [28] KAIMAL, J. C., AND GAYNOR, J. E. Another Look at Sonic Thermometry. *Boundary-Layer Meteorology* 56, 4 (Sep 1991), 401–410.
- [29] KOYAMADA, Y., IMAHAMA, M., KUBOTA, K., AND HOGARI, K. Fiber-Optic Distributed Strain and Temperature Sensing With Very High Measurand Resolution Over Long Range Using Coherent OTDR. *J. Lightwave Technol.* 27, 9 (May 2009), 1142–1146.
- [30] LIU, S., AND HE, T. SmartLight: Light-Weight 3D Indoor Localization Using a Single LED Lamp. In *Proc. of the 15th ACM SenSys* (2017), pp. 1–14.
- [31] LO, C. P., QUATTROCHI, D. A., AND LUVALL, J. C. Application of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *International Journal of Remote Sensing* 18, 2 (1997), 287–304.
- [32] LU, C. X., LI, Y., ZHAO, P., CHEN, C., XIE, L., WEN, H., TAN, R., AND TRIGONI, N. Simultaneous Localization and Mapping with Power Network Electromagnetic Field. In *Proc. of 24th ACM MobiCom* (2018), pp. 607–622.
- [33] MANDAL, J., PAL, S., SUN, T., GRATAN, K. T. V., AUGUSTI, A. T., AND WADE, S. A. Bragg grating-based fiber-optic laser probe for temperature sensing. *IEEE Photonics Technology Letters* 16, 1 (Jan 2004), 218–220.
- [34] MAO, W., HE, J., ZHENG, H., ZHANG, Z., AND QIU, L. CAT: High-Precision Acoustic Motion Tracking. In *Proc. of the 22th ACM MobiCom* (2016), pp. 69–81.
- [35] MAO, W., WANG, M., AND QIU, L. AIM: Acoustic Imaging on a Mobile. In *Proc. of the 16th ACM MobiSys* (2018), pp. 468–481.
- [36] MAO, W., ZHANG, Z., QIU, L., HE, J., CUI, Y., AND YUN, S. Indoor Follow Me Drone. In *Proc. of the 15th ACM MobiSys* (2017), pp. 345–358.
- [37] MENDELSON, R., NORDHAUS, W. D., AND SHAW, D. The Impact of Global Warming on Agriculture: A Ricardian Analysis. *The American Economic Review* 84, 4 (1994), 753–771.
- [38] NANDAKUMAR, R., GOLLAKOTA, S., AND WATSON, N. Contactless Sleep Apnea Detection on Smartphones. In *Proc. of the 13th ACM MobiSys* (2015), pp. 45–57.
- [39] PARINUSSA, R. M., HOLMES, T. R. H., YILMAZ, M. T., AND CROW, W. T. The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Journal of Hydrology and Earth System Sciences* 15, 10 (2011), 3135–3151.
- [40] PEACOCK, G. R. Temperature Sensors: Contact or Noncontact? <https://www.sensorsmag.com/components/temperature-sensors-contact-or-noncontact>, 2018.
- [41] PEREIRA DE SOUZA NETO, E., CUSTAUD, M.-A., FRUTOSO, J., SOMODY, L., GHARIB, C., AND FORTRAT, J.-O. Smoothed Pseudo Wigner-Ville Distribution as an Alternative to Fourier Transform in Rats. *Autonomic Neuroscience* 87, 2 (2001), 258–267.
- [42] ROBERT, P. C. Precision Agriculture: A Challenge for Crop Nutrition Management. In *Progress in Plant Nutrition: Plenary Lectures of the XIV International Plant Nutrition Colloquium: Food security and sustainability of agro-ecosystems through basic and applied research*, W. J. Horst, A. Bürkert, N. Claassen, H. Fleßa, W. B. Frommer, H. Goldbach, W. Merbach, H.-W. Olfs, V. Römhild, B. Sattelmacher, U. Schmidhalter, M. K. Schenk, and N. v. Wirén, Eds. Springer, 2002, pp. 143–149.
- [43] ROY, N., AND ROY CHOUDHURY, R. Listening through a Vibration Motor. In *Proc. of the 14th ACM MobiSys* (2016), pp. 57–69.
- [44] SCHUMANN, A., PLOENNIGS, J., AND GORMAN, B. Towards Automating the Deployment of Energy Saving Approaches in Buildings. In *Proc. of ACM BuildSys* (2014), pp. 164–167.
- [45] SHAKER, G., TENTZERIS, M., AND SAFAVI-NAEINI, S. Low-cost antennas for mm-Wave sensing applications using inkjet printing of silver nano-particles on liquid crystal polymers. In *Proc. of 2010 IEEE Antennas and Propagation Society International Symposium* (July 2010), pp. 1–4.
- [46] SHU, Y., SHIN, K. G., HE, T., AND CHEN, J. Last-Mile Navigation Using Smartphones. In *Proc. of the 21st ACM MobiCom* (2015), pp. 512–524.
- [47] SONG, C., ZHOU, X., AND LIU, J. Investigation of Human Thermal Comfort in Sleeping Environments Based on the Effects of Bed Climate. *Procedia Engineering* (2015), 1126–1132.
- [48] SOZZI, R., AND FAVARON, M. Sonic Anemometry and Thermometry: Theoretical Basis and Data-Processing Software. *Environmental Software* 11, 4 (1996), 259–270.
- [49] SUN, K., WANG, W., LIU, A. X., AND DAI, H. Depth Aware Finger Tapping on Virtual Displays. In *Proc. of the 16th ACM MobiSys* (2018), pp. 283–295.
- [50] SUN, K., ZHANG, T., WANG, W., AND XIE, L. VSkin: Sensing Touch Gestures on Surfaces of Mobile Devices Using Acoustic Signals. In *Proc. of the 24th ACM MobiCom* (2018), pp. 591–605.
- [51] TEXAS INSTRUMENTS (TI). Design Considerations for Measuring Ambient Air Temperature. <http://www.ti.com/lit/an/snoa966b/snoa966b.pdf>, 2020.
- [52] TUNG, Y.-C., AND SHIN, K. G. Expansion of Human-Phone Interface By Sensing Structure-Borne Sound Propagation. In *Proc. of the 14th ACM MobiSys* (2016), pp. 277–289.
- [53] WANG, A., SUNSHINE, J. E., AND GOLLAKOTA, S. Contactless Infant Monitoring Using White Noise. In *Proc. of 25th ACM MobiCom* (2019), pp. 1–16.
- [54] WANG, J., TAN, N., LUO, J., AND PAN, S. WOLoc: WiFi-only Outdoor Localization Using Crowdsensed Hotspot Labels. In *Proc. of the 36th IEEE INFOCOM* (2017), pp. 1–9.
- [55] WANG, Z., CHEN, Z., SINGH, A., GARCIA, L., LUO, J., AND SRIVASTAVA, M. UWHEar: Through-wall Extraction and Separation of Audio Vibrations Using Wireless Signals. In *Proc. of the 18th ACM SenSys* (2020), pp. 1–14. <https://doi.org/10.1145/3384419.3430772>.

- [56] WU, C., ZHANG, F., FAN, Y., AND LIU, K. J. R. RF-Based Inertial Measurement. In *Proc. of the 31th ACM SIGCOMM* (2019), pp. 117–129.
- [57] XIE, B., TAN, G., AND HE, T. SpinLight: A High Accuracy and Robust Light Positioning System for Indoor Applications. In *Proc. of the 13th ACM Sensys* (2015), pp. 211–223.
- [58] XU, C., FIRNER, B., ZHANG, Y., AND HOWARD, R. E. The Case for Efficient and Robust RF-Based Device-Free Localization. *IEEE Transactions on Mobile Computing* 15, 9 (2016), 2362–2375.
- [59] XU, Q., ZHENG, R., AND HRANILOVIC, S. IDyLL: Indoor Localization Using Inertial and Light Sensors on Smartphones. In *Proc. of ACM Ubicomp* (2015), pp. 307–318.
- [60] XU, X., SHEN, Y., YANG, J., XU, C., SHEN, G., CHEN, G., AND NI, Y. PassiveVLC: Enabling Practical Visible Light Backscatter Communication for Battery-Free IoT Applications. In *Proc. of the 23rd ACM MobiCom* (2017), pp. 180–192.
- [61] XU, X., YU, J., CHEN, Y., ZHU, Y., KONG, L., AND LI, M. BreathListener: Fine-Grained Breathing Monitoring in Driving Environments Utilizing Acoustic Signals. In *Proc. of the 17th ACM MobiSys* (2019), pp. 54–66.
- [62] YAN, T., MARZILLI, M., HOLMES, R., GANESAN, D., AND CORNER, M. mCrowd: A Platform for Mobile Crowdsourcing. In *Proc. of the 7th ACM SenSys* (2009), pp. 347–348.
- [63] YANG, Y., HAO, J., LUO, J., AND PAN, S. CeilingCast: Energy Efficient and Location-Bound Broadcast Through LED-Camera Communication. In *Proc. of the 35th IEEE INFOCOM* (2016), pp. 1–9.
- [64] ZHANG, C., KUPPANNAGARI, S. R., KANNAN, R., AND PRASANNA, V. K. Building HVAC Scheduling Using Reinforcement Learning via Neural Network Based Model Approximation. In *Proc. of the 6th ACM BuildSys* (2019), pp. 287–296.
- [65] ZHANG, C., SUBBU, K., LUO, J., AND WU, J. GROPING: Geomagnetism and cROwd-sensing Powered Indoor NaviGation. *IEEE Trans. on Mobile Computing* 14, 2 (2015), 387–400.
- [66] ZHEN, Q., HEINSCH, M. A., ZHAO, M., AND W.RUNNING, S. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment* 111, 4 (2007), 519 – 536.
- [67] ZHOU, B., ELBADRY, M., GAO, R., AND YE, F. BatTracker: High Precision Infrastructure-Free Mobile Device Tracking in Indoor Environments. In *Proc. of the 15th ACM SenSys* (2017), pp. 1–14.
- [68] ZHOU, W., PENG, B., SHI, J., WANG, T., DHITAL, Y. P., YAO, R., YU, Y., LEI, Z., AND ZHAO, R. Estimating High Resolution Daily Air Temperature Based on Remote Sensing Products and Climate Reanalysis Datasets over Glacierized Basins: A Case Study in the Langtang Valley, Nepal. *Remote Sensing* 9, 9 (2017).