Smartphone Sonar-Based Contact-Free Respiration Rate Monitoring

XUYU WANG, California State University, Sacramento, CA, USA
RUNZE HUANG, Tongxin Microelectronics Co., Ltd., Beijing, China
CHAO YANG and SHIWEN MAO, Auburn University, Auburn, AL, USA

Vital sign (e.g., respiration rate) monitoring has become increasingly more important because it offers useful clues about medical conditions such as sleep disorders. There is a compelling need for technologies that enable contact-free and easy deployment of vital sign monitoring over an extended period of time for healthcare. In this article, we present a SonarBeat system to leverage a phase-based active sonar to monitor respiration rates with smartphones. We provide a sonar phase analysis and discuss the technical challenges for respiration rate estimation utilizing an inaudible sound signal. Moreover, we design and implement the SonarBeat system, with components including signal generation, data extraction, received signal preprocessing, and breathing rate estimation with Android smartphones. Our extensive experimental results validate the superior performance of SonarBeat in different indoor environment settings.

CCS Concepts: • **Applied computing** → **Health informatics**; • **Human-centered computing** → **Smartphones**;

Additional Key Words and Phrases: Channel state information, health sensing, healthcare Internet of Things (IoT), respiration rate monitoring, smartphone apps

ACM Reference format:
Xuyu Wang, Runze Huang, Chao Yang, and Shiwen Mao. 2021. Smartphone Sonar-Based Contact-Free Respiration Rate Monitoring. *ACM Trans. Comput. Healthcare* 2, 2, Article 15 (February 2021), 26 pages.
https://doi.org/10.1145/3436822

1 INTRODUCTION

With the rapid development of mobile techniques and the growth in living standards, healthcare has become one of the main application areas for the Internet of Things (IoT) [Wang et al. 2017a; Islam et al. 2015]. The healthcare IoT architecture mainly consists of three layers: (1) the sensing layer for monitoring vital signals, such as body temperature, heart rate, respiration rate, and blood pressure; (2) the gateway layer for collecting data from the sensing layer and transmitting them to the third layer, the cloud layer [Cicalo et al. 2016]; and (3) the cloud layer consisting of data centers in the cloud to store, process, and analyze multimodal medical datasets

This work is supported in part by the NSF under Grant ECCS-1923163, and through the Wireless Engineering Research and Education Center (WEREC) at Auburn University. This work was conducted when R. Huang was pursuing an MS degree in ECE and X. Wang was pursuing a PhD degree in ECE at Auburn University. This work was presented in part at the 26th International Conference on Computer Communications and Networks (ICCCN 2017), Vancouver, Canada, July/Aug. 2017 [Wang et al. 2017a].
Authors' addresses: X. Wang, 6000 J Street, Sacramento, CA 95819 USA; R. Huang, No. 3 A DaTun Rd, ChaoYang Dist, Beijing, China 10010; C. Yang and S. Mao, 200 Broun Hall, Auburn University, Auburn, AL 36849-5201 USA.

2637-8051/2021/02-ART15 $15.00
https://doi.org/10.1145/3436822

[Moon and Lee 2017; Müeller and Unay 2017; Baytas et al. 2016; Henriquez et al. 2017; Yuan et al. 2016] and deliver the analysis results to medical centers. Particularly, the respiration signal is one of the key vital signs to be collected in the first layer, which is indispensable for physical health monitoring, since such vital signals can offer important information for personal health problems such as sudden infant death syndrome (SIDS) [Hunt and Hauck 2006]. Traditional systems in the sensing layer require a person to wear special devices such as a pulse oximeter [Shariati and Zahedi 2005] or a capnometer [Mogue and Rantala 1988] to monitor breathing rates, which are not convenient for monitoring vital signals for elders and infants, and are hard to be used for an extended period of time. Thus, technologies that can enable contact-free, easy deployment, and long-term vital sign monitoring are highly desirable for healthcare provisioning.

Existing vital signal monitoring systems are mainly focused on radio frequency (RF)-based techniques, which leverage RF signals to capture breathing and heart movements [Lubecke et al. 2000]. The existing techniques can be classified into (1) radar-based and (2) WiFi-based approaches. Examples of radar-based vital sign monitoring include Doppler radar [Droitcour et al. 2009; Nguyen et al. 2016; Lubecke et al. 2002; Droitcour et al. 2001], ultra-wideband radar [Salmi and Molisch 2011], and frequency-modulated continuous wave (FMCW) radar [Adib et al. 2015], all of which require a piece of customized hardware working on high frequency. Existing WiFi-based techniques include UbiBreathe [Abdelnasser et al. 2015] and mmVital [Yang et al. 2016], which exploit the received signal strength (RSS) of 2.4 GHz WiFi and 60 GHz millimeter wave (mmWave) signals (i.e., 802.11ad), respectively. Liu et al. [2015a] employ the amplitude of WiFi channel state information (CSI) data to track vital signs for a sleeping person, while our prior works PhaseBeat Wang et al. [2017b], Wang et al. [2020] and TensorBeat [Wang et al. 2017c] exploit the CSI phase difference data for vital sign monitoring for single and multiple persons, respectively. Although RF-based techniques work over a relatively long distance, they could be susceptible to environmental change, such as the movements of other persons nearby.

To this end, the smartphone can serve as an excellent platform for vial sign monitoring by exploiting its built-in sensors, such as the accelerometer, gyroscope [Aly and Youssef 2016], and microphone [Ren et al. 2015]. Usually the smartphone should be placed near the body, or the person needs to wear special types of sensors that connect to the smartphone. The device-free and contact-free monitoring techniques aim to relieve the burden of attached sensors. Nandakumar et al. [2015] propose to use an active sonar built into the smartphone by leveraging the FMCW technique for respiration monitoring. The scheme is shown to work well, but the FMCW-based technique requires an accurate estimation of the distance between the smartphone and the chest. When the body suddenly moves (e.g., rolling over in bed), the system needs to detect the new smartphone-chest distance, thus leading to a large time complexity. Alternatively, the Low-Latency Acoustic Phase (LLAP) system employs continuous-wave (CW) radar to measure distance and achieves device-free hand tracking using sonar phase information [Wang et al. 2016], while Nandakumar et al. [2016] use the phase of acoustic orthogonal frequency-division multiplexing (OFDM) signals for finger tracking.

Motivated by these interesting studies, we employ sonar phase data with a smartphone implementation to monitor the periodic signal caused by the rises and falls of the chest (i.e., inhaling and exhaling). We find that the sonar phase information can effectively track the periodic signal of the breathing rate with high accuracy. Compared with other existing systems such as Doppler shift and FMCW [Nandakumar et al. 2015], the sonar phase-based scheme has a lower latency and complexity. In addition, the sonar phase data is highly robust to different orientations, distances, and respiration rates of different persons.

Specifically, we first present a rigorous sonar phase analysis, which proves that the sonar phase information can accurately capture the respiration rate with the same frequency. Built upon analysis, we design SonarBeat, a smartphone-based active **Sonar** phase for **Breat**hing rate monitoring system. The name also indicates the SonarBeat design on detecting the periodic chest movements (like heartbeat) for respiration rate estimation. The SonarBeat system consists of four modules, including signal generation, data extraction, received signal preprocessing, and breathing rate estimation. First, it transmits an inaudible sound signal in the frequency range of 18 to 22 KHz from the smartphone speaker, which serves as a CW radar. Then, the reflected signal from the

chest of the monitored subject is received by the microphone of the same smartphone. The received signal is then calibrated and the breathing signal will be recovered. We implement SonarBeat with an Android smartphone and validate its performance with extensive experiments that involve five persons over a period of 3 months in three different environments, including an office scenario, a bedroom scenario, and a movie theater scenario. The experimental results show that SonarBeat can achieve a low estimation error for breathing rate estimation, with a medium error of 0.2 breaths per minute (bpm) in most experiments. We also find that SonarBeat is highly robust to different experimental parameters and settings.

The main contributions of this article are as follows:

- —Through analysis and experiments, we validate the feasibility of leveraging the active sonar phase information for breathing rate estimation. To the best of our knowledge, this is the first work to employ active sonar phase information for breathing monitoring with smartphones.
- —We design SonarBeat based on the analysis and address the technical challenges in using the active sonar phase. We implement several signal processing algorithms, including signal generation, data extraction, received signal preprocessing, and breathing rate estimation. Specially, we propose an adaptive median filter approach to remove the static vector in the received signal, which allows to effectively extract the inaudible phase information.
- —We prototype the SonarBeat system with Android smartphones and validate its superior performance by comparing it with an existing scheme in three different indoor scenarios. Our extensive experimental results demonstrate the superior performance of SonarBeat under different environmental factors and different experimental parameters.

In the remainder of this article, Section 2 reviews related work. Then, we present the sonar phase analysis and technical challenges in Section 3. We describe the SonarBeat design in Section 4 and validate its performance in Section 5. Section 6 concludes this article.

2 RELATED WORK

The work is related to the prior works on sensing systems and mobile health systems based on audible signals with smartphones. We review the key related works in this section.

Mobile sensing systems with audible signals have attracted great attention [Li et al. 2017; Stowell et al. 2015]. Such systems are increasingly convenient for people's lives and healthcare. In the meantime, by using mobile audible sensing systems with smartphones, people do not need to pay extra money for new devices. Traditionally, mobile audible sensing systems can be classified into two categories, including passive and active sensing systems. First, the passive audible sensing systems mainly focus on how to leverage the microphone to sense and recognize the surround audible signal [Fu et al. 2015]. The work on AAmouse leverages an inaudible sound pulse at different frequencies to transform a mobile device into a mouse by exploiting the Doppler shift, speed, and distance estimation [Yun et al. 2015]. Moreover, the CAT system implements a distributed FMCW for tracking devices, such as Virtual Reality/Augmented Reality (VR/AR) headsets. This work mainly is focused on synchronizing two smartphones and using the microphone as a mouse, which can interact with VR/AR headsets for more accurate localization [Mao et al. 2016]. In addition, an audible-based sensing technique is used for wireless a virtual keyboard with smartphones. Keystroke snooping [Liu et al. 2015b] and Ubik systems [Wang et al. 2014] can obtain the sound signal with a smartphone's single or dual microphones and leverage the time-difference-of-arrival (TDOA) measurements to monitor finger strokes on the table. Then, the strokes are transformed into related alphabets in the same position as a computer keyboard. SilentWhistle is a lightweight indoor localization system using acoustic sensing for obtaining users' locations [Qiu and Mutka 2017]. Another work, Dhwani, builds an acoustics-based NFC system with smartphones, using a technology called JamSecure. It can provide a

secure communication channel between devices, which is an OFDM channel for audible signals [Nandakumar et al. 2013].

On the other hand, active inaudible sensing systems transfer a smartphone to an active sonar using ultrasonic sound waves at 18 KHz to 22 KHz, which is closely related to the proposed SonarBeat system [Nandakumar and Gollakota 2017]. OFDM-based sensing systems such as FingerIO can track the finger movement in a 2-D domain through tracking echoes from the finger that are received by the microphone to measure the finger position [Nandakumar et al. 2016]. BatMapper employs an acoustic sensing-based system for fast and accurate floor plan construction using commodity smartphones [Zhou et al. 2017]. Moreover, LLAP leverages the principle of CW radar to measure distance and implement device-free hand tracking [Wang et al. 2016]. This work is closely related to SonarBeat, because both systems use the phase-based CW signal to sense movements. The difference between the two systems is that SonarBeat is more robust to different environments with the adaptive median filter technique. On the other hand, the AudioGest system employs a pair of built-in speakers and microphones to send inaudible sound and leverage the echoes to sense the hand movement [Ruan et al. 2016]. Our SonarBeat system is motivated by the active inaudible acoustic sensing systems and transforms a smartphone into active sonar with ultrasonic sound waves.

Mobile health applications and research have become an important part of the IoT [Dimoulas 2016]. Smartphones and other wearable devices can provide people with a more convenient way to monitor their health conditions without the need for professional equipment [Bernal et al. 2017]. The work on Burnout leverages accelerometers to sense skeletal muscle vibrations, which does not require one to wear a suit embedded with sensors [Mokaya et al. 2016]. Yang et al. [2017] propose to capture the depth video of a human subject with Kinect 2.0 to monitor heart rate and rhythm. Moreover, wearable devices for monitoring exercise and body are widely available. For example, the FEMO system achieves an integrated free-weight exercise monitoring service with RFID tags on the dumbbells and leverages the Doppler shift for recognition and assessment of free-weight exercise [Mokaya et al. 2016]. For vital signal monitoring, fine-grained sleep monitoring using a microphone in the earphone can record human breathing sounds to monitor people's heath signals when they are sleeping [Ren et al. 2015], which adopts a passive audible method. The Apnea system uses an active sonar with smartphone to monitor the breathing signal [Nandakumar et al. 2015]. This work leverages the FMCW technique for respiration signal monitoring, which requires the system to estimate the distance between the smartphone and the chest of the subject.

3 SONAR PHASE ANALYSIS AND TECHNICAL CHALLENGES

3.1 Sonar Phase Analysis

We propose to use smartphones to monitor respiration signals by utilizing an inaudible sound signal, where the speaker and microphone of the smartphone emulate an active sonar system. In particular, the speaker transmits an inaudible sound signal in the frequency range of 18 to 22 KHz, in the form of a CW signal as $C(t) = A\cos(2\pi ft)$, where A is the amplitude and f is the frequency of the sound. Then the signal is reflected by the chest of the test subject and received by the microphone. One unique advantage of the smartphone-based design is that, because the speaker and microphone use the same frequency, there are no carrier frequency offset (CFO) errors between the sender and receiver. Thus, we can exploit the phase of the received inaudible signal to accurately estimate the vital sign.

In communications theory, a modulated signal $S(t)$ can be written as $S(t) = I \times \cos(2\pi ft) + Q \times \sin(2\pi ft)$, where I is the In-phase Component and Q is the Quadrature component. A demodulator can be used to recover the I and Q functions, which are then used to recover the baseband signal. To extract the phase of the CW signal, we need to design a coherent detector to down-convert the received sound signal $R(t)$ to the I-component and Q-component of a baseband signal in Figure 1. The I-component and Q-component of the baseband signal can represent a complex vector, which can be used to obtain the amplitude and phase information of the baseband

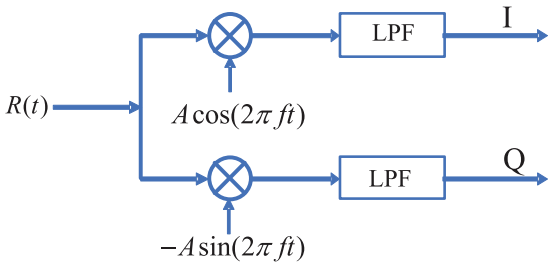

Fig. 1. I/Q demodulation for the received signal.

signal. For the SonarBeat system, we mainly exploit the phase information extracted from the I-component and Q-component of the baseband signal, thus capturing the period signal caused by the movement of the chest such as inhaling and exhaling. The SonarBeat design is to first split the received sound signal into two identical copies. Then, these two copies are multiplied with the transmitted signal $C(t) = A\cos(2\pi ft)$ and its phase-shifted version $C'(t) = -A\sin(2\pi ft)$. Finally, the corresponding In-phase and Quadrature signals are obtained by using a low-pass filter (LPF) to remove the high-frequency components.

We first present a simple analysis for the ideal case that there is no multipath effect (or, for the high signal-to-noise (SNR) regime, where the line-of-sight (LOS) component is the dominant part of the received signal). Under this assumption, the inaudible signal travels through a single path (i.e., from the speaker to the chest and then back to the microphone) and the propagation delay can be modeled as

$$d(t) = (D_0 + D\cos(2\pi f_b t))/c, \tag{1}$$

where D_0 is the constant distance of the reflected path; D and f_b are the amplitude and frequency of the chest movements, respectively; and c is the speed of sound. The received inaudible signal from this path can be modeled as $R(t) = A_r\cos(2\pi ft - 2\pi fd(t) - \theta)$, where A_r is the amplitude of the received inaudible signal and θ is a constant phase offset due to the delay in audio recording and playing. To estimate the phase of the inaudible signal, we need to remove the high-frequency components. Multiplying the received signal with $C(t) = A\cos(2\pi ft)$, we have

$$\begin{aligned} &A_r\cos(2\pi ft - 2\pi fd(t) - \theta) \times A\cos(2\pi ft) \\ &= \frac{A_rA}{2}(\cos(4\pi ft - 2\pi fd(t) - \theta) + \cos(-2\pi fd(t) - \theta)). \end{aligned} \tag{2}$$

The first term in Equation (2) has a high frequency of $2f$, which can be removed with a properly designed low-pass filter. Thus, the I-component of the baseband is extracted as $I = \frac{A_rA}{2}\cos(-2\pi fd(t) - \theta)$. With a similar approach (i.e., multiplying by $C'(t)$ and removing the high-frequency component), we can estimate the Q-component of the baseband signal as $Q = \frac{A_rA}{2}\sin(-2\pi fd(t) - \theta)$. We then demodulate the phase of the inaudible signal data as

$$\begin{aligned} \varphi(t) &= \arctan(Q/I) = -2\pi fd(t) - \theta \\ &= -2\pi f(D_0 + D\cos(2\pi f_b t))/c - \theta. \end{aligned} \tag{3}$$

Note the phase signal $\varphi(t)$ has the same frequency as the respiration signal. In the general case, the received inaudible signal is a complex signal, which includes a *static component* and a *dynamic component* due to the multipath effect in indoor environments. For example, Figure 2 shows complex I/Q traces for the received audio signal, with the static vector and dynamic vector in the I-Q plane. To track the breathing rates, we need to demodulate the phase from the I/Q components by removing the static vector. Figure 3 shows the complex I/Q

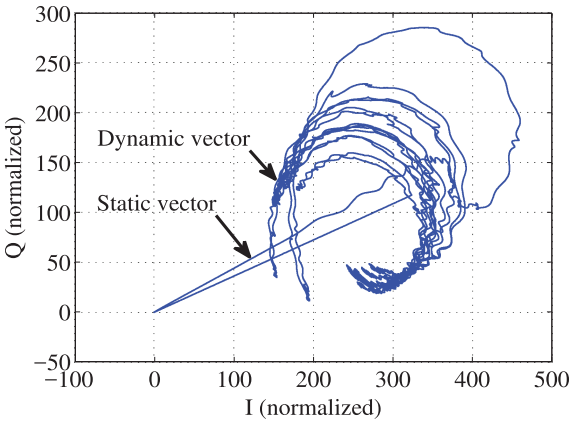

Fig. 2. Complex I/Q traces of the received audio signal.

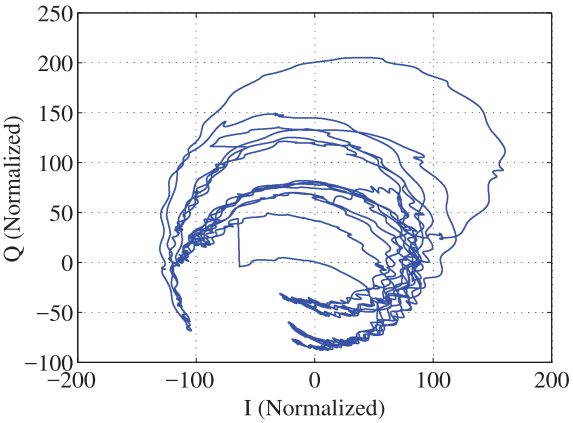

Fig. 3. Complex I/Q traces of the received audio signal after removing the static vector effect.

traces of the received audio signal after removing the static vector. It is noticed that the demodulated phase is a good indicator of the breathing-caused chest movements.

3.2 Technical Challenges

3.2.1 Mitigate the Static Vector Effect. The main challenge for respiration monitoring with phase-modulated data is to mitigate the static vector effect, which directly influences the sensitivity and correctness of the phase data. The larger the stationary component, the larger the error in the extracted phase data. This is because the SNR at the receiver will become low when there is a large static component, making it hard to demodulate the phase data. Wang et al. [2016] adopt the local extreme value detection (LEVD) to remove the stationary components for hand tracking. However, this method may not be effective for respiration monitoring, because the LEVD method needs to set an empirical threshold for each different environment. In this article, we propose an adaptive median filter method to deal with this challenge, which is shown to be effective for removing the stationary component in different scenarios.

3.2.2 Adapt to Body Movements and Environment Noise. The second challenge for tracking breathing signals is adapting to body movements and environment noise. Body movement is unavoidable in the monitoring phase

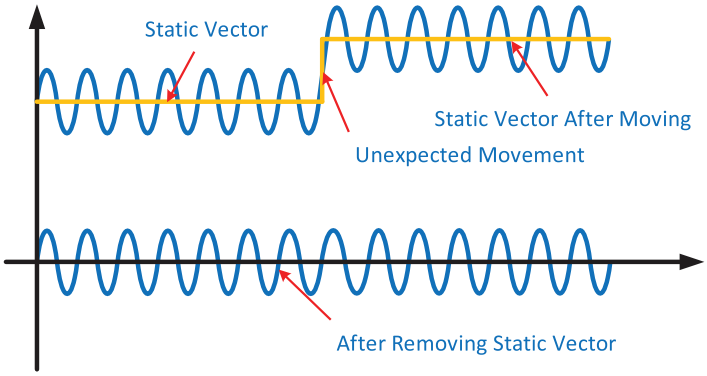

Fig. 4. Illustration of adapting to body movements by eliminating the static vector.

(e.g., during sleep), and its impact should be mitigated. The FMCW-based scheme in Nandakumar et al. [2015] requires estimating the distance between the smartphone and chest before respiration monitoring. When the body suddenly moves, the system needs to seek the new distance, thus leading to large time complexity. The proposed sonar phase-based approach is effective in adapting to body movements. For example, Figure 4 illustrates the idea of adapting to body movements by eliminating the static vector in SonarBeat. When there is an unexpected small body movement, the magnitude of the breathing signal becomes larger, leading to a smaller SNR. After eliminating the static vector, we can mitigate the effect of body movement and still obtain a neat respiration signal, as shown in the lower part of Figure 4.

Furthermore, consider the case of multiple persons in the testing environment. Not only do their movements cause interference to the reflected respiration signal, but also the background noise could be high (e.g., when they are talking). We employ coherent I/Q demodulation in SonarBeat to remove the environment noise from external audio sources.

3.2.3 Real-Time Monitoring with Lower Delay. For a vital sign monitoring system to be really useful, it should work in real time, with good interactions with the user. Real-time monitoring is challenging since most smartphones have a high sampling rate of 48 KHz, which leads to 96,000 multiplication operations per second for the coherent detector to down-convert the received sound signal to the baseband. To address this challenge, we perform down-sampling for I/Q demodulation, which can reduce the computation complexity while still capturing the breathing rate.

For better interaction with the user, SonarBeat operates in three stages. In the first stage of 15 seconds, it performs respiration monitoring in real time without FFT-based breathing rate estimation. In the second stage of 15 seconds, SonarBeat analyzes the data collected in a 15-second sliding window to extract the respiration signal and plots the respiration signal on the screen to give the user some preliminary testing results. In the final stage, after 30 seconds, SonarBeat applies FFT to all the captured phase data to achieve an accurate breathing rate estimation.

4 THE SONARBEAT SYSTEM

4.1 SonarBeat System Architecture

According to the sonar phase analysis, SonarBeat can effectively exploit sonar phase information to monitor respiration signals. First, the phase information can track the periodic breathing rates with high accuracy, and the phase information is sensitive to the small breathing-induced chest movements. Second, compared with other traditional methods, such as Doppler shift and FMCW [Adib et al. 2015], the phase-based approach has a lower latency and complexity. Finally, the sonar phase data is robust to different orientations, different distances,

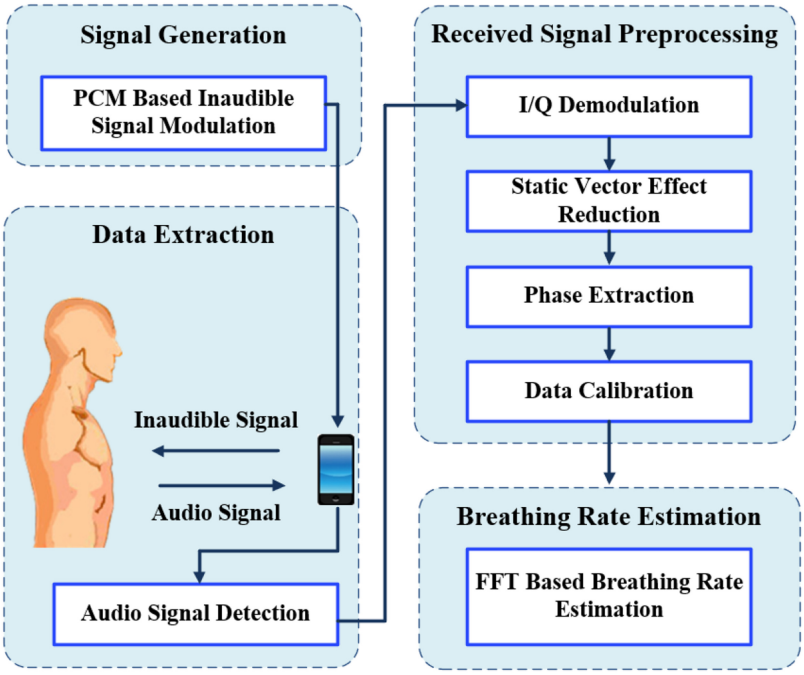

Fig. 5. The SonarBeat system architecture.

different cloth thicknesses, and different breathing rates of different persons. It is also robust to large body movements, which only leads to a change of the stationary component of the phase data, which can be effectively removed with the proposed adaptive median filter method.

Figure 5 presents the SonarBeat system architecture, which includes four basic modules: (1) Signal Generation, (2) Data Extraction, (3) Received Signal Preprocessing, and (4) Breathing Rate Estimation. The *Signal Generation* module mainly implements a Pulse-code Modulation (PCM) of the inaudible signal, where a CW signal at 18 KHz to 22 KHz is generated and modulated with the PCM technique. The *Data Extraction* module is to detect the audio signal, which employs a Short-Time Fourier transform (STFT) for audio signal detection. A threshold-based method is proposed for detecting the beginning part of the received inaudible signal. The *Received Signal Preprocessing* module consists of (1) I/Q demodulation, which implements down-sampling followed by a coherent phase detector; (2) static vector effect reduction, which implements the proposed adaptive median filter method to remove the static vector of the In-phase and Quadrature signals; (3) phase extraction, where the sonar phase information is extracted and calibrated; and (4) data calibration, where a median filter is applied as a simple Low-Pass finite impulse response (FIR) filter to remove noise. The *Breathing Rate Estimation* module employs an FFT-based method to estimate the breathing rate.

4.2 Signal Generation

The signal generation module uses one speaker of the smartphone as the transmitter, to produce the inaudible signal. We implement the signal generation module as a PCM-based modulator on the Android platform. Specifically, the speaker generates an inaudible sound signal in the frequency range of 18 to 22 KHz in the form of a CW signal, that is, $C(t) = A\cos(2\pi f t)$. We produce the sampled analogy signal and then use PCM to digitally represent the sampled CW signal. To generate a PCM stream, the amplitude of the analog CW signal is sampled

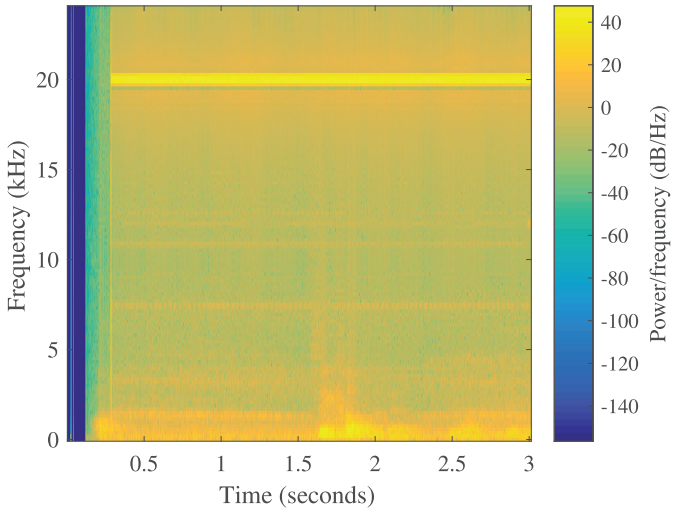

Fig. 6. STFT-based method for audio signal detection.

at uniform intervals, where each sampled value is quantized. The PCM-based inaudible signal modulation is implemented with the AudioTrack class.

4.3 Data Extraction

We use the speaker of the smartphone to transmit the CW inaudible signal at 18 kHz to 22 kHz, which does not contain any user information (just a cosine wave). The microphone of the smartphone is used to receive the inaudible signal reflected from the chest with a sampling rate of 48 KHz. The microphone may also record other signals at different frequencies from the surrounding environment. Although the human voice and other audible sounds are both captured by the microphone in the raw acoustic signal, these audible signals are considered as noise and filtered out by the low-pass filters as illustrated in Figure 1. Thus, the privacy of the user is well protected because no audible sound is recorded by the system. We implement an audio signal detection method to identify the beginning of the desired signal as follows.

The proposed audio signal detection method is based on STFT. At the beginning of the signal, there will be a drastic increase of power at the carrier frequency. A threshold-based method is used to detect the beginning of the inaudible signal. Figure 6 illustrates the STFT-based method for audio signal detection, where the carrier frequency is 20 KHz. We can see that before 0.25 seconds, the microphone only receives audio frequencies from the surrounding environment. After 0.25 seconds, the microphone detects the inaudible signal, since the magnitude of the 20 KHz spectrum becomes much stronger than other audio frequencies (see the bright yellow, horizontal strip at 20 KHz). We adopt a window size of 512 in STFT for estimating the spectrum. Moreover, we set a threshold of 200 for the power change to detect the beginning of the inaudible signal. In fact, if we detect the power change with the threshold method, the beginning of the inaudible signal can be set as the end of the STFT chirp.

4.4 Received Signal Preprocessing

We describe the four components of the Received Signal Preprocessing module in this section, including I/Q Demodulation, Static Vector Effect Reduction, Phase Extraction, and Data Calibration.

4.4.1 I/Q Demodulation. Before I/Q Demodulation, we need to down-sample the received signal $R(t) = A_r \cos(2\pi f t - 2\pi f d(t) - \theta)$ for reducing computation complexity, which is necessary for real-time monitoring. The original system with sampling frequency of 48 KHz is reduced to 480 Hz with a down-sampling ratio

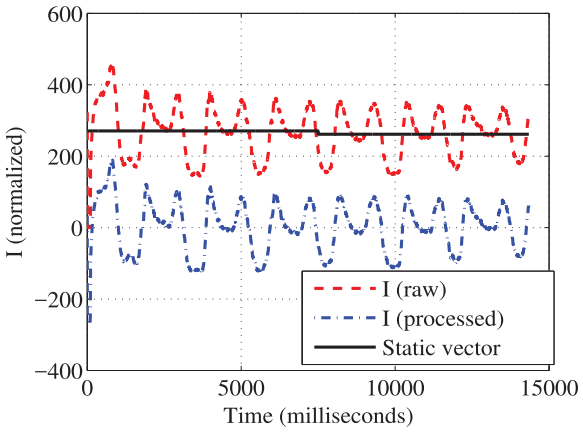

Fig. 7. The adaptive median filter for removing the static vector in the baseband I-component.

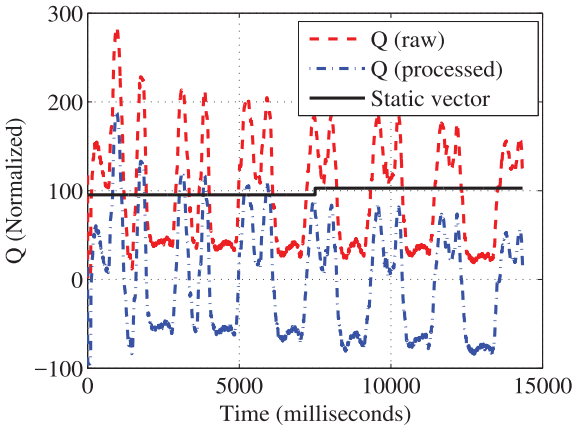

Fig. 8. The adaptive median filter for removing the static vector in the baseband Q-component.

of 100. Then, we implement the I/Q demodulation to obtain the I-component and Q-component of the baseband signal using a coherent detector. The design is to split the received audio signal into two identical copies. Due to the down-sampling ratio of 100, these two copies should be multiplied with the signal $A\cos(2\pi\frac{f}{100}t)$ and its phase-shifted version $-A\sin(2\pi\frac{f}{100}t)$ to obtain the I-component and Q-component of the baseband signal, respectively. Because we use phase modulation for breathing monitoring, down-sampling only reduces the number of samples of the amplitude of the breathing signal, rather than the phase information.

Finally, an LPF is employed to obtain the corresponding In-phase and Quadrature signals, which has a cutoff frequency of 1 Hz, a sampling rate of 480 Hz, and a resonance of 2. This setting has been shown to be effective for removing the high-frequency components and environment noises. In Figures 7 and 8, we plot the raw I-component and Q-component of the baseband signal, respectively, after the LPF (the dashed curves), which, however, still include their static vectors.

4.4.2 Static Vector Effect Reduction. As discussed, the performance of SonarBeat largely depends on how the effect of the static vector is mitigated in multipath environments. This is because usually the static vector is much stronger than the dynamic vector that represents the small chest movements. It is difficult to detect the weak

ALGORITHM 1: The Adaptive Median Filter Method

Input: One baseband signal component: $X(n)$, $n = 0, 1, \ldots, N-1$, and the window size w;
Output: The baseband signal component with static vector removed: $O(n)$, $n = 0, 1, \ldots, N-1$;
//Initialization
n_w: number of windows;
r: number of remaining elements of $X(n)$, which cannot form a full window of size w;
$W[1, 2, \ldots, w]$: sublists with window size w;
$R[1, 2, \ldots, r]$: sublist with the remaining r elements;
//Find the median for each window
for $i = 0 : n_w$ **do**
  if $((i+1) * w) <= N$ **then**
    $W[1, 2, \ldots, w] \leftarrow X((w * i)$ to $((i+1) * w - 1))$;
    $M \leftarrow$ the median of $W[1, 2, \ldots, w]$;
    for $j = w * i : (i+1) * w$ **do**
      $O(j) = X(j) - M$;
    end
  end
  else if $r \neq 0$ **then**
    $R[1, 2, \ldots, r] \leftarrow X((w * i)$ to $(w * i + r - 1))$;
    $M \leftarrow$ the median of $R[1, 2, \ldots, r]$;
    for $j = w * i : (i+1) * w$ **do**
      $O(j) = R(j) - M$;
    end
  end
end
Return $O(n)$, $n = 0, 1, \ldots, N-1$;

breathing signal if we directly use the received sound signal. Recently, there were two methods proposed for static vector effect mitigation. The Dual-Differential Background Removal approach is used for hand tracking with 60 GHz mmWave signals [Wei and Zhang 2015]. The method is susceptible to environment noise and has large latency, which is not effective for real-time respiration monitoring. The second scheme, termed Local Extreme Value Detection (LEVD) [Wang et al. 2016], was also developed for tracking hand movements. The method requires an empirical threshold for detecting the static vector, which is not robust for different environments and different test subjects.

In Algorithm 1, we present an adaptive median filter method for removing the static vector, which has a low latency and is robust to different environments. The idea is to use a window to obtain the median for estimating the static vector. The only parameter is window size w, which is robust for different environments. For baseband signal component $I(n)$ or $Q(n)$, $n = 0, 1, \ldots, N-1$, we partition it into multiple non-overlapping sublists, each denoted by $W[1, 2, \ldots, w]$) with window size w, and a single sublist $R[1, 2, \ldots, r]$ with window size $r < w$, where r is the number of remaining elements of $I(n)$ or $Q(n)$ not included in the previous W sublists. The sublists $W[1, 2, \ldots, w]$ and sublist $R[1, 2, \ldots, r]$ are used to estimate the medians for the first $n_w - 1$ windows and the last window, respectively, where $n_w = \lfloor N/w \rfloor$. Finally, the output $O(n)$, $n = 0, 1, \ldots, N-1$, can be obtained as in Steps 14 and 21 by removing the static vector. The proposed method is simple and robust for real-time processing of received data in different environments with a low delay.

Figure 7 and Figure 8 illustrate how the adaptive median filter method removes the static vectors in the baseband signal components. We can see that the estimated static vector can represent well the average amplitude

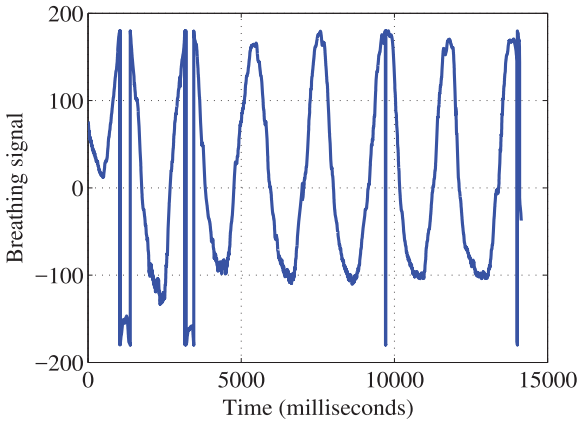

Fig. 9. Respiration curve for phase data with removing static vector effect.

information of the baseband signal components. After the adaptive median filter, the components I and Q are roughly centered at zero; the improved SNR makes it easier for extracting the breathing signals they carry.

4.4.3 Phase Extraction. After removing the static vector, we next extract the phase data in the I-Q plane, which only includes the dynamic breathing component. Let $O_I(t)$ and $O_Q(t)$ denote the outputs of Algorithm 1. The phase of the inaudible signal can be computed with Equation (3), that is:

$$\varphi(t) = \arctan\left(\frac{O_Q(t)}{O_I(t)}\right). \tag{4}$$

With Equations (3) and (4), we find the phase value $\varphi(t)$ for the respiration signal reflected from the chest. Although the reflected respiration signal may still have multipath components, these multipath signals have the same breathing frequency but with different phase shifts, each of which is a constant. Thus, the breathing rate will not be affected by the dynamic multipath effect. This is different from hand tracking, which requires only one path from the smartphone and the hand. Thus, our SonarBeat can estimate the breathing rate using a single subcarrier rather than multiple subcarriers.

Figure 9 presents the respiration curve obtained from the phase data with static vectors removed. It is noticed that the magnitude of the breathing signal is large, which is periodic if we can remove the sudden phase changes. Thus, we need to implement a data calibration scheme for the demodulated phase data with better periodicity.

4.4.4 Data Calibration. We implement a phase unwrapping scheme for recovering the correct phase values, as well as a median filter for reducing the environment noise. To estimate breathing rates, we need to obtain the right breathing curve for phase data. Because the phase value will have a change of 2π for every wavelength distance, we implement a phase unwrapping scheme to process the demodulated phase data.

Figure 10 shows the respiration curve obtained from the phase data after phase unwrapping. It is a clear breathing signal, but still with smaller environment noises. Moreover, we adopt the median filter method to remove the environment noise, where the filter window size is set to 300. Figure 11 presents the respiration curve for the unwrapped phase data after the median filter, which is next used for accurate breathing rate estimation.

4.5 Breathing Rate Estimation

SonarBeat operates in three stages for breathing monitoring, for better interactions with the user. In the first stage of 15 seconds, we cannot effectively estimate the breathing rate but just record the extracted respiration

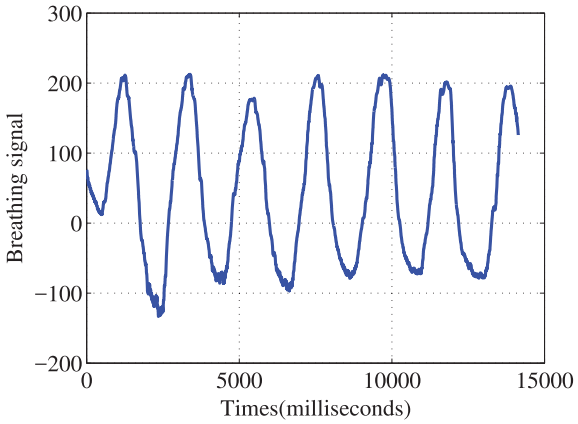

Fig. 10. Respiration curve for phase data after unwrapping the phase data.

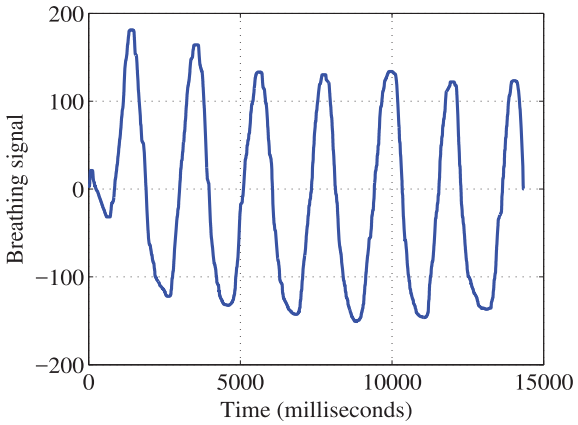

Fig. 11. Respiration curve for unwrapped phase data after the median filter method.

signal, because the phase data in a window can continually change when new data arrive. In the second stage of 15 seconds, we analyze the collected phase data in a 15-second sliding window to extract the breathing signal and plot it on the smartphone screen. In the final stage, after 30 seconds, we use all the collected phase data for breathing rate estimation with FFT, which can achieve a higher estimation accuracy (since more data is available now). In fact, the frequency resolution depends on the window size of FFT. If the window size becomes larger, the estimation accuracy will be higher, but a larger window size also leads to a lower time domain resolution. Thus, for online breathing rate estimation, we use the same window size as that of the STFT-based method. It balances the tradeoff between the frequency domain resolution and the time domain resolution. Figure 12 illustrates the FFT-based respiration rate estimation. We can see that the estimated breathing frequency is 0.23 Hz, which is approximately the same as the true breathing rate measured by the NEULOG Respiration Monitor Belt Logger Sensor during the experiment. In our experiments, no thoracic breathing (i.e., breath performed only by the chest when the abdomen does not move) is involved, so the abdomen movement captured by the NEULOG belt can be used as accurate ground truth in the system evaluation.

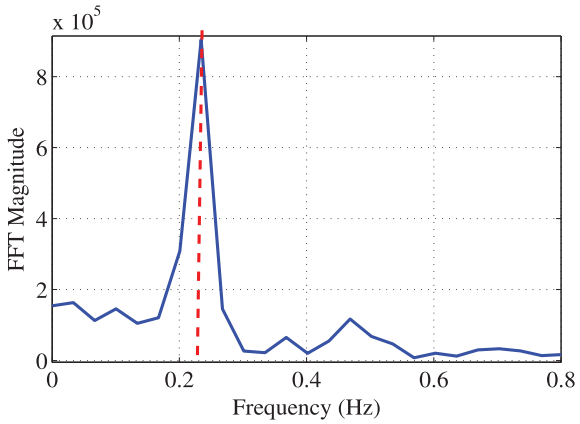

Fig. 12. Respiration rate estimation based on FFT.

Table 1. Participant Characterization

Subject Index	*Gender*	*Age*	*BMI*
Subject 1	Male	31	21
Subject 2	Male	28	26
Subject 3	Male	29	24
Subject 4	Male	28	24
Subject 5	Female	29	19

5 EXPERIMENTAL STUDY AND DISCUSSIONS

5.1 Experiment Configuration

We prototype the SonarBeat system on the Android platform in Java and the Android SDK, as a smartphone app. The first edition of SonarBeat is implemented with the minimum version of Android 5.1.1 OS (API 21). So it works with all the more recent Android systems such as Android 6.0 and Android 7.0. The app is evaluated with Samsung Galaxy S6 and Samsung Galaxy S7 Edge smartphones. For respiration monitoring, we use one speaker and one microphone to transmit and receive the inaudible audio data, respectively, while the microphone and speaker are fixed at the bottom of the smartphone. Furthermore, we use the AudioTrack class to play inaudible sound and the AudioRecord class to record sound. The buffer of the recording thread is set to 1,920 points with a sampling rate of 48 KHz. Therefore, we set the real-time signal processing unit to 1,920 points, which is about 40 ms.

We conduct extensive experiments using SonarBeat with five persons over a period of 3 months. The characterization of all the subjects is presented in Table 1. The test scenarios include an office, a bedroom, and a movie theater, as shown in Figures 13, 14, and 15, respectively. The *office* is a 4.5×8.8 m^2 room, which is crowded with tables and PCs, forming a complex propagation environment. In this office environment, we test SonarBeat under different parameter settings. The second environment is a *bedroom* of 3.9×6 m^2, where we test breathing monitoring for a single person. The third setup is a *movie theater* of a large 27×40 m^2 area, where many people are watching a movie, with strong audio interference from the movie and other people. The office and bedroom scenarios are implemented over longer periods with multiple times in different days as compared to the movie cinema, which is tested in 1 hour for watching a movie. During these periods, multiple experiments are conducted, each lasting for 30 seconds. In each experiment, the smartphone is either held in hand or placed on a

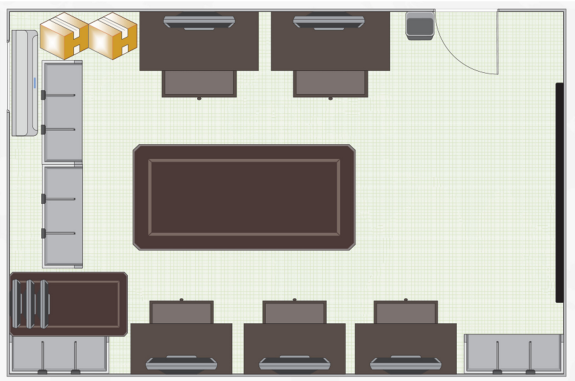

Fig. 13. Experimental setup in the *office* scenario.

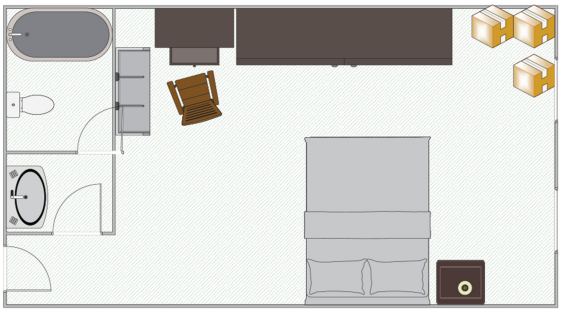

Fig. 14. Experimental setup in the *bedroom* scenario.

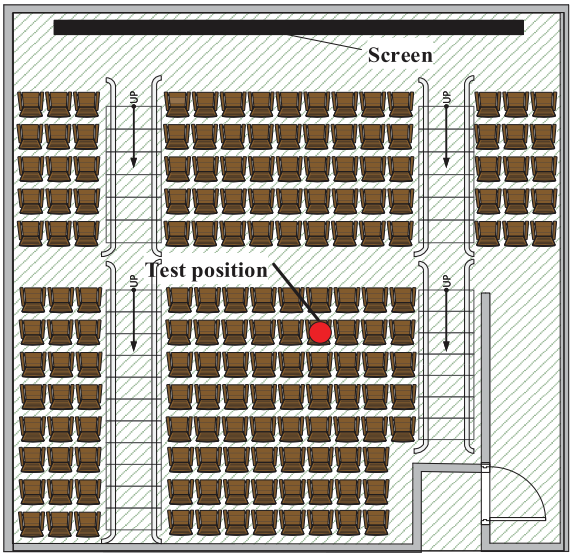

Fig. 15. Experimental setup in the *movie theater* scenario.

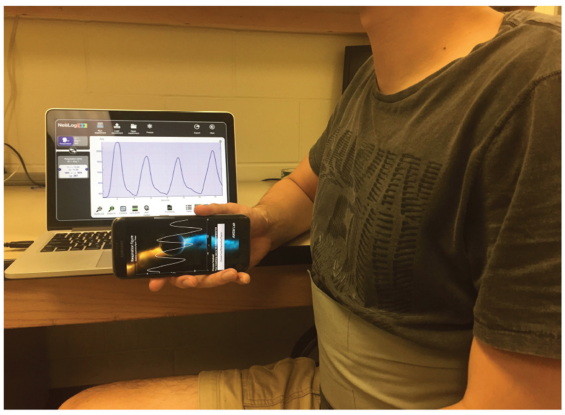

Fig. 16. The office experiment, where the NEULOG Respiration Monitor Belt Logger Sensor records the ground truth (shown on the laptop screen).

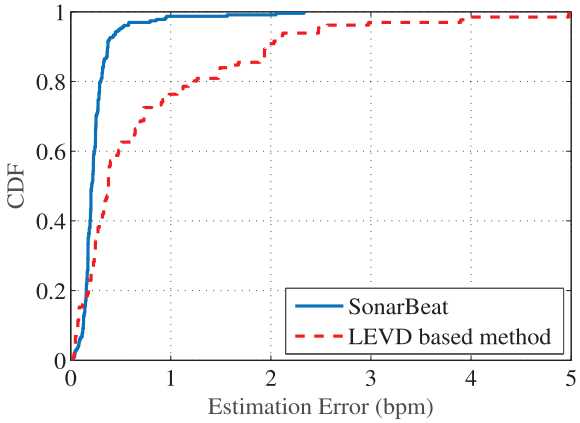

Fig. 17. CDFs of estimation errors in breathing rate estimation.

desk, with the speaker and microphone facing the subject. Moreover, the proposed system mainly focuses on estimating the breathing rate for a single person, where we consider other persons to have above 60 cm distance away from the smartphone. For comparison purposes, we use the NEULOG Respiration Monitor Belt Logger Sensor to record the ground truth of the breathing rate (see Figure 16).

For breathing rate estimation each time, we use all the collected data for breathing rate estimation with FFT for 30 seconds. In each experiment, we measure the error in respiration rate in bpm by comparing to the ground truth. Moreover, we use the cumulative distribution function (CDF) of breathing errors as the measurement metric, which can be employed to account for the random factors in the measurement. We also consider the mean estimation error as another evaluation metric for measuring the impact of various environmental factors and the impact of various system parameters.

5.2 Performance of Breathing Rate Estimation

Figure 17 presents the CDFs of estimation errors in breathing rate estimation with SonarBeat. Besides, we also compare our SonarBeat system with the LEVD-based system [Wang et al. 2016], where the LEVD method is used for estimating the static vector and all other signal processing methods are the same as in SonarBeat. The figure

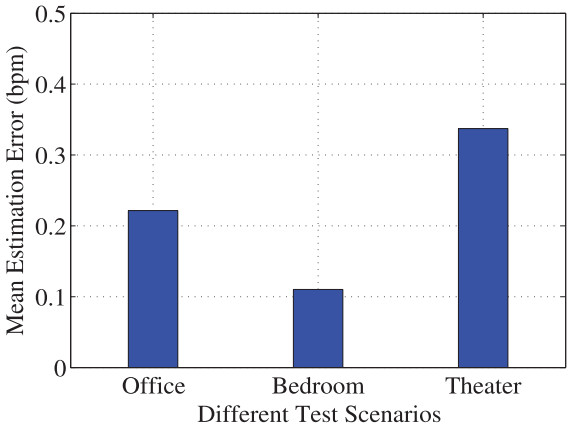

Fig. 18. Mean estimation error for three different scenarios: office, bedroom, and movie theater.

shows that the median errors of SonarBeat and the LEVD-based method are 0.2 bpm and 0.3 bpm, respectively. This illustrates that both systems can effectively estimate breathing rates. However, it is worth noting that for SonarBeat, 95% of the test results are under 0.5 bpm, while only 60% of the test results with the LEVD-based method are under 0.5 bpm. Moreover, the maximum estimation error of SonarBeat and the LEVD-based method are 2.4 bpm and 5 bpm, respectively. This is because the LEVD-based method requires setting the empirical threshold based on the standard deviation of the baseband signal in a static environment. It is not very robust in varying environments where the same threshold will not work. However, the adaptive median filter method used in SonarBeat is more robust to environmental interference, which leads to a higher and more stable breathing rate estimation accuracy than the LEVD-based baseline scheme.

Figure 18 presents the mean estimation errors for the three different scenarios, which are 0.22 bpm, 0.11 bpm, and 0.33 bpm for the office, bedroom, and movie theater scenarios, respectively. We plot the 95% confidence intervals as error bars. The mean estimation error of the bedroom case is the minimum, because the bedroom is a better environment, with smaller noise and no sound interference from other persons. This shows that SonarBeat is suitable for breathing monitoring during sleeping, which helps to detect apnea or other sleeping problems. For breathing monitoring in the office, the performance is worse than the bedroom. This is because the propagation environment is more complex and there is interference from other people. Furthermore, higher noises from computers, air conditioners, and other equipment in the lab also influence the received inaudible signal. The movie theater test has the largest mean error and variance because of the more complex environment and stronger noises. In fact, breathing monitoring in the theater is still quite accurate given the extremely adverse environment. These experiments validate that SonarBeat is highly accurate and robust in different scenarios.

5.3 Impact of Various Environmental Factors

Figure 19 shows the impact of different persons in the office scenario. In the experiment, we test five persons including three men and two women. The purpose is to show the impact of different subjects on the SonarBeat performance. Each result in the figure is obtained with a different experiment. Every volunteer wears the NEULOG Respiration unit to record the ground truth for breathing data. From Figure 19, we can see that Persons 3 and 5 have a relatively lower mean error. This is because they work out quite often and have a stronger respiration, leading to stronger breathing signals. On the other hand, the other three persons have weaker breathing magnitudes, but their estimation errors are still under 0.5 bpm, which is acceptable. Thus, we can see that SonarBeat is adaptive for different persons.

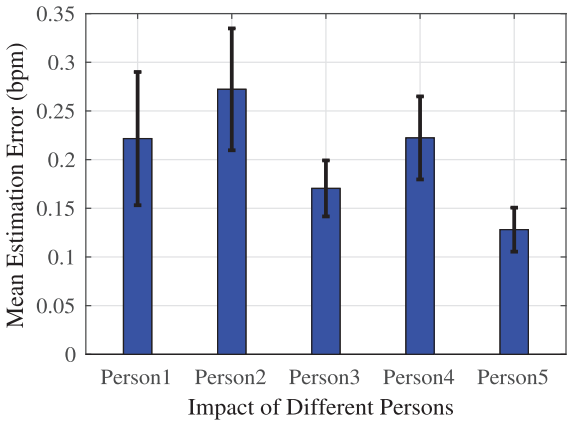

Fig. 19. Breathing rate results for five different persons.

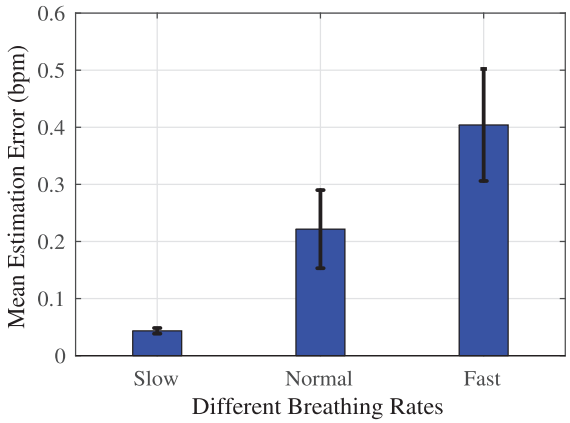

Fig. 20. Impact of different breathing rates.

Figure 20 shows the impact of different breathing rates in the office scenario, where the test subject controls his or her breathing at slow, normal, and fast breathing rates, which are in the ranges from 6 bpm to 10 bpm, 13 bpm to 18 bpm, and above 30 bpm, respectively. It is noticed that with the increase of breathing rate, the mean estimation error will be increased. The reason is that with a higher breathing rate, the stability of the breathing signal becomes weaker. In other words, with fast breathing, the chest movements are more irregular, leading to large variations in the captured breathing curve. Moreover, we adopt the FFT-based breathing estimation. When there are multiple breathing frequencies embedded in the captured breathing signal, FFT does not produce good frequency estimation. Nevertheless, SonarBeat can still effectively capture the different breathing rates with a mean error of a little over 0.4 bpm in the fast breathing case.

Figure 21 illustrates the impact of the distance between the chest and the smartphone. When the distance is increased, the accuracy of breathing estimation becomes lower. Particularly, we can see that for a distance of 55 cm, the mean estimation error becomes 1 bpm with a large variance. In this experiment, we find that the ultrasound wave in 18 KHz to 22 KHz experiences a large attenuation, and the microphone will receive lower power from the chest reflection if the distance is increased beyond 50 cm. Moreover, breathing rate estimation with SonarBeat depends on I/Q demodulation. The magnitude of the I/Q components becomes weaker when the

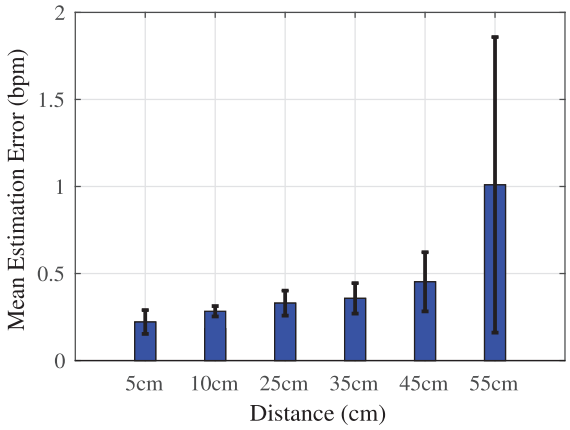

Fig. 21. Impact of the distance between the test subject and the smartphone.

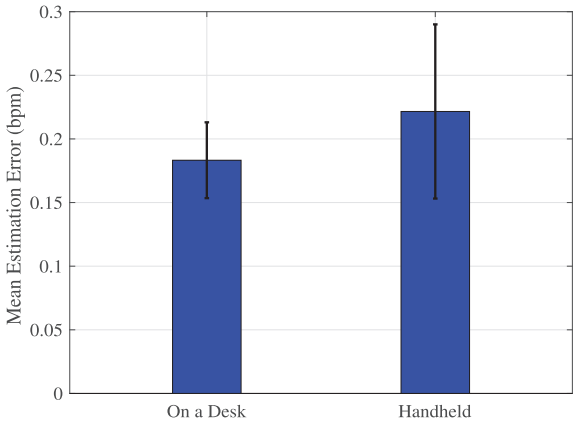

Fig. 22. Breathing rate results when the smartphone is held in hand or put on a desk.

distance is increased, leading to higher errors. To improve the measurement coverage, we leverage the parameter resonance of the low-pass filter to strengthen the amplitude of the inaudible signal near the cutoff frequency, thus improving the magnitudes of the I/Q components. In the experiment, we set the cutoff frequency to 40 Hz for a sampling rate of 48 KHz. We can see that under 50 cm, the proposed system can achieve very high accuracy. The effective range of the system implies that both chest and abdomen can be effectively captured by the system, as long as the smartphone is close to the upper body of the subject. Thus, SonarBeat can work well whenever the subject is having abdominal breathing or thoracic breathing.

Figure 22 presents the breathing rate errors when the smartphone is held in hand or put on a desk. We find the error is low in both cases, which is 0.22 bpm when the smartphone is held in hand and 0.16 bpm when it is put on a desk. Moreover, the variance of breathing rate estimation errors with a handheld smartphone is larger. This is because the small hand movements will affect the reflected signal. Although the small hand movements do not influence the basic estimation accuracy, it still causes a larger variance of the estimation error. On the other hand, because we use the adaptive median filter to effectively remove the static vector, the mean breathing rate estimation results can be guaranteed for both cases.

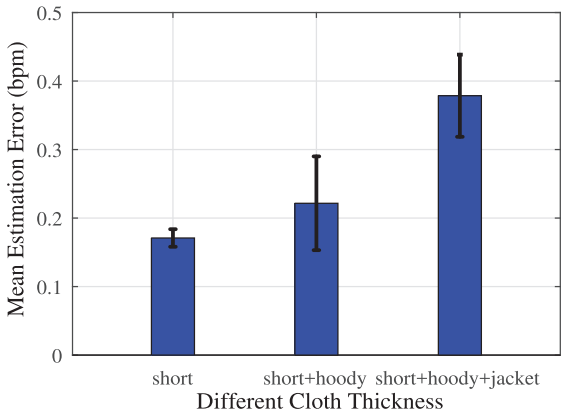

Fig. 23. Impact of cloth thickness.

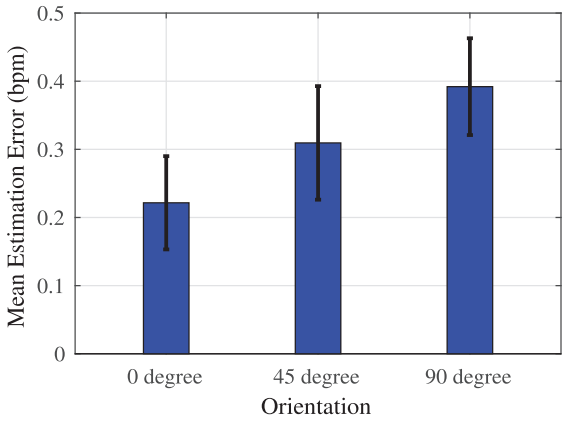

Fig. 24. Impact of user orientation relative to the smartphone.

Figure 23 shows the impact of cloth thickness. In the experiment, the test subject wears clothes of different types and thicknesses. The distance between the user and the smartphone is kept between 10 cm and 15 cm. It is noticed that, with the increase of cloth thickness, the mean error becomes larger. This is because the ultrasound wave at 18 KHz to 22 KHz experiences larger attenuation when the clothes get thicker, which leads to a smaller received breathing signal and a lower SNR. In fact, the maximum breathing estimation error is about 0.37 bpm in this experiment, which is still acceptable.

Figure 24 shows the impact of chest orientation relative to the smartphone in the office scenario, where we consider three cases of 0°, 45°, and 90°. It is noticed that at 0° direction with the front orientation relative to the smartphone, we can obtain the minimum mean estimation error, which is about 0.22 bpm. At the 90° direction, the maximum mean estimation error becomes 0.39 bpm. The received inaudible signal is the strongest when the person faces the smartphone speaker.

Figure 25 presents the impact of different poses, including sitting, standing, and sleeping, in the bedroom scenario. For the sitting case, the smartphone is put on a desk and the distance between the smartphone and the chest is about 15 cm. For the standing case, the person holds the smartphone at the same distance. For the sleeping case, the person wearing a night-gown is lying in the bed, facing the smartphone on the desk at the same distance. We can see that for the sitting and standing poses, the mean estimation errors are smaller than

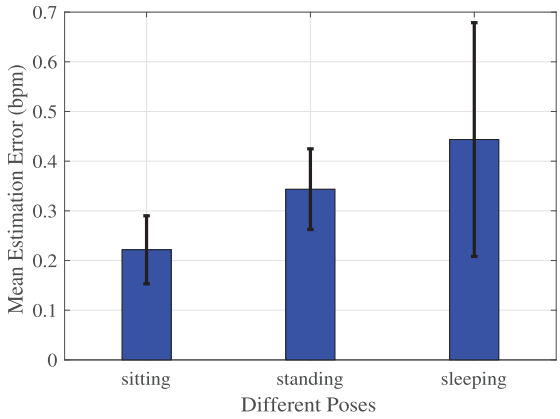

Fig. 25. Impact of different poses.

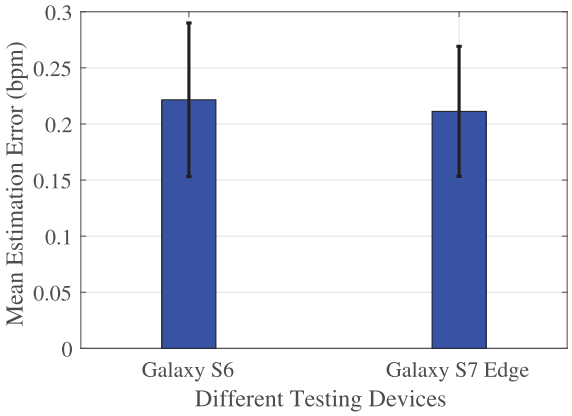

Fig. 26. Estimation error results with two different smartphone platforms.

that of the sleeping pose. This is because the strength of the received signal in the sleeping scenario is smaller than that of the other two cases.

5.4 Impact of Various System Parameters

We evaluate the impact of various system parameters in this section. Figure 26 presents the estimation errors with a Samsung Galaxy S6 with Android 6.0 and a Samsung Galaxy S7 Edge working with the latest version, Android 7.0. The speaker and microphone are on the bottom of both smartphones. We can see that the Samsung Galaxy S7 Edge has a similar performance as the Samsung Galaxy S6, with mean error of 0.21 bpm and 0.22 bpm, respectively. We find that the Samsung Galaxy S7 Edge has stronger processing power, and thus it can obtain better real-time performance than the Samsung Galaxy S6.

Figure 27 shows the impact of different frequencies for the CW signal, including 18 KHz, 20 KHz, and 22 KHz. With the increase of frequency, the mean error also gets slightly larger. The maximum mean estimation error is 0.22 bpm, while the minimum mean estimation error is 0.17 bpm. This shows that SonarBeat is robust to different frequencies. On the other hand, we also know that for ultrasound signals propagating in the same transmission medium, the power attenuate becomes larger for higher signal frequencies, resulting in a smaller SNR for the received breathing signal. Thus, the higher the frequency, the larger the mean breathing rate estimation error.

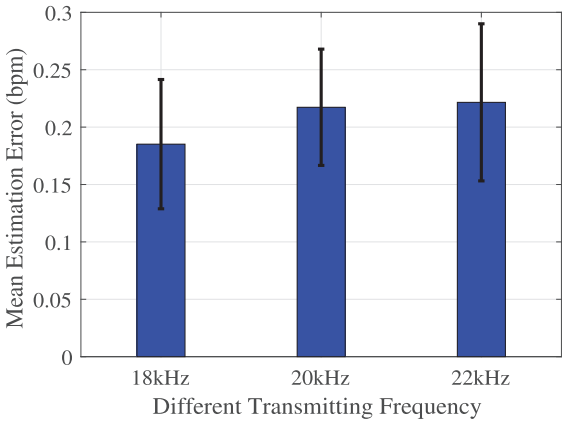

Fig. 27. Impact of the frequency of the inaudible acoustic signal.

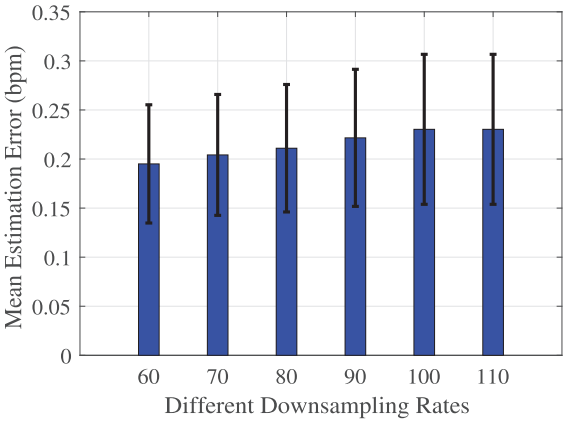

Fig. 28. Impact of different down-sampling rates.

Figure 28 shows the impact of different down-sampling rates. When the down-sampling rate is increased from 60 to 100, the mean breathing rate estimation error will be increased from 0.19 bpm to 0.23 bpm, a small increase. Thus, we choose a down-sampling rate of 100 for SonarBeat. Because SonarBeat uses a sampling frequency of 48 KHz, we can reduce it to 480 Hz by a down-sampling rate of 100. As shown before, the down-sampling operation does not affect the breathing rate embedded in the phase-modulated signal. Thus, down-sampling the I- and Q-components with a rate of 100 can not only reduce the computational complexity for real-time breathing monitoring with the smartphones but also achieve a high accuracy.

Figure 29 presents the results with different window sizes for the adaptive median filter. Recall that the window size is used in the static vector effect mitigation stage. Breathing rate estimation with SonarBeat mainly depends on reduction of the static vector effect. The only parameter of the proposed adaptive median filter approach is the window size, which should be robust for different tests. From Figure 29, we can see that when the window size is increased from 7,400 to 7,700, the mean breathing rate estimation error is only increased slightly from 0.21 bpm to 0.26 bpm. Moreover, the larger window sizes from 7,400 to 7,700 are almost half of the signal points, which is effective for removing the static vector. Thus, we choose a window size of 7,500 for SonarBeat, which can achieve the best breathing estimation accuracy.

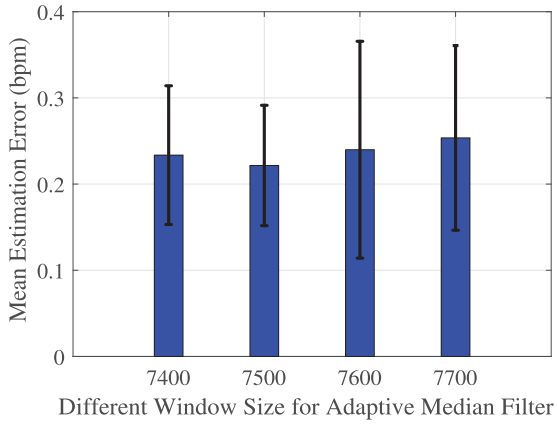

Fig. 29. Impact of different window sizes of the adaptive median filter.

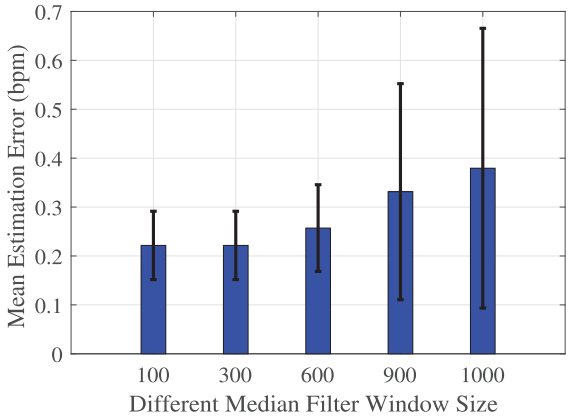

Fig. 30. Impact of different sizes of the median filter window.

Figure 30 shows the impact of different median filter window sizes on the mean error, which is used in the data calibration stage of SonarBeat. We find that when the median filter window size is set to 100 or 300, a similar mean error of 0.21 bpm is achieved. On the other hand, when the median filter window size is increased from 600 to 1,000, the mean estimation error is increased from 0.25 bpm to 0.38 bpm. This is because a smaller window size can process the local breathing curve, to effectively remove the environment noise. On the other hand, a larger median filter window size omits the local breathing noise, causing a higher error. Based on this experiment, we set the median filter window size to 300, which can not only achieve a higher breathing estimation accuracy but also lead to the better breathing curves for real-time breathing monitoring.

5.5 Discussions

In this section, we discuss the limitations of SonarBeat, as well as potential solution directions. First, SonarBeat works well when the distance between the chest and the smartphone is under 50 cm, due to the large attenuation of acoustic signals. Second, SonarBeat works well for a single person but is not suitable for monitoring multiple subjects simultaneously. Figure 21 shows that when the distance between the testing subject and the smartphone is longer than 50 cm, the breath signal will be too weak to be extracted. Thus, the testing subjects should be within 50 cm away from the smartphone to be monitored. The distance will be too close/unlikely for daily life. Third,

SonarBeat does not work well when the sleeping subject is covered with blankets, which will greatly reduce the SNR of the received signal. These limitations are mainly due to the fact that ultrasound waves in 18 KHz to 22 KHz experience a large attenuation, and the microphone will receive low power from the chest reflection when the distance is large or when there are obstacles. Note that SonarBeat uses the off-the-shelf smartphone system with fixed transmit power and maximum frequency. The performance can be significantly improved if the transmit power or the frequency of the signal can be increased.

In addition, the performance of SonarBeat largely depends on how the effect of the static vector is mitigated in multipath environments (see Figure 4). This is because the static vector is much stronger than the dynamic vector induced by the small chest movements, and the static component from the speaker to the microphone will greatly influence the SNR of the received signal. When the subject is moving, the static component suffers from considerable interference caused by the moving human body. The current median filter method cannot effectively deal with the unstable static component, so it will be difficult to detect the weak breathing signal directly from the continuously moving subject.

To further improve the performance of SonarBeat, three schemes will be considered in our future work. First, a new static vector removal algorithm should be developed. For example, we can employ two microphones and use the difference for self-interference cancellation. The second scheme is to consider modulated signals, such as FMCW and OFDM. Finally, signal decomposition techniques should be considered, such as tensor decomposition [Wang et al. 2017c] and independent component analysis (ICA). These techniques help to separate small signals, such as the heartbeat signal, from the received signal. Moreover, they can also be used to estimate breathing rates for multiple subjects.

6 CONCLUSIONS

In this article, we presented SonarBeat, a system that exploits phase-based sonar to monitor breathing rates with smartphones. We first provided a rigorous sonar phase analysis and proved the sonar phase-based method can obtain breathing signals. Also, we discussed the technical challenges for breathing estimation based on active sonar signal, including removing the static vector effect, adaption for body movements and environment noise, and online breathing monitoring with lower delay. We then described the SonarBeat design in detail, including signal generation, data extraction, received signal preprocessing, and breathing rate estimation. Finally, we implemented SonarBeat with two different smartphones and conducted an extensive experimental study with three setups. The experimental results validated that SonarBeat can achieve superior performance on respiration rate estimation for different factors and parameters.

REFERENCES

Heba Abdelnasser, Khaled A. Harras, and Moustafa Youssef. 2015. Ubibreathe: A ubiquitous non-invasive WiFi-based breathing estimator. In *the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'15)*. 277–286.

Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart homes that monitor breathing and heart rate. In *the ACM CHI Conference on Human Factors in Computing Systems (CHI'15)*. 837–846.

Heba Aly and Moustafa Youssef. 2016. Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In *the IEEE International Conference on Computer Communications (INFOCOM'16)*. 1–9.

Inci M. Baytas, Kaixiang Lin, Fei Wang, Anil K. Jain, and Jiayu Zhou. 2016. PhenoTree: Interactive visual analytics for hierarchical phenotyping from large-scale electronic health records. *IEEE Trans. Multimedia* 18, 11 (Nov. 2016), 2257–2270.

Edgar A. Bernal, Xitong Yang, Qun Li, Jayant Kumar, Sriganesh Madhvanath, Palghat Ramesh, and Raja Bala. 2018. Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors. *IEEE Trans. Multimedia* 20, 1 (2018), 107–118.

Sergio Cicalo, Matteo Mazzotti, Simone Moretti, Velio Tralli, and Marco Chiani. 2016. Multiple video delivery in m-health emergency applications. *IEEE Trans. Multimedia* 18, 10 (Oct. 2016), 1988–2001.

Charalampos A. Dimoulas. 2016. Audiovisual spatial-audio analysis by means of sound localization and imaging: A multimedia healthcare framework in abdominal sound mapping. *IEEE Trans. Multimedia* 18, 10 (Oct. 2016), 1969–1976.

Amy Droitcour, Victor Lubecke, Jenshan Lin, and Olga Boric-Lubecke. 2001. A microwave radio for Doppler radar sensing of vital signs. In *Proc. 2001 IEEE MTT-S International Microwave Symposium Digest (Cat. No. 01CH37157)*. 175–178.

Amy D. Droitcour, Olga Boric-Lubecke, and Gregory T. A. Kovacs. 2009. Signal-to-noise ratio in Doppler radar system for heart and respiratory rate measurements. *IEEE Trans. Microw. Theory Technol.* 57, 10 (Oct. 2009), 2498–2507.

Biying Fu, Jakob Karolus, Tobias Grosse-Puppendahl, Jonathan Hermann, and Arjan Kuijper. 2015. Opportunities for activity recognition using ultrasound doppler sensing on unmodified mobile phones. In *Proc. 2nd International Workshop on Sensor-Based Activity Recognition and Interaction*. ACM, 8:1–8:10.

Pedro Henriquez, Bogdan J. Matuszewski, Yasmina Andreu-Cabedo, Luca Bastiani, Sara Colantonio, Giuseppe Coppini, Mario D'Acunto, Riccardo Favilla, Danila Germanese, Daniela Giorgi, et al. 2017. Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization. *IEEE Trans. Multimedia* 19, 7 (July 2017), 1467–1481.

Carl E. Hunt and Fern R. Hauck. 2006. Sudden infant death syndrome. *Can. Med. Assoc. J* 174, 13 (April 2006), 1309–1310.

S. M. Riazul Islam, Daehan Kwak, M. D. Humaun Kabir, Mahmud Hossain, and Kyung-Sup Kwak. 2015. The Internet of Things for health care: A comprehensive survey. *IEEE Access* 3 (June 2015), 678–708.

Fan Li, Huijie Chen, Xiaoyu Song, Qian Zhang, Youqi Li, and Yu Wang. 2017. CondioSense: High-quality context-aware service for audio sensing system via active sonar. *Personal Ubiquitous Comput.* 22, 1 (Feb. 2017), 17–29.

Jian Liu, Yan Wang, Yinging Chen, Jie Yang, Xu Chen, and Jerry Cheng. 2015a. Tracking vital signs during sleep leveraging off-the-shelf WiFi. In *the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'15)*. 267–276.

Jian Liu, Yan Wang, Gorkem Kar, Yingying Chen, Jie Yang, and Marco Gruteser. 2015b. Snooping keystrokes with mm-level audio ranging on a single phone. In *the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 142–154.

V. M. Lubecke, O. Boric-Lubecke, G. Awater, P. W. Ong, P. Gammel, R. H. Yan, and J. C. Lin. 2000. Remote sensing of vital signs with telecommunications signals. In *Proc. World Congress on Medical Physics and Biomedical Engineering (WC'00)*.

Victor Lubecke, Olga Boric-Lubecke, and Eric Beck. 2002. A compact low-cost add-on module for Doppler radar sensing of vital signs using a wireless communications terminal. In *Proc. 2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No. 02CH37278)*. 1767–1770.

Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-precision acoustic motion tracking. In *the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'16)*. ACM, New York, NY, 491–492.

L. Robert Mogue and Boerje Rantala. 1988. Capnometers. *J. Clin. Monit.* 4, 2 (April 1988), 115–121.

Frank Mokaya, Roland Lucas, Hae Young Noh, and Pei Zhang. 2016. Burnout: A wearable system for unobtrusive skeletal muscle fatigue estimation. In *the 15th International Conference on Information Processing in Sensor Networks (IPSN'16)*. IEEE, 1–12.

Seong-Eun Moon and Jong-Seok Lee. 2017. Implicit analysis of perceptual multimedia experience based on physiological response: A review. *IEEE Trans. Multimedia* 19, 2 (Feb. 2017), 340–353.

Henning Müeller and Devrim Unay. 2017. Retrieval from and understanding of large-scale multi-modal medical datasets: A review. *IEEE Trans. Multimedia* 19, 9 (Sept. 2017), 2093–2104.

Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, Venkat Padmanabhan, and Ramarathnam Venkatesan. 2013. Dhwani: Secure peer-to-peer acoustic NFC. *ACM SIGCOMM Comput. Commun. Rev.* 43, 4 (2013), 63–74.

Rajalakshmi Nandakumar and Shyamnath Gollakota. 2017. Unleashing the power of active sonar. *IEEE Pervasive Comput.* 16, 1 (2017), 11–15.

Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *the 13th International Conference on Mobile Systems, Applications, and Services (MobiSys'15)*. ACM, 45–57.

Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *the ACM CHI Conference on Human Factors in Computing Systems (CHI'16)*. ACM, 1515–1525.

Phuc Nguyen, Xinyu Zhang, Ann Halbower, and Tam Vu. 2016. Continuous and fine-grained breathing volume monitoring from afar using wireless signals. In *the IEEE International Conference on Computer Communications (INFOCOM'16)*. 1–9.

Chen Qiu and Matt W. Mutka. 2017. Silent whistle: Effective indoor positioning with assistance from acoustic sensing on smartphones. In *the 18th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'17)*. IEEE, 1–6.

Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen. 2015. Fine-grained sleep monitoring: Hearing your breathing with smartphones. In *the IEEE International Conference on Computer Communications (INFOCOM'15)*. 1194–1202.

Wenjie Ruan, Quan Z. Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal. In *the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'16)*. ACM, 474–485.

Jussi Salmi and Andreas F. Molisch. 2011. Propagation parameter estimation, modeling and measurements for ultrawideband mimo radar. *IEEE Trans. Microw. Theory Technol.* 59, 11 (Nov. 2011), 4257–4267.

Nastaran H. Shariati and Edmond Zahedi. 2005. Comparison of selected parametric models for analysis of the photoplethysmographic signal. In *Proc. 1st IEEE Conf. Comput., Commun. Signal Process.* 169–172.

Dan Stowell, Dimitrios Giannoulis, Emmanouil Benetos, Mathieu Lagrange, and Mark D. Plumbley. 2015. Detection and classification of acoustic scenes and events. *IEEE Trans. Multimedia* 17, 10 (Oct. 2015), 1733–1746.

Junjue Wang, Kaichen Zhao, Xinyu Zhang, and Chunyi Peng. 2014. Ubiquitous keyboard for small mobile devices: Harnessing multipath fading for fine-grained keystroke localization. In *the 12th International Conference on Mobile Systems, Applications, and Services (MobiSys'14)*. ACM, 14–27.

Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'16)*. ACM, New York, NY, 82–94.

Xuyu Wang, Runze Huang, and Shiwen Mao. 2017a. SonarBeat: Sonar phase for breathing beat monitoring with smartphones. In *the 26th International Conference on Computer Communication and Networks (ICCCN'17)*. 1–8.

Xuyu Wang, Chao Yang, and Shiwen Mao. 2017b. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In *the 37th IEEE International Conference on Distributed Computing Systems (ICDCS'17)*. 1–10.

Xuyu Wang, Chao Yang, and Shiwen Mao. 2017c. TensorBeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi. *ACM Trans. Intell. Sys. Technol.* 9, 1 (Sept. 2017), 8:1–8:27.

Xuyu Wang, Choa Yang, and Shiwen Mao. 2020. On CSI-based vital sign monitoring using commodity WiFi. *ACM Trans. Comput. Healthcare* 1, 3 (April 2020), 12:1–12:27.

Teng Wei and Xinyu Zhang. 2015. mTrack: High-precision passive tracking using millimeter wave radios. In *the ACM Annual International Conference on Mobile Computing and Networking (MobiCom'15)*. ACM, 117–129.

Cheng Yang, Gene Cheung, and Vladimir Stankovic. 2017. Estimating heart rate and rhythm via 3D motion tracking in depth video. *IEEE Trans. Multimedia* 19, 7 (July 2017), 1625–1636.

Zhicheng Yang, Parth H. Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2016. Monitoring vital signs using millimeter wave. In *the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'16)*. 211–220.

Xingliang Yuan, Xinyu Wang, Cong Wang, Jian Weng, and Kui Ren. 2016. Enabling secure and fast indexing for privacy-assured healthcare monitoring via compressive sensing. *IEEE Trans. Multimedia* 18, 10 (Oct. 2016), 2002–2014.

Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a mobile device into a mouse in the air. In *the 13th International Conference on Mobile Systems, Applications, and Services (MobiSys'15)*. 15–29.

Bing Zhou, Mohammed Elbadry, Ruipeng Gao, and Fan Ye. 2017. BatMapper: Acoustic sensing based indoor floor plan construction using smartphones. In *the 15th International Conference on Mobile Systems, Applications, and Services (MobiSys'17)*. ACM, 42–55.

Received July 2019; revised October 2020; accepted November 2020