SST: Software Sonic Thermometer on Acoustic-enabled IoT Devices

Chao Cai, *Student Member, IEEE,* Henglin Pu, Menglan Hu, Rong Zheng, *Senior Member, IEEE,* and Jun Luo, *Senior Member, IEEE*

Abstract—Temperature is an important data source for weather forecasting, agriculture irrigation, anomaly detection, etc. While temperature measurement can be achieved via low-cost yet standalone hardware with reasonable accuracy, integrating thermal sensing into ubiquitous computing devices is highly non-trivial due to the design requirement for specific heat isolation and proper device layout. In this paper, we present the first integrated thermometer using commercial-off-the-shelf acoustic-enabled devices. Our Software Sonic Thermometer (SST) utilizes on-board dual microphones on commodity mobile devices to estimate sound speed, which has a known relation with temperature. To precisely measure temperature via sound speed, we propose a chirp mixing approach to circumvent low sampling rates on commodity hardware and design a pipeline of signal processing blocks to handle channel distortions. SST, for the first time, empowers ubiquitous computing devices with thermal sensing capability. It is portable and cost-effective, making it competitive with current thermometers using dedicated hardware. SST is potential to facilitate many interesting applications such as large-scale distributed thermal sensing, yielding high temporal/spatial resolutions with unimaginable low costs. We implement SST on a commodity platform and results show that SST achieves a median accuracy of $0.5°C$ even at varying humidity levels.

Index Terms—Acoustic sensing, commodity hardware, temperature, sound speed.

✦

1 INTRODUCTION

Temperature is a type of useful empirical data for a myriad of applications such as weather forecasting [1], agriculture irrigation [2], anomaly detection [3], climate change prediction [4], and health care [5]. Specially, near-surface air temperature obtained through distributed thermal sensing, plays a pivotal role in land surface and hydrological modeling in meteorology [6] due to its impacts on earth surface system. Currently, thermal sensing are accomplished by dedicated temperature sensors. Though the temperature sensors are very cheap and can achieve precise measurement, it is not ubiquitous right now. It is much desirable to empower temperature measurement on a commercial-off-the-shelf (COTS) pervasive devices and thus enable many interesting applications such as large-scale distributed thermal sensing. However, it is non-trivial to integrate a temperature sensing into a ubiquitous computing devices due to the design requirement for specific heat isolation and proper device layout [7], [8]. Therefore, alternative thermal sensing approaches are called for.

Nowadays, smartphones and Internet of Things (IoT) devices enjoy explosive growth. In 2019, mobile users worldwide are projected to reach 4.68 billion, among which 3 billion use smartphones [9]. Smart devices now have rich on-board sensors and powerful processors, enabling ubiquitous computing and sensing [10], [11], [12], [13]. Most if not all of these devices are network ready and in many cases (e.g., smartphones) can provide their geo-locations. *In this case, can we turn pervasive mobile devices into a ubiquitous thermometer using existing on-board sensors?*

In this paper, we design the first Software Sonic Thermometer (SST) on COTS IoT devices. SST empowers acoustic-enabled IoT devices, smartphones, and wearables to perform accurate ambient air temperature sensing. We leverage on-board dual microphones to estimate sound speed, which has a known relationship with temperature, thus accomplishing ambient air temperature sensing. SST is portable and cost-effective, making it competitive with traditional thermometers for ubiquitous sensing. With recent advances in mobile crowdsourcing, SST has the potential to leverage the ubiquitous IoT devices to implement large-scale distributed temperature sensing and yield high temporal and spatial resolutions with extremely low costs. Such an approach may reshape the landscape for large-scale distributed temperature acquisition.

However, to implement SST on COTS devices is non-trivial. First, in order to exploit the relationship between acoustic speed and ambient temperature, microsecond timing resolution is required. Consider the time-difference-of-arrival (TDoA) estimate between two microphones on either end of a Samsung S5 phone from a single acoustic source [1]. At a room temperature of

- *C. Cai, H. Pu, and M. Hu are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: {cchust, puhenglin, humenglan}@hust.edu.cn. (Corresponding author: Menglan Hu.)*
- *R. Zheng is with the Dept. of Computing and Software, McMaster University, Hamilton L8S 4L8, Canada. E-mail: rzeng@mcmaster.ca.*
- *J. Luo is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798. E-mail: junluo@ntu.edu.sg.*

1. Acoustic speed can be determined as the ratio between the distance of the two microphones and TDoA.

25 °C (= 298.15 K), a 10-degree increase in temperature roughly corresponds to a 3.3% change in acoustic speed if we assume a TDoA of 471 μs (= 0.16 m/340 m/s). To detect the 10-degree change, a timing resolution of 15 μs is required. For changes within 1-degree, one needs a timing resolution of 1.5 μs. However, acoustic modules on smartphones typically can only sample up to 48 kHz or equivalently at sampling intervals around 21 μs. Consequently, the sampling rate in smartphones is insufficient to achieve satisfactory thermal sensing resolution. Second, acoustic modules on COTS devices are typically optimized for the audible range [14], peaking around 8 kHz and decaying at lower and higher frequency. As a result, their frequency responses tend to be non-flat. In addition, acoustic sensors, either across different devices or on the same device, exhibit different properties known as device diversity. The non-flat frequency response and device diversity result in channel distortions, inevitably affecting sensing results.

To circumvent the insufficient sampling rate and break the resolution limit, we adopt a chirp mixing approach. This approach transforms the temperate associated timing information into a more fine-grained frequency domain, thereby yielding highly accurate temperature resolution. To deal with device diversity, non-flat frequency response, and other severe channel distortions, we design a signal processing pipeline, mitigating channel distortions and achieving accurate and stable thermal readings. Simply put, this paper makes the following contributions:

- We design the first Software Sonic Thermometer (SST) on commodity hardware, achieving sub-centigrade resolution.
- We propose a signal processing pipeline which can successfully detect sound speed changes at a granularity of decimeter per second, enabling sub-centigrade thermal sensing.
- We have implemented SST on commodity hardware and conducted extensive experiment studies to demonstrate the robustness of SST under different environmental conditions.

The rest of our paper is organized as follows: Section 2 introduces the related work. Section 3 the basic principle of SST and detailed design. Implementation and experimental evaluation are given in Section 6. We discuss several issues and future works in Section 7. Finally, Section 8 concludes the paper.

2 RELATED WORK

There are two well-known categories of thermal sensing approaches, contact and non-contact methods [15]. Contact sensing methods require sensors to physically contact the device under test (DUT). These sensors, however, can only probe a specific temperature point and the measurement process may significantly change the temperature of the DUT. In addition, this line of approaches has to wait for a heat equilibrium between

sensors and DUT. As a result, they usually incur long latencies before precise temperature readings can be obtained. Also, instruments to facilitate contact sensing usually require special designs. As documented by chip vendors including Texas Instrument [7] and Analog Device [8], dedicated devices to measure ambient temperature needs special designs including heat isolation, proper device layout, etc. However, today's smartphones are becoming more and more compact, leaving no room for the luxurious isolation design. Therefore, it is difficult to integrate these sensors on the ubiquitous mobile devices for accurate ambient air temperature sensing.

To cope with the shortcomings of the contact methods, non-contact approaches have been developed in the literature. This line of work has fast response time but are more vulnerable to environmental noises especially nearby heat radiations [16]. As the most widely known non-contact method, infrared-based approaches [17] measure the temperature of a DUT in its vicinity by gauging the radiated heat from the DUT, which however, is unsuitable for measuring the air temperature. Other non-contact thermal sensing techniques such as laser [18] or ultra-high frequency based methods [19] are costly and are more commonly used in high-end infrastructures.

Acoustic signals have been used to measure air temperature and speed in meteorology [20], [21]. These technologies also exploit the relationship between sound speed and temperature. Nevertheless, such solutions often target high-end devices and require either high sampling rates or ultrasonic signals, all infeasible on commodity mobile devices. In contrast, our SST is portable, non-contact, and cost-effective when compared with existing acoustic solutions.

To the best of our knowledge, we are the first to develop a purely software sonic thermometer on commodity devices. SST builds upon recent advances in acoustic sensing where acoustic signals have been used to facilitate localization [22], [23], [24], tracking [10], [25], [26], sensing [27], [28], and gesture recognition [29]. For instance, the authors in [29] leverage the phase of acoustic reflections to achieve high-precision gesture tracking and recognition. Mao et al., [10], [25] utilize a chirp mixing approach to accomplish sub-centimeter level tracking. ForcePhone [27] uses the properties of structure-born sound propagation to estimate applied force on commodity smartphones. Acoustic sensors can be used in synthetic aperture radars to inspect the shape of an object [30]. These exciting developments, especially the chirp mixing approaches in [25], [31], inspire us to investigate acoustic thermal sensing on commodity hardware. However, SST utilizes a much wide-band signals and suffer from more severe channel distortions. Consequently, the techniques in state-of-the-art work are not directly applicable for our system.

3 OVERVIEW OF SST

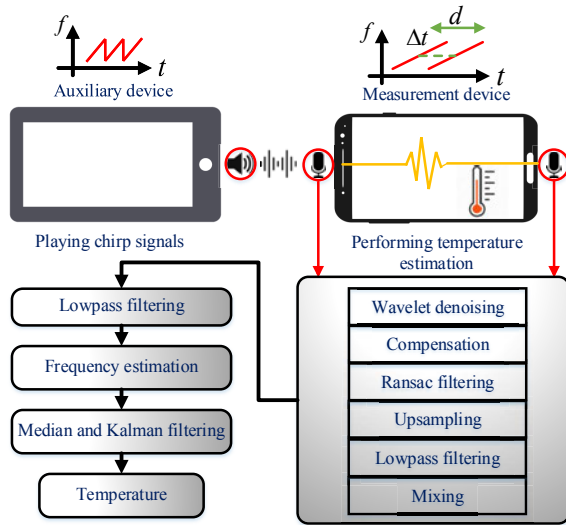SST consists of two devices: an auxiliary device that transmits chirp signals and a measurement device that

Fig. 1: System architecture of SST

receives and processes the audio samples for thermal sensing. An overview of the system architecture is shown in Fig. 1. The auxiliary device can be any device that is capable of playing back specified acoustic signals. The measurement device should have two separated microphones for audio inputs. Such a requirement is satisfied by many COTS mobile devices for the purpose of stereo recording and noise cancellation.

To perform thermal sensing, first, the auxiliary device is turned on to transmit chirp signals. To mitigate inter chirp interference as well as multipath reverberations, a guard interval is inserted between successive transmissions. Second, the measurement devices begin to capture the audio signals from its stereo channel. It synchronizes to one audio channel and extracts a chirp segment, over the same time window of which, it extracts the same amount of samples from the other channel. Since the two channels are strictly synchronized, the chirp segment extracted from one channel is a delayed version of that from the other channel. The delay that reflects the travel speed of acoustic signals, can be more precisely estimated via chirp mixing. Before mixing, the received signals from the two channels undergo a signal processing pipeline including wavelet denoising, compensation, RANSAC filtering, mitigating the channel distortions and yielding high resolution. Following that, the signals are mixed. The mixed results, after a low pass filter, can be used to extract frequency components, from which fine-grained temperature can be derived. We further apply median and Kalman filters to smooth the results.

The principle behind the above operation is the relation between sound speed c and air temperature T [20],

$$c^2 = 403T \left(1 + 0.32e/p\right), \tag{1}$$

where c is the sound speed (m/s) in air, T is the temperature (in Kelvin), e is vapor pressure of water in air, p is the absolute atmospheric pressure. In (1), $\frac{e}{p}$ has a close relation with humidity. Theoretically, the term

$\frac{e}{p}$ in (1) can be removed only in dry air. But previous work [21] and our experiments reveal that this term has negligible impacts on the final sensing results. Therefore, we further simplify Eq. (1) as follows,

$$c^2 = 403T. \tag{2}$$

4 BREAKING THE RESOLUTION LIMIT

In this section, we first present the basic idea behind chirp mixing and then discuss the key parameters and rationales for the chirp signal design. Finally, we present a feasibility study on a customized platform. From Eq. (2), we have $403T = c^2 = \left(\frac{d}{\Delta t}\right)^2$, where d is the *fixed* distance between two microphones and Δt is the TDoA of the acoustic signal from the auxiliary device to the two microphones of the measurement device. In time domain, to improve the resolution of T, we need to increase the resolution of Δt, which however, is limited by the sampling rate f_s of the device. To break the resolution limit, we adopt the chirp mixing approach [25], [10].

4.1 Chirp Mixing

Let the chirp signal emitted by the auxiliary device be $s = \cos\left(2\pi f_{\min}t + \pi kt^2\right)$, where $f_{\min}$ is the initial frequency and k denotes the modulation coefficient. The modulation coefficient is defined by $k = \frac{B}{D}$, where B is the bandwidth and D is the duration of the signal. A measurement device captures the emitted chirp signal using its two microphones separated by distance d. The received signals of the two channels are given by,

$$\begin{aligned} r_1 &= \alpha_1 \cos\left(2\pi f_{\min}t + \pi kt^2\right) \\ r_2 &= \alpha_2 \cos\left(2\pi f_{\min}\left(t - \tfrac{d}{c}\right) + \pi k\left(t - \tfrac{d}{c}\right)^2\right), \end{aligned} \tag{3}$$

where $\frac{d}{c}$ is the delay due to propagation; α_1 and α_2 are complex attenuations. We mix r_1, r_2 at the receiver and obtain the following results:

$$\begin{aligned} r_{\mathrm{mix}} &= r_1 \times r_2 \\ &= \alpha_1\alpha_2 \cos\left(2\pi f_{\min}t + \pi kt^2\right) \times \\ &\qquad \cos\left(2\pi f_{\min}\left(t - \tfrac{d}{c}\right) + \pi k\left(t - \tfrac{d}{c}\right)^2\right) \\ &= \tfrac{\alpha_1\alpha_2}{2} \cos\left(2\pi f_{\min}\tfrac{d}{c} + \pi k\left(2t\tfrac{d}{c} - \left(\tfrac{d}{c}\right)^2\right)\right) + \text{(terms with} \\ &\qquad\qquad\qquad\qquad \text{high frequency).} \end{aligned} \tag{4}$$

After passing through a low pass filter and ignoring the attenuations, the mixed result becomes:

$$r_{\mathrm{f}} = \cos\left(2\pi f_{\min}\frac{d}{c} + \pi k\left(2t\frac{d}{c} - \left(\frac{d}{c}\right)^2\right)\right). \tag{5}$$

The frequency components of r_{f} can be obtained by taking the derivative of the phase component in (5) with respect to t. Since all the constant terms are removed, the frequency component is calculated as f:

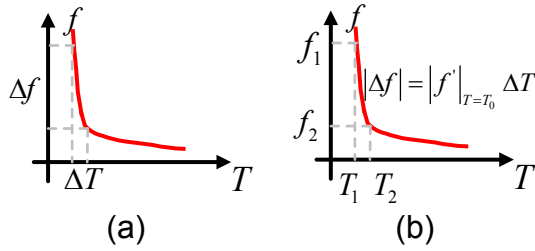$$f = \frac{1}{2\pi}\frac{\partial\theta}{\partial t} = \frac{kd}{c} \tag{6}$$

Fig. 2: Illustration to achieve the optimal chirp design. The design goal is to achieve a high temperature resolution so that when f spans a wide range, the corresponding changes for T should be small.

Fig. 3: Testbed for feasibility measurement

Since c, k, and d are all constants, and f is a scalar, consequently, r_f is a pure tone signal. Under such circumstance, f can be obtained with Discrete Fourier Transform (DFT) analysis and c can be obtained by:

$$c = \frac{kd}{f}. \tag{7}$$

Plugging Eq. (2) into Eq. (7), T can be calculated by:

$$T = \frac{k^2 d^2}{403 f^2}. \tag{8}$$

Eq. (8) gives us some tunable parameters, namely, k and f, to circumvent the insufficient sampling rate thereby achieving highly accurate thermal sensing.

4.2 Chirp Signal Design

The design of the chirp signal is crucial to system performance. In determining the design parameters including bandwidth B and duration D, a number of potentially conflicting factors should be taken into account.

First, the frequency range of the chirp signals should be kept small if possible to account for non-flat frequency responses of microphones and speakers (more details to be discussed in Section 5). Therefore, B should be small. Second, a high-resolution frequency estimation for f is desirable. The maximum resolution using Fourier transformation to estimate f is determined by the effective samples, which is equal to the multiplication between sampling rate f_s and the duration of chirp signal D. Since f_s cannot be changed, the larger the duration D, the higher the frequency resolution can be achieved.

However, D should not be set arbitrarily large as there is a trade-off here. First, to achieve a high temperature resolution, it is beneficial to maximize k. Consequently, a large chirp signal bandwidth B but a small D is profitable. Moreover, a large signal duration D may lead to severe self-inference and low signal-to-noise ratio (SNR) due to echoes in the environments. Finally, a large D also prolongs the measurement time. To help understand how to choose an optimal parameter for T, we first express f as a function of T from Eq. (8) as $f = \frac{kd}{403\sqrt{T}}$. Next, we take the derivate of f with respect
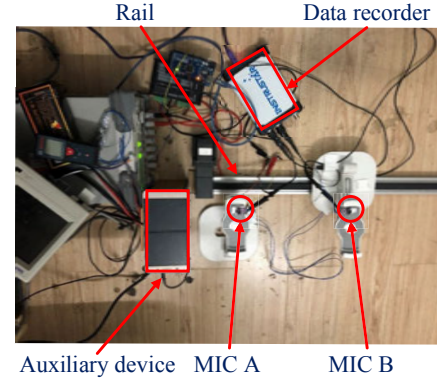
to T and have:

$$\frac{\Delta f}{\Delta T} \approx -\frac{k \times d}{806\sqrt{T^3}} = -\frac{B \times d}{806 D \sqrt{T^3}}. \tag{9}$$

As shown in Fig. 2, a large absolute derivative indicates that small variations in T result in large (detectable) changes in f, indicating smaller D is better.

Based on the aforementioned analysis as well as experiment evaluation, we choose a chirp signal that has a bandwidth of $B = 20$ kHz, ranging from $2 - 22$ kHz, and a duration of 0.01 s. Given that the typical temperature range is between -30 to $40\ °$C [1], these parameters offer us a resolution of $\frac{\Delta T}{\Delta f} = \frac{40-(-30)}{932.8-822.6} \approx 0.63\ °$C/Hz,[2] which is sufficient for most applications.

It should be noted that a larger two-microphone distance d, denoting the distance difference between the acoustic source and the two microphones, would be much desirable since it can improve the sensing accuracy. Therefore, it is recommended that the direction of the acoustic transmission from the auxiliary device be carefully placed so that the effective two-microphone distance d can be maximized (to the distance between the two microphones on the auxiliary device) and thus the accuracy can be boosted. This could be achieved by placing the speaker of the auxiliary device and two microphones of the measurement device in a straight line. Another thing worth noting is that the relative placements of the auxiliary device and the measurement device should be kept the same during the calibration phase and the runtime phase. This could be properly handled, for instance, by aligning the two devices to a common straightedge shaped by a wall or other concrete objects, or by manually placing marks [32] at the calibration phase to guarantee consistence between multiple placements.

4.3 Feasibility Measurement

To demonstrate the feasibility of SST, we have conducted a measurement on a customized platform. as shown in Fig. 3. This platform has multiple components: a data

2. This resolution is achieved by the premise that the distance difference d from the speaker to two microphones is $d = 0.146$ m.
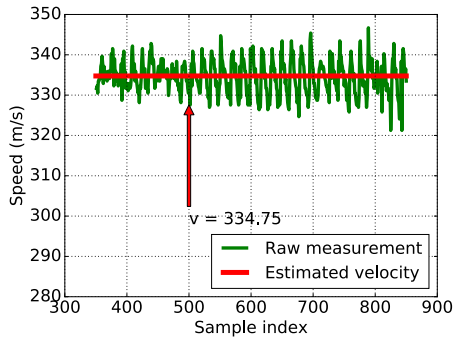
Fig. 4: Estimated sound speed using our testbed

recorder that has dual channels and flexible sampling rates up to 1 MHz, a rail that is controlled by an arduino board taking commands from a PC through serial interfaces, a smartphone with its output connected to a 3W classic-AB audio power amplifier, and two commodity microphone sensors with MAX9814, an Automatic Gain Control (AGC) to improve signal-to-noise-ratio. We wrote a MATLAB script to move the rail back and forth at a precision of 0.5 mm and developed an Android APP to transmit chirp signals of various settings. To measure the ground truth temperature, we use an expensive commercial thermometer that achieves $0.1\,^{\circ}$C and $1.5\,\%$RH accuracy in estimating temperature and humidity.

In this experiment, the smartphone acting as auxiliary device transmits a chirp signal incessantly. The chirp signal has a initial frequency of $f_{\min} = 15$ kHz, 7 kHz frequency bandwidth, and a duration of 0.02 s. We first record a set of measurements from the two microphone A and B at a sampling rate of 50 kHz. Using current two channel data, the temperature can be estimated by:

$$T = \frac{k^2 d_1^2}{403 f_1^2},\qquad(10)$$

where d_1 is the ground truth distance between A and B, f_1 is the estimated frequency via DFT analysis. We keep A fixed, move B 52.5 mm away, and record another set of audio signals. Now we can get:

$$T = \frac{k^2 d_2^2}{403 f_2^2},\qquad(11)$$

The ground truth distance d_1 and d_2 are non-trivial to obtain since a microphone sensor has centimeter size. However, their distance can be precisely obtained in our platform as $\Delta d = d_2 - d_1 = 52.5$ mm. Combining Eq. 10, 11 with the knowledge of k, f_1, f_2, and $\Delta d = d_2 - d_1$, the temperature can be finally computed.

Fig. 4 depicts the estimated sound speed using our testbed. The mean value of sound speed is 334.75 m/s, indicating a temperature of $5.5\,^{\circ}$C. The reading from the thermometer is $5.1\,^{\circ}$C. Therefore, the estimation error is only $0.4\,^{\circ}$C. Considering the measurement noise, the estimation results successfully demonstrate the feasibility of our proposal.

5 MITIGATING CHANNEL DISTORTION AND NOISE

In this section, we present our channel distortion mitigation techniques, namely, wavelet denoising, compensation, and RNSAC filtering. Apart from common background noise, the acoustic signals suffer from severe channel distortions. Without further distortion mitigation, using these unfiltered measurements cannot achieve satisfactory results. One major distortion called non-flat frequency responses or frequency selectivity, wherein acoustic sensors have different channel gains at different frequency bands, can lead to unstable measurements and large variances in temperature estimation [10]. Another one is called device diversity. Device diversity describes a phenomenon where different devices may exhibit different properties. SST requires two microphones to capture the transmitted signals. The two microphones on commodity hardware especially smartphones typically have different sensitivity. The microphone on the back often has higher channel gains at high frequency, which is used to remove background noise. The bottom one, which is used to record human voices, has higher channel gains under 8 kHz [33], [14]. The device diversity can also result in large measurement variance. To understand these distortions, we first present a measurement study.

5.1 Measurement Study

To characterize the degree and understand sources of frequency selectivity and device diversity, we perform measurement studies on a smartphone device. A comprehensive in-depth understanding of channel distortion can serve as guidelines to optimize system designs. For instance, we may disable parts of the hardware that introduce severe distortions and thereby improve system performance.

To begin with, we introduce the typical hardware diagram for acoustic signal processing. Fig. 5 (a) and (b) depict the typical pipelines of a sound recording and emitting system [33], [34], respectively. A sound recording system converts mechanical signals into digital samples while a sound emitting system reverses this process. In a recording system, sound signals are first converted into voltage signals by a microphone. An Automatic Gain Control (AGC) circuit or Programmable Gain Amplifier (PGA) then amplifies the voltage signals to surpass the quantization level of the posterior Analog-to-Digital Converter (ADC). Next, the amplified signals pass through a Low Pass Filter (LPF), also known as an anti-aliasing filter, and become band-limited signals. The cut-off frequency of the LPF is $f_s/2$, where f_s is the sampling rate. The filtered signals go through a buffer and finally become digital samples via ADC conversion. A sound emitting system reverses the above process via a different circuit. Digital samples are first interpolated, then fed into a Digital-to-Analog Converter (DAC) and
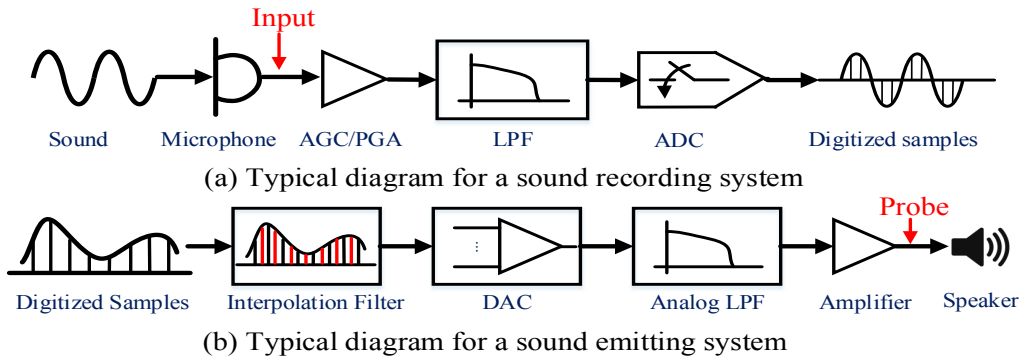
(a) Typical diagram for a sound recording system

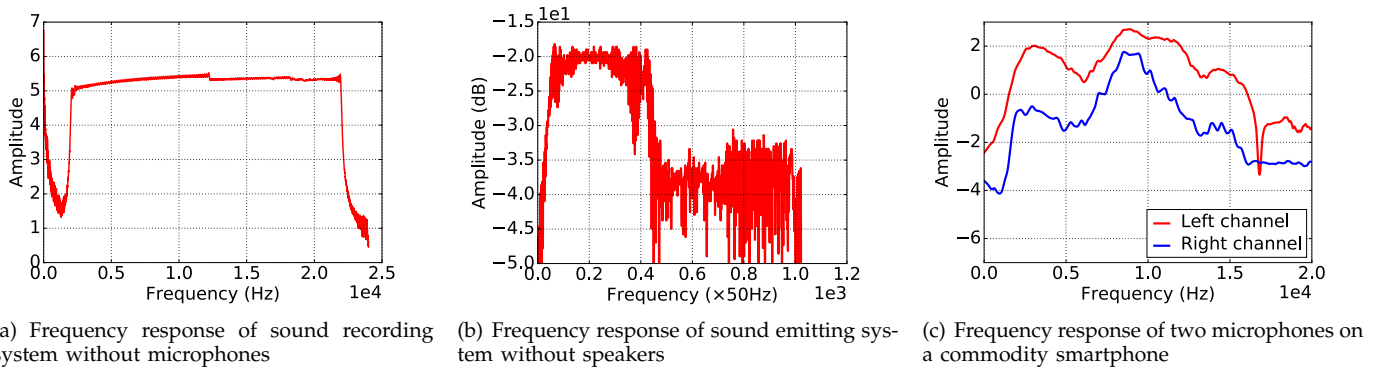(b) Typical diagram for a sound emitting system

Fig. 5: Hardware diagram of sound recording and emitting system

(a) Frequency response of sound recording system without microphones

(b) Frequency response of sound emitting system without speakers

(c) Frequency response of two microphones on a commodity smartphone

Fig. 6: Measurements on commodity hardware: (a) Frequency response of sound recording system without microphones, (b) Frequency response of sound emitting system without speakers, (c) The two microphones on a commodity smartphone exhibit different frequency response.

become analog signals, which after amplification are finally converted into sound waves by a speaker.

We have modified the hardware of a commodity s- martphone (Samsung Galaxy S5) to inspect where channel distortions are introduced. First, we remove the connections between a microphone and its posterior circuit. We use RIGOL DG4162, a 160 MHz arbitrary waveform generator, to synthesize a chirp signal with $f_{min} = 2$ kHz, $B = 20$ kHz, $D = 0.01$ s and connect the output of this signal generator to the input of AGC as depicted in Fig. 5 (a). The sampling rate is set to 48 kHz and sufficient measurements are collected for analysis. Then, we synthesize a wave audio file that contains the chirp signal in the previous experiment, play the audio files through the sound emitting system, and use Tektronix TDS 2024C, a 200 MHz high-performance oscillograph, to probe the frequency response before the speaker as depicted in Fig. 5 (b). Fig. 6(a) and Fig. 6(b) depict the frequency response of the remaining sound recording system without the microphone and the sound emitting system without the speaker, respectively. From the figure we observe that the frequency responses in absence of microphones and speaker are quite flat, indicating that non-flat frequency responses are primarily introduced by the acoustic sensors. The results reveal that we cannot expect to lower channel distortions by simply disabling the internal hardware modules.

Next, we let one smartphone transmit a chirp signal continuously and use two microphones on another smartphone to record the signal. Fig. 6(c) shows the frequency response of the two microphones. Clearly, there are different gains and features. Both the non-flat frequency response of microphones and device diversity would introduce distortions on measurement results, leading to degraded system performance. In the next section, we will show how to mitigate these artifacts.

5.2 Channel Distortion Compensation

To compensate channel distortion, we propose a signal processing pipeline consisting of wavelet denoising, compensation, and RANSAC filtering. Existing work addresses the frequency selectivity problem via a compensation filter whose frequency response is reciprocal of that of the acoustic channels [10]. However, such a solution is not applicable in SST since the employed signals occupy a wide bandwidth. A small variation in the severely attenuated frequency band can cause significant noise if compensation is performed directly.

5.2.1 Wavelet Denoising

Wavelet filter [35] in SST is used to filter in-band noise. Wavelet filters are well-known for their ability

(a) Raw signals   (b) After wavelet filtering   (c) After compensation   (d) After RANSAC filtering
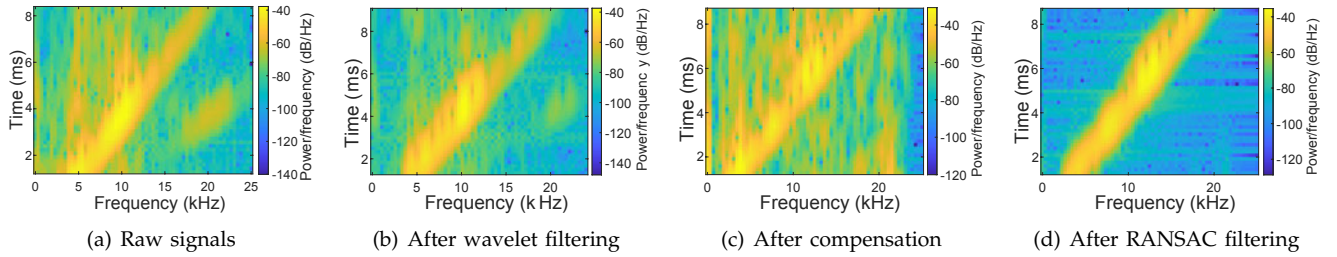
Fig. 7: Spectrogram of signals in different processing stages: (a) raw signals (b) after wavelet filtering (c) after compensation (c) after RANSAC filtering
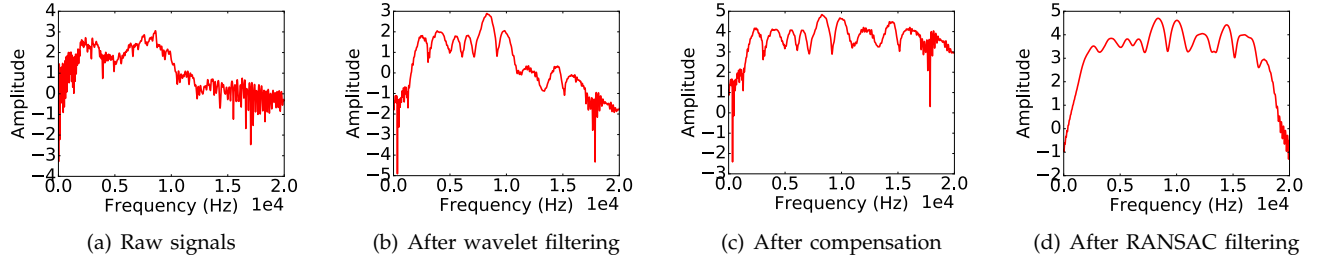

(a) Raw signals   (b) After wavelet filtering   (c) After compensation   (d) After RANSAC filtering

Fig. 8: Frequency response of signals in different processing stages: (a) raw signals (b) after wavelet filtering (c) after compensation (d) after RANSAC filtering

to preserve local features and remove outliers. It has many advantages such as low entropy, multi-scale resolution, and design flexibility. However, directly applying wavelet filtering on chirp signals is not appropriate due to their non-smooth properties, making it challenging to set appropriate wavelet coefficients to achieve good performance [36]. However, chirp signals are smoother in frequency domain. Therefore, we apply wavelet filters in frequency domain. In the design of a specific wavelet filter, choosing the appropriate wavelet basis is crucial. To achieve optimal design, several factors should be taken into consideration. In SST, phase information needs to be preserved since it contains critical time information. Therefore, a wavelet filter that has linear-phase property is desired. In addition, it should be compactly supported for fine resolution. Therefore, we choose bi-orthogonal basis of compactly supported wavelets [37].

We utilize wavelet packet shrinkage denoising [38], [39] to filter in-band noises. For a received signal $x(t)$, we first obtain its frequency domain representation $X(f)$ via DFT. We preserve its phase information $\theta(f)$ and extract the amplitude for later processing. Next, the wavelet filter is applied to improve SNR. Then, the phase information $\theta(f)$ is added to the filtered signal. Finally, Inverse Discrete Fourier Transform (IDFT) is used to restore a time-domain signal, $\hat{x}(t)$. The comparison between Fig. 7(a) and Fig. 7(b) or Fig. 8(a) and Fig. 8(b) demonstrate the effectiveness of wavelet filter. As depicted in Fig. 7(b), the intensity of the out-band noises indicated by the yellow color are much lighter than that of the original signals shown in Fig. 7(a). Meanwhile, the ripples in the original signals presented in Fig. 8(a) are attenuated after wavelet filtering as depicted in Fig. 8(b),

indicating less noises.

Since $\hat{x}(t)$ has less noise, we subsequently compensate $\hat{x}(t)$ to address the frequency selectivity problem. It should be noted that solutions to the frequency selectivity problem can also handle the device diversity problem. For simplicity, we only discuss the frequency selectivity hereafter.

5.2.2 Compensation

The compensation step is critical to address the frequency selectivity problem. As discussed earlier, frequency selectivity is caused by non-flat channel responses introduced by microphones, and leads to unstable measurements. Let $H(f)$ be the channel response in frequency domain. The goal of compensation is to find a $H'(f)$ that is reciprocal to $H(f)$ so that $H(f)H'(f) = 1, \forall f$. Apparently, we can derive $H'(f)$ by $H'(f) = \frac{1}{H(f)}$. To handle random noises in the severely attenuated bandwidth, we add a small constant σ (default to 0.1) to the denominator in $H'(f) = \frac{1}{H(f)}$. Therefore, the compensation filter has the following frequency response:

$$H'(f) = \frac{1}{H(f) + \sigma} \tag{12}$$

The effects of compensation after wavelet filtering are shown in Fig. 7 (c) and Fig. 8 (c). Before compensation, as we can observe from Fig. 7 (b) and Fig. 8 (b), the channel responses above 15 kHz are severely attenuated. However, after compensation, the frequency response is much flatter than before. Therefore, the frequency selectivity problem is mitigated. Nevertheless, such compensation leads to severe noises as depicted in Fig. 7 (c), where the yellow color indicates that the signal intensity apart from
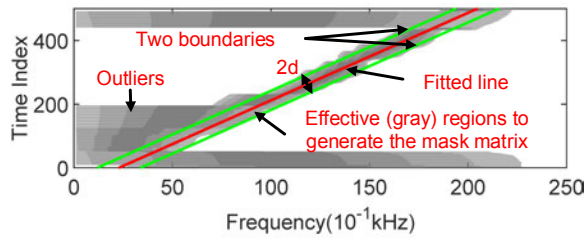
Fig. 9: RANSAC filtering illustration: We use the RANSAC fitted line and two boundaries with distance d to the fitted line to form a region. The samples within the regions are effective samples while outside ones are noisy measurements and should be filtered out.

the chirp region is more intense after compensation. To this end, we further suppress the noises via a Random Sample Consensus (RANSAC) filtering approach.

5.2.3 RANSAC Filtering

RANSAC filtering is used to further suppress in-band noise. RANSAC filtering originates from the RANSAC fitting algorithm [40], which is used to estimate regression parameters from noisy samples with outliers. The intuition behind the proposed RANSAC filtering algorithm is that we can use RANSAC fitting to identify the main regions of chirp signals in the spectrogram and set the remaining regions in the spectrogram to a minimal value. Doing so removes undesirable in-band noises. The pseudo-algorithm of the proposed RANSAC filtering is summarized in Algorithm 1.

In the implementation, we set the threshold value $p = 1$, the number of iteration as 10000, the inlier distance threshold as 10; and the inlier number threshold as $0.5 * N$, respectively, where N is the number of the input samples.

Algorithm 1 RANSAC filtering algorithm

Data: Time series acoustic samples s
Result: Filtered signals $\hat{s}$

1: Perform STFT on s and preserve the corresponding amplitude A and phase θ information;
2: Binarize the spectrogram via a threshold p, $A\left(A > p\right) = 1$;
3: Perform RANSAC fitting on the binarized spectrogram;
4: Define two boundaries that are parallel to the fitted line with a boundary-line distance d (see Fig. 8);
5: Minimize the values outside the boundaries;
6: Utilize the values within the boundaries to form a matrix mask M_{m};
7: Multiply M_{m} by the original spectrogram formed by A and θ, and get the filtered spectrogram;
8: Perform ISTFT on the filtered spectrogram to obtain time-series $\hat{s}$.

Fig. 7(d) and Fig. 8(d) demonstrate the effectiveness of the RANSAC filtering algorithm. From Fig. 7(d) we
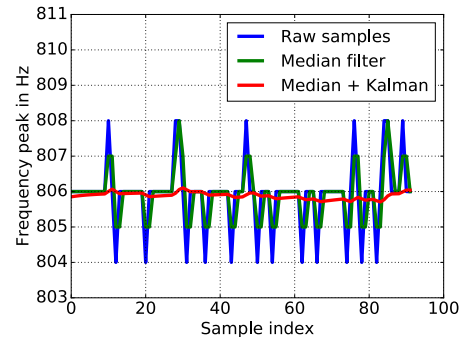

Fig. 10: The effect of filters on the raw frequency peaks over a time window. The maximum deviation of the frequency peak before filters can be over 8 Hz. However, after median and Kalman filter, the maximum deviation is within 0.5 Hz.

can see that, the in-band noise (depicted as yellow regions apart from the desirable chirp region in Fig. 7 (c)) after RANSAC filtering is greatly reduced. We can also observe the effectiveness of the filter from Fig. 8 (d) and Fig. 8 (c). Compared with Fig. 8 (c), the frequency response depicted in Fig. 8 (d) has fewer ripples, indicating much weaker noises.

5.3 Smoothing

After compensation and filtering, the signals from the two channels can finally be applied to perform mixing. Following that, we find the peak frequency component f from the mixed result and use Eq. (8) to obtain the temperature results. To further smooth the estimated temperature results over a time window, the raw estimates of frequency peaks are filtered by a median filter and a Kalman filter in tandem. The filter performance can be observed in Fig. 10. As shown in Fig. 10, the maximum deviation of the frequency peak before filters can reach 8 Hz. However, after median and Kalman filter, the maximum deviation is within 0.5 Hz.

6 PERFORMANCE EVALUATION

6.1 Implementation

We have implemented SST on a commodity smartphone, i.e., Samsung Galaxy S5, and a custom-built platform. The customized platform has three modules: the sensor, codec, and Raspberry Pi modules. The customized platform was previously designed for acoustic based localization in indoor environments [41] and can be thought of as an example acoustic-enabled IoT devices. We mount an INMP411 high-performance acoustic sensor on the sensor board. INMP411 is chosen due to its optimized performance in high frequency range. The amplifying circuit on the sensor board is transistor-based and there is no AGC. The codec board mounts a WM8731 codec chip that has dual channel inputs with a maximum of 96 kHz sampling rate. It also has an on-board TDA2822M,
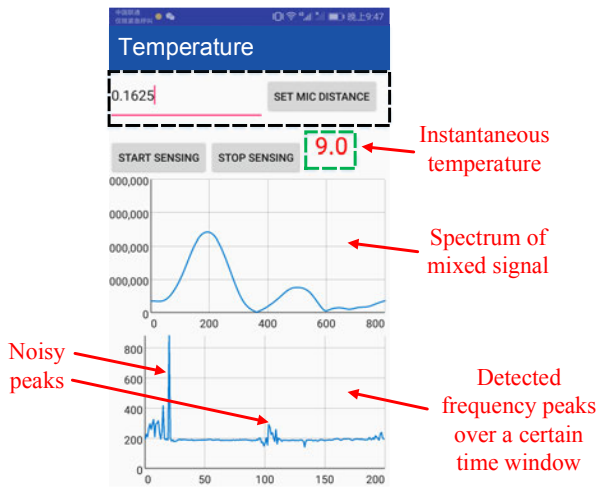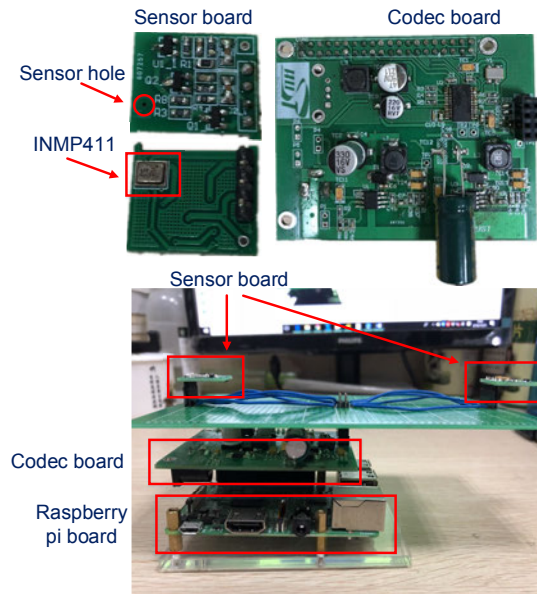
Fig. 11: Snapshot of the Android application

Fig. 12: Platform two: A high-end customized acoustic sensing platform that supports two channel inputs with a maximum of 96 kHz sampling rate, and stereo output up to 3W.

a 3W classic-AB audio power amplifier. The employed chirp signal has 20 kHz bandwidth and 0.01 s duration with an initial frequency of 2 kHz. The guard interval between successive transmission is 0.03 s and the sampling rate is set to 48 kHz. The detailed settings are shown in Table. 1.

SST works in two modes: stand-alone and client-

TABLE 1: Settings for the transmitted signals

Parameters	Value
Sampling rate	48 kHz
Chirp bandwidth	20 kHz
Chirp duration	0.01 s
Initial frequency	2 kHz
Guard interval	0.03 s

(a) Pi testbed　　　　　(b) Phone testbed

Fig. 13: Testbed setup

server. On the Android platform, we execute a stand-alone app. In the app, most of the complex signal processing modules are implemented via NDK APIs. The final results are visualized using an open-source graphic toolkit [42]. A snapshot of the Android app is shown in Fig. 11. With this app, one can preset the microphone distance, obtain the instantaneous temperature data, and inspect the intermediate results including the mixed spectrum and the estimated frequency peaks over a time window.

A client-server application was developed for the customized platform. We use MATLAB to synthesize an audio file that contains the designed chirp signals. The Pi plays this audio files, generating expected wide-band chirp signal to be played back. The signals recorded by the microphone modules are then streamed to a server using a UDP socket. The server processes the captured frames in real-time and performs temperature estimation. To measure ground truth values, we use an arduino-compatible module with dedicated sensors named GY-213V-SI7021. It can achieve 0.05 °C and 2 %RH accuracy in measuring temperature and humidity, respectively.

6.2 Results

We conduct extensive measurements to evaluate the performance of SST.

6.2.1 Stability of SST

Stability indicates whether the thermometer outputs stable readings over a certain time window when the temperature and humidity do not have dramatic changes. Unstable measurements prohibit the practical use of a thermometer since a user cannot obtain reliable readings. We use the maximum deviation of the estimated temperature values or frequency peaks to quantify the stability. Clearly, a small deviation is much desirable.

We conduct experiments using the customized platform and smartphones shown in Fig. 13. Fig. 14(a) depicts the results on the customized platform. It can be observed that the deviation of the frequency peaks is within 1 Hz and the maximum deviation is 0.6583 Hz. This is equivalent to a temperature deviation around 0.26 °C. Fig. 14(b) depicts the results on the smartphone. The results reveal a maximum frequency deviation of 1.62

(a) Stability test on a customized platform   (b) Stability test on a commodity smartphone   (c) AGC incurs dynamic channel gains
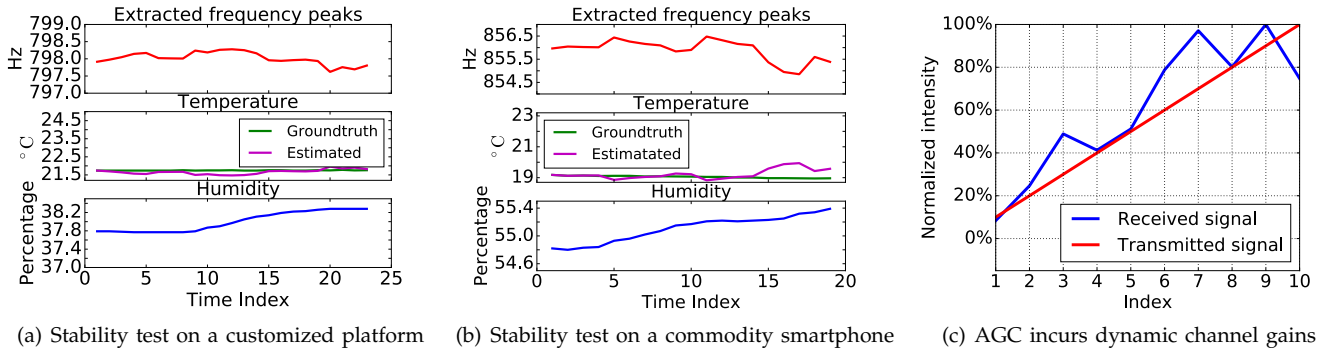
Fig. 14: Stability evaluation on (a) a customized platform and (b) a smartphone. During the measurement period, both temperature and humidity remain stable. The results on the customized platform is much better than that on the smartphone. Fig. (c) Demonstrates that the channel gain exhibits variations due to AGC

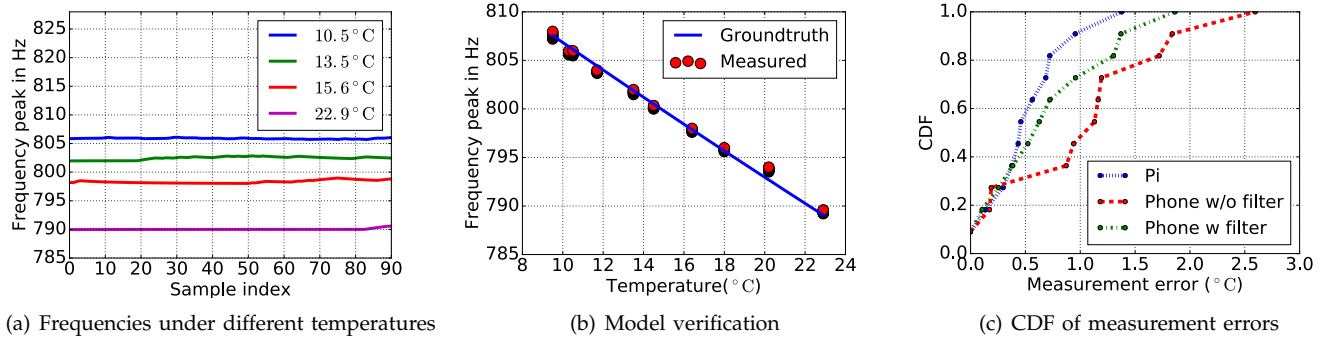(a) Frequencies under different temperatures   (b) Model verification   (c) CDF of measurement errors

Fig. 15: Evaluation results: (a) Measured frequency peaks under different temperatures, demonstrating the feasibility of the proposed model (b) Model verification. The blue line indicates the theoretical relation between temperature and frequency peaks. The red dots are measured frequency peaks at discrete temperature points. (c) CDF of the measurement results under different configurations and platforms.

Hz which is not as good compared to the customized platform but still provides acceptable temperature deviations (around 1.1 °C. The degraded performance on the smartphone are due to two major contributing factors. First, the acoustic signals experience non-line-of-sight path to reach two microphones. The NLOS signals are comparatively weaker and more unstable than that of line-of-sight (LOS) signals [27], [10]. This NLOS problem, caused by the physical layout of the two microphones, is inherent in a smartphone[3]. Second, a smartphone has an on-board AGC (depicted in Fig. 5) that makes the acoustic channel exhibit dynamic channel gains, resulting in degraded compensation effects and thus worse thermal sensing performance. Fig. 14(c) illustrates this phenomenon. As we can observe in Fig. 14(c), as the intensity of transmitted signals increases linearly, the intensity of the received signals exhibits over 20% variations. AGC cannot be disabled through Android API [43]. As part of future work, we will investigate native implementation that can disable AGC to improve the stability of SST on smartphones.

---

3. A common microphone placement on a smartphone is: one microphone is at the bottom and the other is on the top. Therefore, it is impossible for both microphones to be at LOS from the auxilary device.

6.2.2 Accuracy of SST

To evaluate the temperature estimation accuracy of SST, we conduct the experiments in a car. We use the air conditioner in the car to change the temperature. The enclosed environment makes it relatively easy to change the temperature in a short period of time. The achievable range of the temperature is 9 - 30 °C. We randomly select one temperature point to calibrate d and then use Eq. (8) to perform estimation at other temperature points.

Fig. 15(a) depicts the measured frequency peaks over different temperatures. Evidently, different temperatures result in different frequency peaks. Fig. 15(b) plots frequency peaks vs temperature. The blue line plots the theoretical relationship between the temperature and frequency peak, and the red points are measured frequency peaks at discrete temperature points. We can clearly observe that the measurements are in good agreement with the theoretical ones, verifying the correctness of the proposed models. Fig. 15(c) depicts the CDF of the estimated temperature. We can observe that SST can achieve a median accuracy of 0.5 °C and a 80-percentile of 0.7 °C on our Pi platform. On smartphones, SST achieves a median accuracy of 0.6 °C and a 80-percentile of 1.2 °C. Without the proposed compensation filter, the 80-percentile error of could reach 1.4 °C.
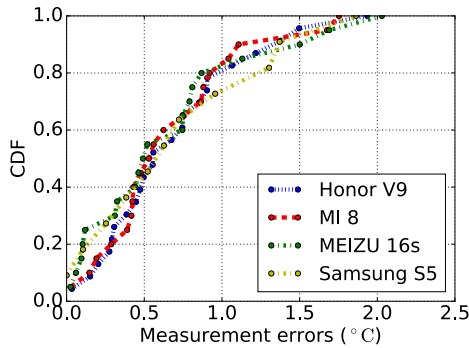
Fig. 16: Measurement results of different phone models

Fig. 17: Testbed setup to explore the impact of humidity on the sensing results

We have also conducted experiments using different phone models. The phones used in the experiment include HUAWEI Honor V9, Xiao MI 8, MEIZU 16s, and Samsung S5. The frequency responses of the acoustic sensors on these phones are different. The two microphones on these smartphones are all in opposite directions with one on the top and the other on the bottom (common microphone placement on smartphones). However, the distances between the microphones and their relative orientations across different smartphones are different. We use another HUAWEI honor phone as the auxiliary device. The measurement device and the auxiliary device are placed along the long side. Though the relative orientation of the microphones to the speaker are different, the calibrated effective microphone-speaker distances are much the same (around 0.15 m), indicating the same measurement precision. The results are shown in Fig. 16. Apparently, the sensing results across different phones are almost the same.

6.2.3 Effect of Humidity

Humidity is known to affect sound speed in air but in practice has negligible impacts on the estimated temperature. To verify the above claim, we conduct another experiment. The testbed setup is the same with the one in the stability evaluation experiment and is shown in Fig. 17. We use a smartphone to play the designed chirp signals. We plot the frequency peaks, temperature, and humidity in real-time, respectively. At some point during the experiment, we use a water spay to change the humidity of the acoustic channel. The vapor produced by the water spay is fine-tuned so that it only changes the humidity but not the temperature.

Fig. 18(a) depicts the estimated frequency peaks, the estimated temperature, and measured humidity over a certain time window, respectively. From Fig. 18(a) we can observe that, when the humidity experiences a dramatic change, increasing from a minimum of 42.6 %RH to a maximum of 90.99 % RH, the frequency peaks only undergo a change less than 0.89 Hz, or equivalently a change in temperature estimation within 0.5 °C. Given the inherent measurement noises, it is reasonable to conclude that humidity has negligible impacts on the
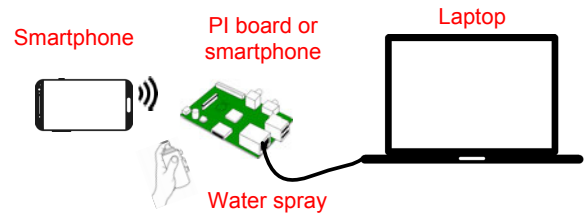
sensing results. This can be further validated using the Pearson Product-Moment Correlation Coefficient (PPM-CC), a well-known statistical approach to measure correlation between data sets, to quantify the relation between the humidity and frequency peaks. The PPMCC [44] is defined by,

$$\rho_{X,Y} = \frac{\text{cov}\,(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X - \mu_X)\,(Y - \mu_Y)\right]}{\sigma_X \sigma_Y}, \quad (13)$$

where μ_X, μ_Y and σ_X, σ_Y are mean and variance, respectively. $cov\,()$ calculates the covariance between two measurements. The PPMCC is within a range between $[-1, 1]$ and if PPMCC is close to zero, two measurements have loose correlation, otherwise strong correlation. In the experiments, the correlation between the frequency peaks and the humidity is $\rho = -0.023963$. In other words, probably the fluctuation in frequency peaks is due to different noisy sources other than humidity.

6.2.4 Usability Study

We have conducted a usability study to evaluate the applicability of SST. In this experiment, we ask four volunteers to use SST under real-world settings. During the calibration phase, they manually marked the placement of the measurement device and the auxiliary device along a side of a common object such as a desk. Then they were asked to take the devices away and recorded the thermal readings in three different time (morning, noon, and night) of the next day. Each participant was asked to repeat the measurements at least at three different places. It should be noted that the four volunteers took turns in using the same devices for measurements. The results are depicted in Fig. 18(b). It is observable that all the measurements from four participants exhibit the same performance. The average errors are all around 0.6°C. The four participants all were excited about the application and gave positive feedback to it.

7   DISCUSSIONS

7.1   Removing Auxiliary Device

Intuitively, one may wonder whether it is viable to run the proposed algorithm on a single device. The answer is no and the underlying reasons are listed as follows. First, the SNR of recorded signals is much worse on a single device. When performing sensing on a single device, the model proposed in this paper (Eq. (8)) could

(a) Measurements at different humidities   (b) Usability study   (c) Measurements under different volumes
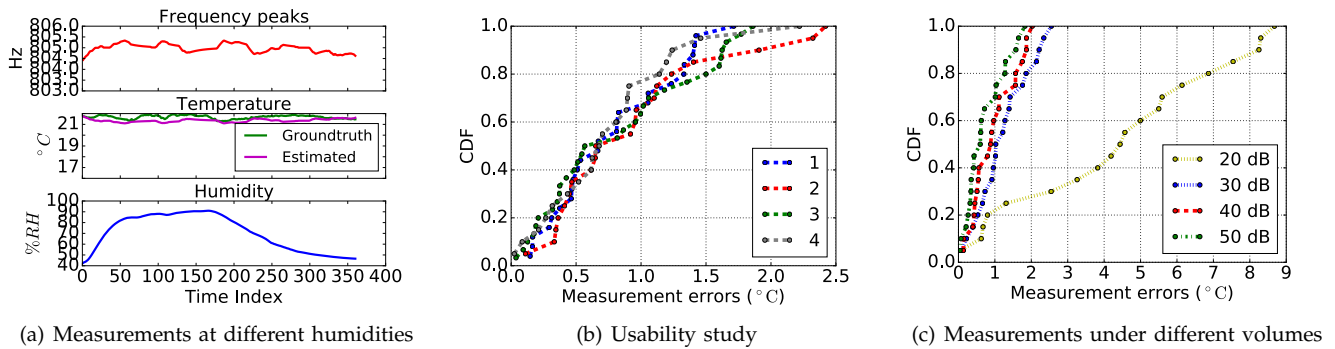
Fig. 18: Evaluation results: (a) A dramatic change in humidity does not affect the estimated sensing results. (b) Usability study. The measurement results from different users reveal the same performance. (c) The measurement results at different transmission volumes.

be hardly verified due to severely signal distortions. Second, the effective microphone distance d decreases sharply. For instance, on Samsung S5, the effective microphone distance almost cuts down by half. Therefore, the corresponding measurement accuracy decays up to $1°C$, which affects its usability. Finally, the sensed spectrum peak is very unstable on a single device, resulting in unstable thermal readings.

However, we believe that the auxiliary device may not be needed in certain conditions. For instance, recent advances in acoustic sensing have shown capabilities in decomposing acoustic reflections [10], [30] and even use it for acoustic imaging [30]. With advanced signal processing techniques, we can perform thermal sensing using acoustic reflections on a single commodity mobile devices. For example, we can place a smartphone in proximity to a concrete wall and inspect the temperature empirically from a specific acoustic reflection. Under such circumstances, the auxiliary device can be removed.

7.2 Effect of Wind Velocity

SST utilizes the relation between temperature and the sound speed in thermal sensing. It is desirable that external wind force is minimal, or otherwise the estimated results would be greatly impacted. To mitigate the impact of wind velocity, we can place a cover on the acoustic channels or add a wind velocity sensor to compensate the wind force effects.

7.3 Audibility

SST needs a wide-band chirp signal that occupies the frequency range from 2 - 22 kHz. Such chirp signals are audible to human ears and can be disturbing. However, since obtaining a thermal reading only takes less than a second, audibility is not a severe issue. The technique of reshaping the signal amplitude in [23] can mitigate the audibility problem but may aggravate the frequency selectivity problem hence distort thermal sensing performance. Another feasible approach to further deal with the audibility is to reduce the volume on the transmission side. This operation would reduce the SNR and

thereby affect the sensing results. To explore how the transmission volume affects the sensing results, we have conducted experiments under different transmission volumes. In the experiment, we use an APP called Sound Meter [45] to measure the average sound intensity. The results are presented in Fig. 18(c).

It can be observed that reducing the volume from 50 dB to 30 dB on the transmission side would slightly decay the system performance. And at 20 dB, the performance become worse with an average accuracy over $4°C$. The degraded performance at 20 dB may be caused by the fact that the transmitted signals are under noise floor or below the quantization level of the receiver end. At 50 dB, the average sensing results could reach $0.6°C$. It should be noted that 50 dB is the sound intensity of common indoor office and 30 dB is in a quite home [46]. A noisy environment should reach 70 dB [46]. Consequently, we believe our proposed method introduce neglectable interference to human life.

8 CONCLUSIONS

In this paper, we have proposed the first Software Sonic Thermometer (SST) using acoustic-enabled IoT devices. We have employed on-board dual microphones on commodity hardware to achieve accurate thermal sensing in a non-contact manner by leveraging the deterministic relation between sound speed and temperature data. To precisely measure sound speed, we have presented a chirp signal mixing technique to circumvent low sampling rates on commodity hardware and designed a pipeline of signal processing blocks to handle channel distortion. Extensive measurements have revealed that SST yields a median accuracy of 0.5 °C and 90% estimation errors below 0.9 °C. With the recent advances in crowdsourcing, SST could leverage the ubiquitous IoT devices to achieve large-scale distributed thermal sensing, which may revolutionize temperature acquisition in meteorology and urban computing.

REFERENCES

[1] (2018) World climate maps. https://www.climate-charts.com/World-Climate-Maps.html.

[2] R. Mendelsohn, W. D. Nordhaus, and D. Shaw, "The impact of global warming on agriculture: A ricardian analysis," *The American Economic Review*, vol. 84, no. 4, pp. 753–771, 1994.

[3] M. Hu, W. Liu, J. Lu, R. Fu, K. Peng, X. Ma, and J. Liu, "On the joint design of routing and scheduling for vehicle-assisted multi-uav inspection," *Future Generation Computer Systems*, vol. 94, pp. 214–223, 2019.

[4] Collaboration, *Observations: Surface and Atmospheric Climate Change*, ser. IPCC Fourth Assessment Report: Climate Change 2007. Working Group I: The Physical Science Basis. Cambridge University Press, 2007, pp. 235–336.

[5] M. Chen, J. Zhou, G. Tao, J. Yang, and L. Hu, "Wearable affective robot," *IEEE Access*, vol. 6, pp. 64766–64776, 2018.

[6] W. Zhou, B. Peng, J. Shi, T. Wang, Y. P. Dhital, R. Yao, Y. Yu, Z. Lei, and R. Zhao, "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*, vol. 9, no. 9, 2017.

[7] (2018) Design considerations for measuring ambient air temperature. http://www.ti.com/lit/an/snoa966b/snoa966b.pdf.

[8] (2018) Temperature measurement theory and practical techniques. https://www.analog.com/media/en/technical-documentation/application-notes/an_892.pdf.

[9] "Number of mobile phone users worldwide from 2015 to 2020 (in billions)," https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/, 2019.

[10] 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, 2017.

[11] J. Hao, Y. Yang, and J. Luo, "Ceilingcast: Energy efficient and location-bound broadcast through led-camera communication," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016.

[12] M. Chen, Y. Hao, C.-F. Lai, D. Wu, Y. Li, and K. Hwang, "Opportunistic task scheduling over co-located clouds in mobile environment," *IEEE Transactions on Services Computing*, vol. 11, no. 3, pp. 549–561, 2018.

[13] M. Hu, W. Liu, K. Peng, X. Ma, W. Cheng, J. Liu, and B. Li, "Joint routing and scheduling for vehicle-assisted multi-drone surveillance," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1781–1790, 2019.

[14] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "Batmapper: Acoustic sensing based indoor floor plan construction using smartphones," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '17, 2017.

[15] "Temperature Sensors: Contact or Noncontact?" https://www.sensorsmag.com/components/temperature-sensors-contact-or-noncontact, 2018.

[16] M. Anderson, J. Norman, G. Diak, W. Kustas, and J. Mecikalski, "A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing," *Remote Sensing of Environment*, vol. 60, no. 2, pp. 195 – 216, 1997.

[17] C. P. Lo, D. A. Quattrochi, and J. C. Luvall, "Application of high-resolution thermal infrared remote sensing and gis to assess the urban heat island effect," *International Journal of Remote Sensing*, vol. 18, no. 2, pp. 287–304, 1997.

[18] J. Mandal, S. Pal, T. Sun, K. T. V. Grattan, A. T. Augousti, and S. A. Wade, "Bragg grating-based fiber-optic laser probe for temperature sensing," *IEEE Photonics Technology Letters*, vol. 16, no. 1, pp. 218–220, Jan 2004.

[19] G. Shaker, M. Tentzeris, and S. Safavi-Naeini, "Low-cost antennas for mm-wave sensing applications using inkjet printing of silver nano-particles on liquid crystal polymers," in *2010 IEEE Antennas and Propagation Society International Symposium*, July 2010, pp. 1–4.

[20] J. C. Kaimal and J. E. Gaynor, "Another look at sonic thermometry," *Boundary-Layer Meteorology*, vol. 56, no. 4, pp. 401–410, 1991.

[21] "Sonic anemometry and thermometry: theoretical basis and data-processing software," *Environmental Software*, vol. 11, no. 4, pp. 259 – 270, 1996.

[22] I. Dokmani and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[23] D. B. Haddad, M. V. S. Lima, W. A. Martins, L. W. P. Biscainho, L. O. Nunes, and B. Lee, "Acoustic sensor self-localization: Models and recent results," *Wireless Communications and Mobile Computing*, vol. 2017, no. 7972146, October 2017.

[24] C. Cai, R. Zheng, J. Li, L. Zhu, H. Pu, and M. Hu, "Asynchronous acoustic localization and tracking for mobile targets," *IEEE Internet of Things Journal*, 2019.

[25] W. Mao, J. He, H. Zheng, Z. Zhang, and L. Qiu, "High-precision acoustic motion tracking: Demo," in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '16, 2016.

[26] C. Cai, M. Hu, D. Cao, X. Ma, Q. Li, and J. Liu, "Self-deployable indoor localization with acoustic-enabled iot devices exploiting participatory sensing," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5297–5311, 2019.

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

[28] C. Cai, X. Ma, M. Hu, Y. Yang, Z. Li, and J. Liu, "Sap: A novel stationary peers assisted indoor positioning system," *IEEE Access*, vol. 6, pp. 76475–76489, 2018.

[29] S. Ke, Z. Ting, W. Wei, and X. Lei, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18, 2018.

[30] W. Mao, M. Wang, and L. Qiu, "Aim: Acoustic imaging on a mobile," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '18, 2018.

[31] C. Cai, M. Hu, X. Ma, K. Peng, and J. Liu, "Accurate ranging on acoustic-enabled iot devices," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3164–3174, 2019.

[32] Y.-C. Tung and K. G. Shin, "Echotag: Accurate infrastructure-free indoor location tagging with smartphones," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15, 2015.

[33] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '17, 2017.

[34] "TI Smartphone Solutions," http://www.ti.com/lit/sl/slyy032a/slyy032a.pdf, 2017.

[35] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1053–1060, Aug 1995.

[36] L. Deng and J. G. Harris, "Wavelet denoising of chirp-like signals in the fourier domain," in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems VLSI (ISCAS'99)*, 1999.

[37] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 17, pp. 485–560, 1992.

[38] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?" *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 2, pp. 301–369, 1995.

[39] R. R. Coifman and D. L. Donoho, *Translation-Invariant De-Noising*, 1995, pp. 125–150.

[40] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[41] Y. Wang, J. Li, R. Zheng, and D. Zhao, "Arabis: an asynchronous acoustic indoor positioning system for mobile devices," in *IEEE International Conference on Indoor Positioning and Indoor Navigation*, 2017.

[42] (2018) Graphview - open source graph plotting library for android. http://www.android-graphview.org/.

[43] Y.-C. Tung, D. Bui, and K. G. Shin, "Cross-platform support for rapid development of mobile acoustic sensing applications," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '18, 2018.

[44] A. Gayen, "The frequency distribution of the product moment correlation coefficient in random samples of any size draw from non-normal universes," in *Biometrika*, vol. 38, 1951, pp. 219 – 247.

[45] (2019) Sound meter. https://www.analog.com/media/en/technical-documentation/application-notes/an_892.pdf.

[46] (2019) Sound intensity and sound level. https://courses.lumenlearning.com/physics/chapter/17-3-sound-intensity-and-sound-level/.
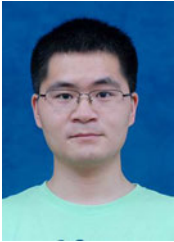
Chao Cai is a Ph.D student with the School of Electronic Information and Engineering at Huazhong University of Science and Technology, Wuhan, China. His current research interests include mobile computing, acoustic sensing, wireless sensing, embedded system, digital signal processing, and deep learning.

Jun Luo received his BS and MS degrees in Electrical Engineering from Tsinghua University, China, and the Ph.D. degree in Computer Science from EPFL (Swiss Federal Institute of Technology in Lausanne), Lausanne, Switzerland. From 2006 to 2008, he has worked as a postdoctoral research fellow in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. In 2008, he joined the faculty of the School Of Computer Science and Engineering, Nanyang Technological University in Singapore, where he is currently an Associate Professor. His research interests include mobile and pervasive computing, wireless networking, applied operations research, as well as network security. More information can be found at http://www.ntu.edu.sg/home/junluo.

Henglin Pu is an undergraduate student at Huazhong University of Science and Technology. His current research interests include acoustic sensing, wireless sensing, and digital signal processing.

Menglan Hu received the BE degree in electronic and information engineering from Huazhong University of Science and Technology, China, in 2007, and the PhD degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2012. He is currently an Associate Professor at the School of Electronic Information and Communications, Huazhong University of Science and Technology, China. His research interests includes cloud computing, mobile computing, parallel and distributed systems, scheduling and resource management, as well as wireless networking.

Rong Zheng received her Ph.D. degree from Dept. of Computer Science, University of Illinois at Urbana-Champaign and earned her M.E. and B.E. in Electrical Engineering from Tsinghua University, P.R. China. She is now a Professor in the Dept. of Computing and Software in McMaster University, Canada. Between 2004 and 2012, she is on the faculty of the Department of Computer Science, University of Houston. Rong Zheng's research interests include wireless networking, mobile computing and machine learning. She serves as an editor of IEEE Transactions on Mobile Computing and IEEE Transactions on Network Science and Engineering, and IEEE Transactions of Wireless Communication. Rong Zheng received Discovery Accelerator Supplement from Natural Science and Engineering Council, Canada in 2019 and the National Science Foundation CAREER Award in 2006. She was Joseph Ip Engineering fellow from 2015 to 2018.