Contact-free Breathing Rate Monitoring with Smartphones: A Sonar Phase Approach

by

Runze Huang

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
Aug 04, 2018

Keywords: Channel State Information; health sensing; healthcare Internet of Things; respiration rate monitoring; smartphone programming.

Approved by

Shiwen Mao, Samuel Ginn Endowed Professor of Electrical and Computer Engineering
Thaddeus Roppel, Associate Professor of Electrical and Computer Engineering
Xiaowen Gong, Associate Professor of Electrical and Computer Engineering

Abstract

Vital sign (e.g., respiration rate) monitoring has become increasingly more important because it offers useful clues of medical conditions such as sleep disorders. There is a compelling need for technologies that enable contact-free and easy deployment vital sign monitoring over extended period of time for healthcare. In this paper, 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.

Acknowledgments

Firstly, I am very grateful to my advisor Prof. Shiwen Mao for his guidance. He inspired me a lot whenever I had problem with my research work. I also appreciate my other committee members, Prof. Thaddus Roppel, and Prof. Xiaowen Gong. They gave me lots of suggestions and helpful comment, so that I can modify and improve my thesis research.

Besides these professors, I would like to thank my team members, particularly, Xuyu Wang, who helped me a lot in my research field. I can't make a success in my research and other projects without his help.

Last but no the least, I want to thank my family members and friends, who give support with their love during my whole study in Auburn.

This work is supported in part by the US NSF under Grant CNS-1702957, and through the Wireless Engineering Research and Education Center (WEREC) at Auburn University.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Auburn Auburn University

LoA List of Abbreviations

Chapter 1

Introduction

With the rapid development of mobile techniques and the growth in the living standard, healthcare has become one of the main application areas for Internet of Things (IoT) [**?**, 1, 2, 3]. The healthcare IoT architecture mainly consists of three layers: (i) the sensing layer for monitoring vital signals, such as body temperature, heart rate, respiration rate, and blood pressure; (ii) the gateway layer for collecting data from the sensing layer, and transmitting them to the third layer, the cloud layer [4]; (iii) the cloud layer consisting of data centers in the cloud to store, process, and analyze multi-modal medical datasets [5, 6, 7, 8, 9], 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) [10]. Traditional systems in the sensing layer require a person to wear special devices, such as a pulse oximeter [11] or a capnometer [12] to monitor breathing rates, which are not convenient for monitoring vital signals for the 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. The existing techniques can be classified into (i) radar based and (ii) WiFi based approaches. Examples of radar based vital sign monitoring include Doppler radar [13, 14], ultra-wideband radar [15], frequency modulated continuous wave (FMCW) radar [16], all of which require a piece of

customized hardware working on high frequency. With the development of wireless techniques [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36], WiFi based methods include UbiBreathe [37] and mmVital [38], which exploit the received signal strength (RSS) of 2.4 GHz WiFi and 60 GHz millimeter wave (mmWave) signals (i.e., 802.11ad), respectively. Recently, the authors in [39] employ the amplitude of WiFi CSI data to track vital signs for a sleeping person, while our prior works PhaseBeat [40] and TensorBeat [41] 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 [42].

To this end, the smartphone can serve as an excellent platform for vial sign monitoring, by exploiting its built-in sensors, such as accelerometer, gyroscope [43], and microphone [44]. 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. In a recent work [45], the authors propose to use an active sonar built in 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 a continuous-wave (CW) radar to measure distance and achieves device-free hand tracking using sonar phase information [46]; while [47] uses the phase of acoustic orthogonal frequency-division multiplexing (OFDM) signals for finger tracking.

1.1 Approach

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 breathing rate with a high accuracy. Compared with other existing systems such as Doppler

shift and FMCW [45], 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 breathing rate with the same frequency. Built upon analysis, we design SonarBeat, an active **Sonar** phase for **Breat**hing rate monitoring system with a smartphone. 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 18-22 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 a Android smartphone and validate its performance with extensive experiments that involve five persons over a period of three 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 beats 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 paper include the 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 on using 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 with comparison with an existing scheme in three different indoor scenarios. Our extensive experimental results demonstrate the superior performance of SonarBeat under different environment factors and different experimental parameters.

1.2 Layout

In the remainder of this paper, 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 paper.

Chapter 2

Related Work

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

2.1 Audible Signal with Smartphone based Sensing Systems

Mobile sensing systems with audible signals have attracted great attention [48, 49]. CondioSense system provides a method to identify the current phone context and detect the microphone sensing states. The system includes two parts with microphone state sensing detection (MSS) and Context detection. MSS detection triggers the vibrator for a few seconds and the microphone records those waves bounced back from the environment. Further, extracting pre-designed features from the recorded signal which are preprocessed with a band-pass filter, framing will be used for machine learning training, in order to estimate the microphone sensing state. From Fig. reffig:CondioSense, the smart phone speaker emits chip sound signal and records it by microphone after it travels through the context. Also, a band-pass filter is used to remove background noise. After that, cross-correlation and auto-correlation methods are used for features extraction. At last, context information can be detected by the machine learning method with the extracted features. Such systems are increasingly convenient for people's life 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 [50]. In Fig. 2.2, like continuous

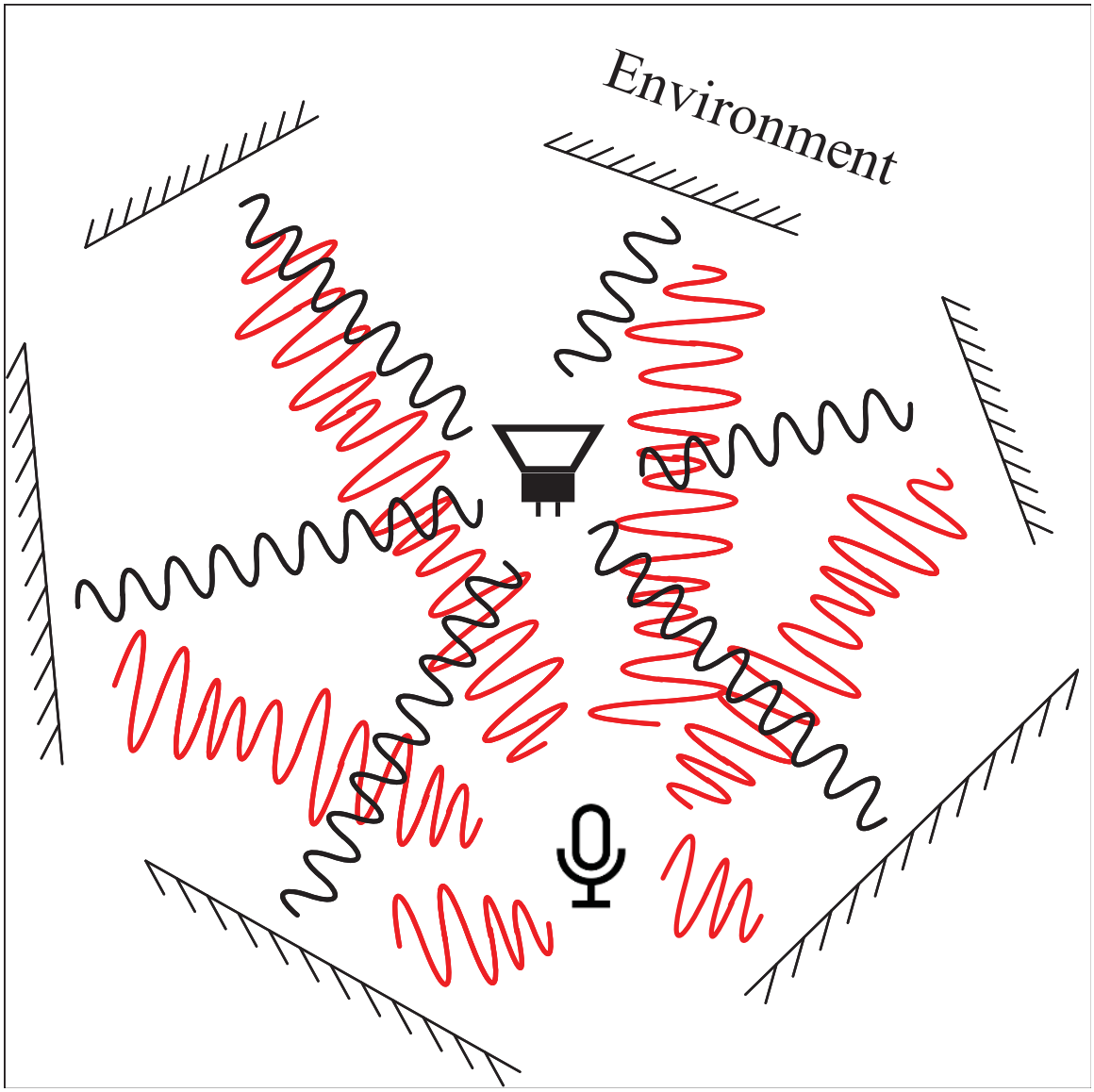

Figure 2.1: CondioSense system.

wave radar, a sine-wave with ultrasound frequency is sent out from the devices speaker. With the phenomenon of Doppler shift of frequency, an approaching target can increase the signal frequency and a departing signal can decrease the signal frequency, The smartphone can use those information to estimate the moving speed, direction and range of the target.

The recent work 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 [51]. The basic principle of AAmouse system is Doppler shift effect of the frequency of the signal in Fig. 2.3. Differently, AAmouse system separates the microphone and the speaker to two different devices. In other words, using two speakers in the other device,

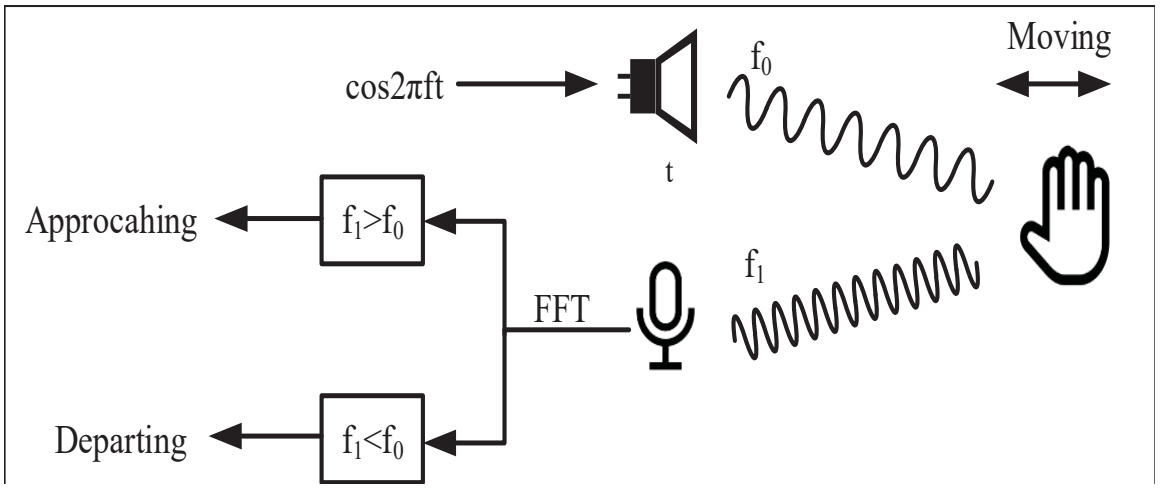

Figure 2.2: Opportunities for activity recognition using ultrasound Doppler sensing on unmodified mobile phones.

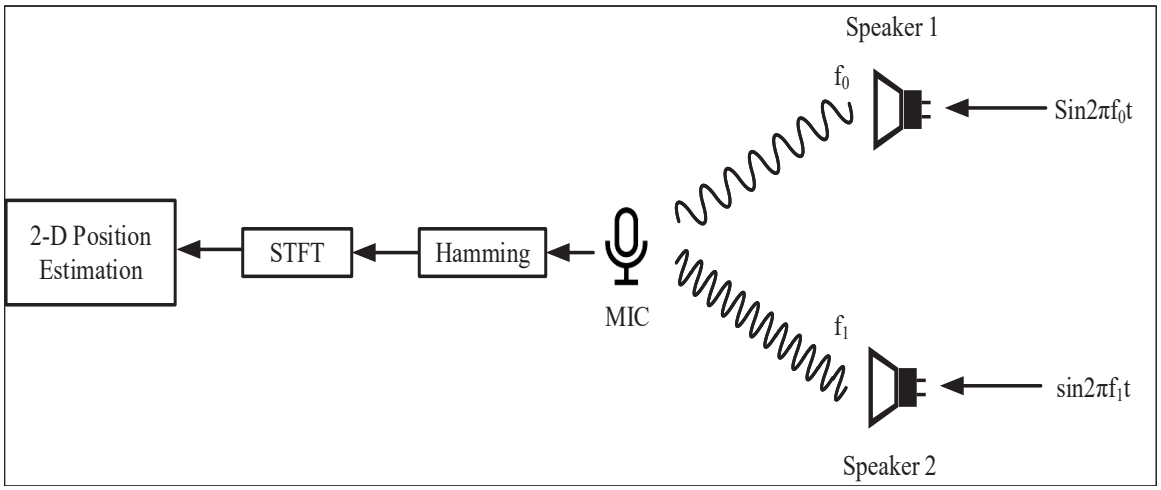

Figure 2.3: AAmouse system.

like TV, the sender emits different frequency audio signals, and the microphone in smartphone is considered as receiver. After receiving those signals, they use Short-term Fourier Transform (STFT) to analyze the change of spectrum over time. Finally, by using the distance between two speakers and the estimated distance between the microphone and speakers, the system can track the position of the smartphone in a 2-D platform with the triangulation method.

Moreover, the CAT system implements a distributed FMCW for tracking devices, such as VR/AR headsets. In Fig. 2.4, 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 [52].

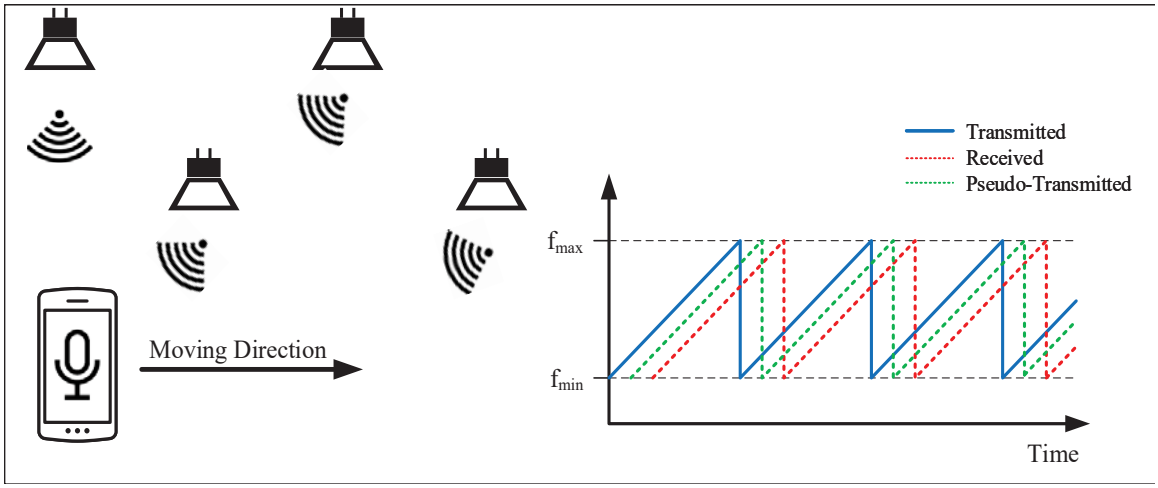

Figure 2.4: CAT.

In addition, audible based sensing technique is also used for wireless virtual keyboard with smartphones. Keystroke snooping [53] and Ubik systems [54] 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 stroke on the table. For Ubik system, When a position is touched, a spike of sound will be recorded in the signal strength in Fig. reffig:Ubik. In fact, the spike may not be received by the two microphones in the same time. Thus, leveraging TDOA of these two spikes can get the distance between the keystroke point and microphones. With the given distance of two microphones, the triangulation method should be used for keystroke localization. Then, the strokes are transformed into related alphabets in the same position as a computer keyboard. SilentWhistle is a light-weight indoor localization system using acoustic sensing for obtaining users' locations [55]. Another work Dhwani builds an acoustics-based NFC system with smartphones, using a technology JamSecure. It can provide a secure communication channel between devices, which is an OFDM channel for audible signals [56].

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 [57]. OFDM based sensing systems such as FingerIO can track the finger movement in a 2-D domain through tracking echoes from the finger that is received by the microphone, to measure the finger position in Fig 2.6. FingerIO splits the frequency range of 0 - 24kHz into 64 subcarriers. Subcarriers within the range of 18-20kHz are set to either +1 or -1, and the rest of the frequency range is set as 0. The speaker input is only allowed as

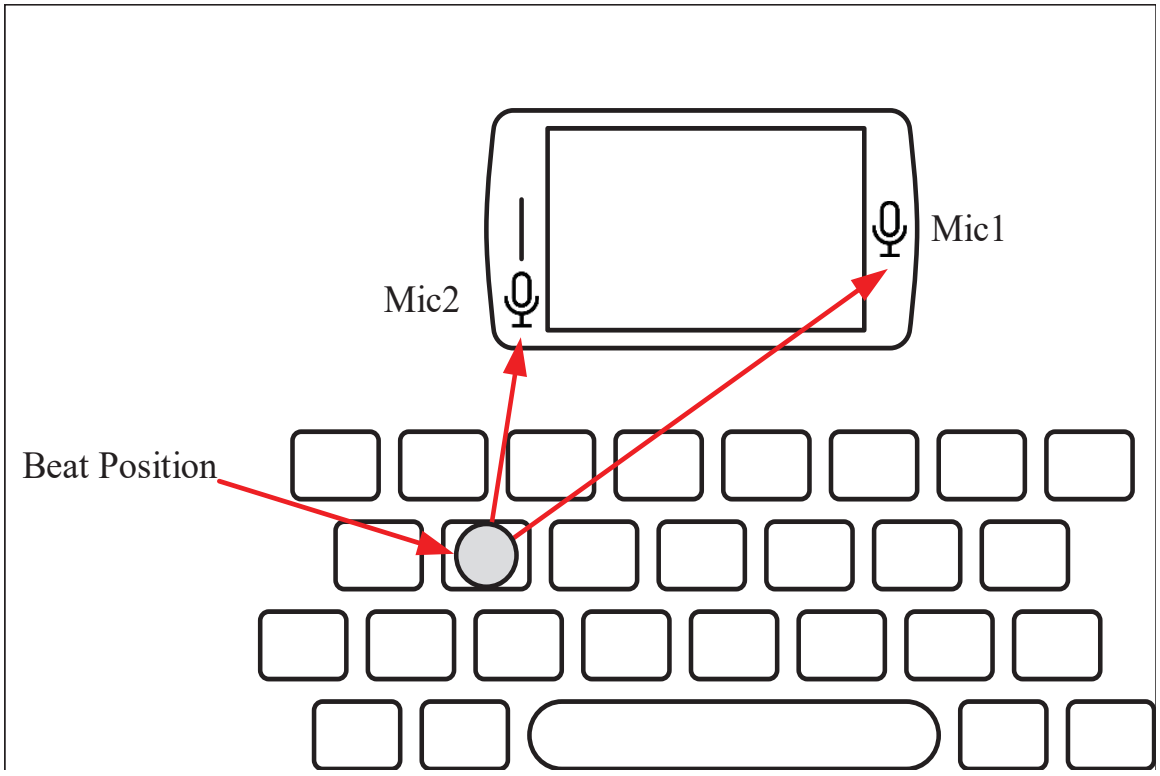

Figure 2.5: Ubik.

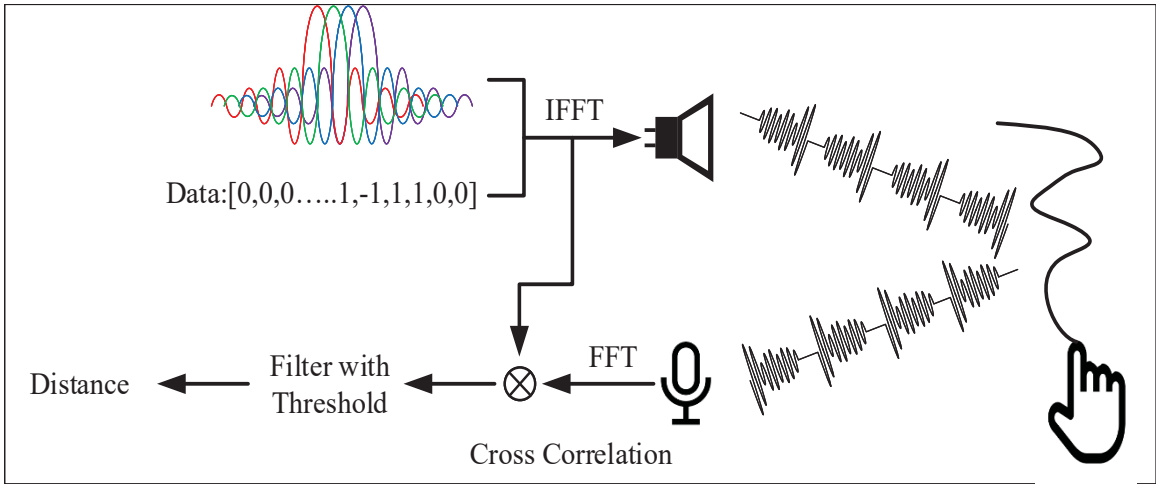

Figure 2.6: FingerIO.

16-bit real number. After computing the IFFT of the transmitting data, those complex numbers should be recalculated to it real representation. Furthermore, the received signal should be correlated with the transmitting signal to create the echo profile. After that, the system can use the threshold to distinguish the echo from the finger with others. The distance between the smartphone and the finger can be estimated. In fact, the system can compute the distance from the finger to two microphones separately and then combine them for the 2-D localization.

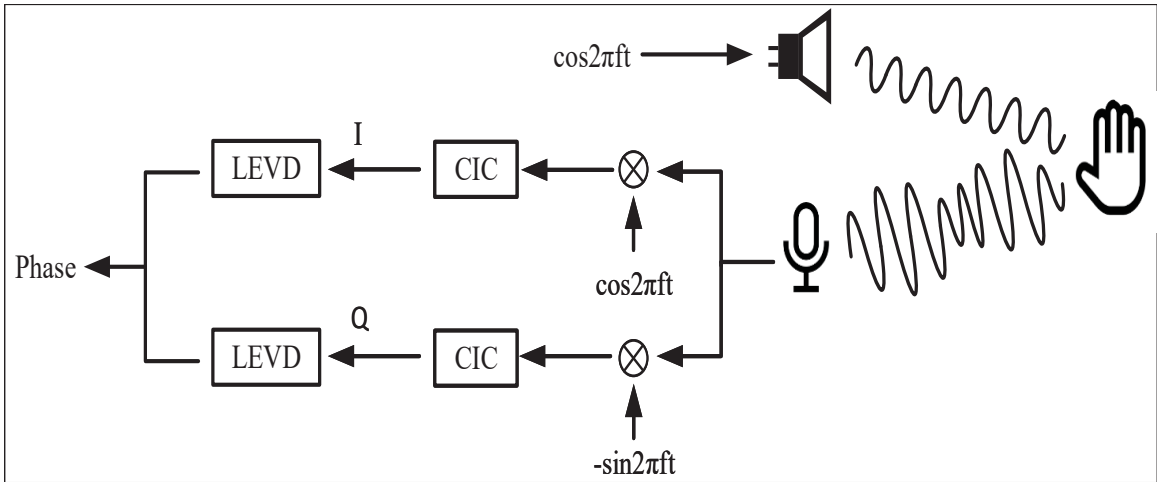

Figure 2.7: LLAP.

Another work BatMapper employs an acoustic sensing based system for fast and accurate floor plan construction using commodity smartphones [58].

Moreover, LLAP system leverages the principal of CW radar to measure distance and implement device-free hand tracking [46]. A cosine wave signal in the frequency range from 17kHz to 23kHz is emitted from the speaker of Smartphone. After that the signal is bounced back from the target, and then LLAP system uses the principle of I/Q-demodulation for extracting phase information in Fig **??**. Instead of a FIR filter, LLAP system uses CIC filter to eliminate the high frequency component formed in demodulation. However, the baseband signal still has a static vector from the sound wave traveling through the LOS path. Also, static object in the environment will heavily impact the estimation of the phase change. LLAP system introduces a heuristic algorithm called Local Extreme Value Detection (LEVD) to estimate the static vector and subtract it from I/Q component separately. After removing the static vector, phase change can be acquired from I divided by Q. This work is closely related to SonarBeat, because both systems use the phased based CW signal to sense movements. The difference between two systems is that SonarBeat is more robust to different environments with the adaptive median filter technique. Moreover, the proposed SonarBeat system is used for breathing rates monitoring, while LLAP system is employed for hand tracking.

On the other hand, the AudioGest system employs a pair of built-in speaker and microphone to send inaudible sound and leverage the echos to sense the hand movement [59]. Our

SonarBeat system is motivated by the active inaudible sensing systems to transform a smartphone into an active sonar with ultrasonic sound waves.

2.2 Audible Signal with Smartphone Based Health Systems

Mobile health applications and research have become an important part of the IoT [60]. Smartphones and other wearable devices can provide people with a more convenient way to monitor their health conditions without the need of professional equipment [61]. The recently work Burnout leverages accelerometers to sense skeletal muscle vibrations, which does not require to wear a suit embedded with sensors [62]. The authors in [63] 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 [62]. For vital signal monitoring, fine-grained sleep monitoring using a microphone in the earphone can record human breathing sound to monitor people's heath signal when they are sleeping [44], which adopts a passive audible method. The Apnea system uses an active sonar with smartphone to monitor the breathing signal [45]. This work leverages the FMCW technique for breathing monitoring, which requires the system to seek the distance between the smartphone and the chest of the person.

2.3 Signal Processing for Acoustic signal

2.3.1 CW Radar

CW radar is a radar system which can continuously transmit a high frequency signal through the air and process it at the same device in the meantime. In fact, the speed or the moving range of the target can be estimated with Doppler shift of the bounced signal in Fig. 2.8. A change in the frequency of the transmitted signal will be produced when it meets a moving object. Thus,

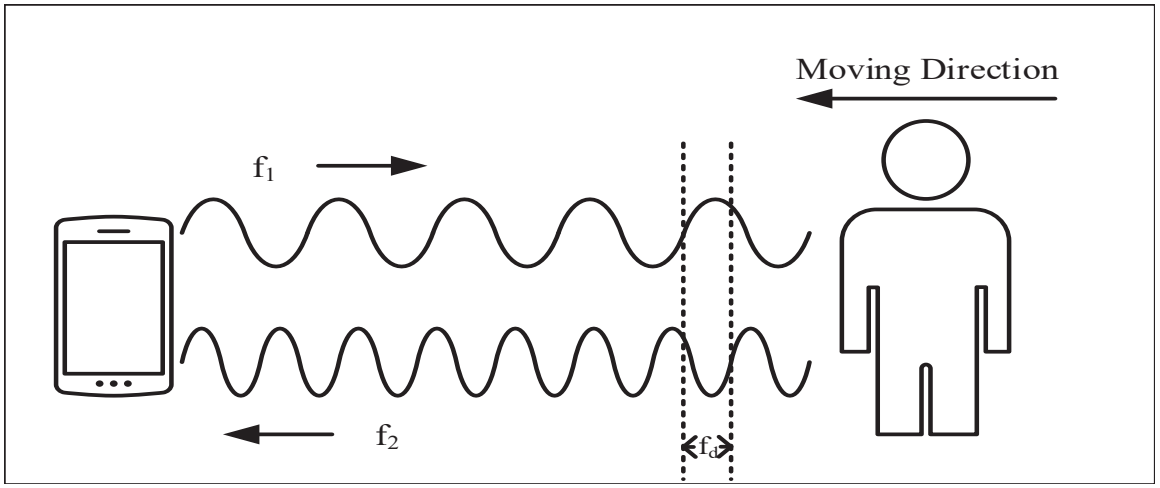

Figure 2.8: Doppler shift.

its moving speed can be estimated by

$$v = (c * f_d)/(2f_t + f_d) \tag{2.1}$$

where f_d is the Doppler frequency, which is also the frequency difference between the transmitting and received signals, c is the speed of sound, f_t is the transmitting frequency. Thus, the moving range of the mobile can be calculated by the moving speed multiplied by the speed of sound.

Due to the natural characteristic, CW radar is much simpler for manufacture and operate. Furthermore, there is not any distance limitation for the measurement range for CW radar. In other words, as long as the signal has sufficient power, it can work in a distance without limitation. However, basic CW radar cannot measure the distance between the source to the target. This is because it lacks the timing mark to precisely determine the transmitting and received cycle, and cannot thus convert it into the range. Moreover, small objects will be overwhelmed by large objects with high possibility and become undetectable in a scenario which exists moving objects Generally, CW radar can be used for speed detection in a scenario in which most objects are stationary.

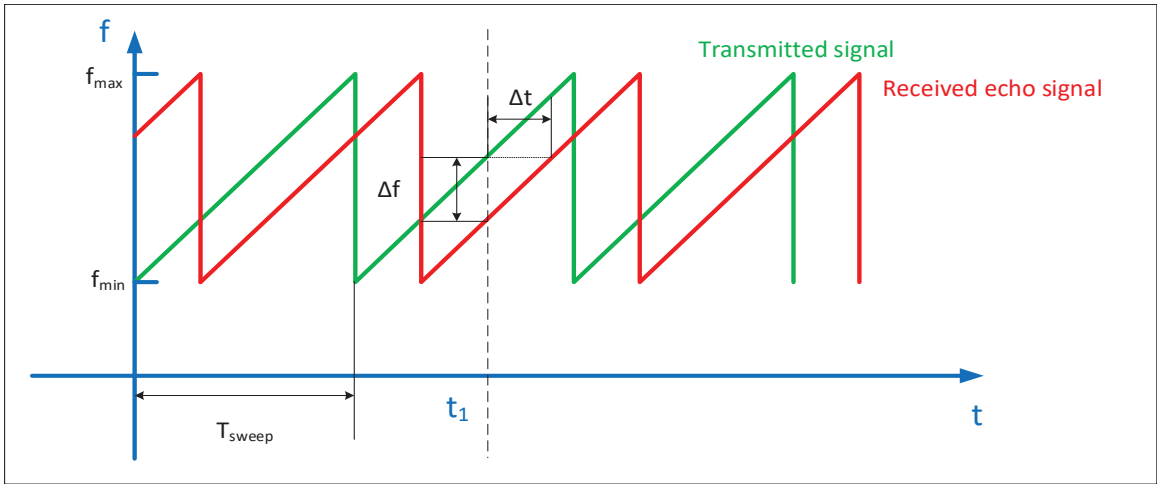

Figure 2.9: FMCW.

2.3.2 FMCW Radar

FMCW radar is a pulsed and modulated continuous wave radar, where transmitting signal will be modulated in frequency increasing or decreasing periodically in Fig. 2.9. Technically, FMCW radar will have continuous radiate modulated power. When an echo is received, comparing with the emitting signal which has the same order of the received signal, a time delay will show in the same frequency. With the speed of sound signal for electromagnetic wave, the distance between the source and the target is predicted. It can be calculated by

$$R = (C_0 * |\Delta t|)/2 \tag{2.2}$$

where C_0 is the speed of sound (343 m/s), Δt is the time difference between the received signal and the transmitting signal in the same frequency. Basically, FWCW radar only can measure distance in a small range, however, in a high accuracy. The maximum non-ambiguous range is determined by the duration of the increasing of the frequency. And the accuracy and the range resolution are highly related to the bandwidth of the transmitted signal. Thus, a higher accuracy is obtained, when the maximum detection range is not too large. Like unmodulated CW radar, FMCW can estimate the target velocity and moving range simultaneously. However, based on the physical principle of FMCW radar, it will continuously transmit a frequency band, which is more susceptible to the interference from other electronic systems in some specific frequency.

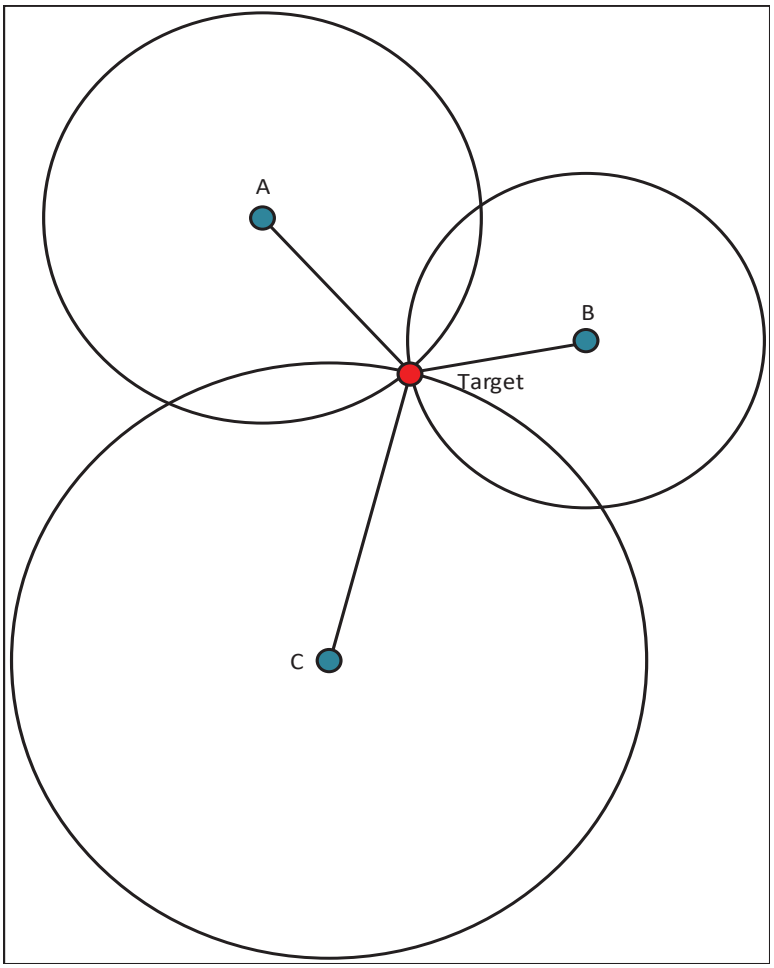

Figure 2.10: Triangulation Positioning.

2.3.3 Time of Arrival

Time of Arrival (TOA) is the most common technology used for localization, such as Global Positioning System(GPS). Knowing the exact time of transmitting and the time of arrival is necessary to estimate the distance from the source to the target. For localization, using the triangulation method, the target will collect distance information from at least three beacons, A , B and C, e.g. satellites in GPS system. After that, using those distances from the target to beacons as different radius, we can draw at least three cycles. The point of intersection will be used for estimating the location of the target, which is marked as the red dot in the middle in Fig 2.10. Also, TOA is used for indoor localization which is implemented by audio signals that are emitted by smartphone speakers and are received by microphones as beacons in the room. However, a prevail design of modern smartphone consists of one speaker and two microphones,

high localization accuracy is restricted under such structure. For example, in a wide place, like arena, the energy of transmitting signal is not enough to reach an obstacle and bounce back. Considering this issue, corridor and small room are the best application scenes for smartphone based indoor localization with TOA technology.

In fact, the accuracy of TOA is highly restricted by the time synchronization of all devices. Especially, in outdoor localization, a microsecond level error will cause 300-meter error in the ground. Normally, network-based clock synchronization is required in modern localization systems.

Chapter 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 18–22 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 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 is 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.

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 I-component and Q-component of a baseband signal in Fig. 3.1. 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 signal. For SonarBeat system, we mainly exploit the phase information extracted from 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.

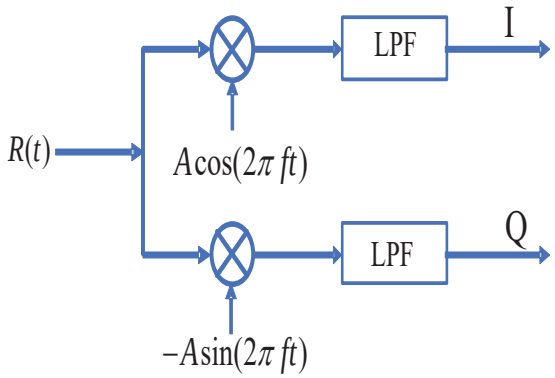

Figure 3.1: I/Q demodulation for the received signal.

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 the 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

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

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 $R(t) = A_r \cos(2\pi f t - 2\pi f d(t) - \theta)$, 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 $C(t) = A\cos(2\pi f t)$, we have

$$\begin{aligned} & A_r \cos(2\pi f t - 2\pi f d(t) - \theta) \times A\cos(2\pi f t) \\ = & \frac{A_r A}{2}(\cos(4\pi f t - 2\pi f d(t) - \theta) + \cos(-2\pi f d(t) - \theta)). \end{aligned} \tag{3.1}$$

The first term in (3.1) 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 $I = \frac{A_r A}{2}\cos(-2\pi f d(t) - \theta)$.

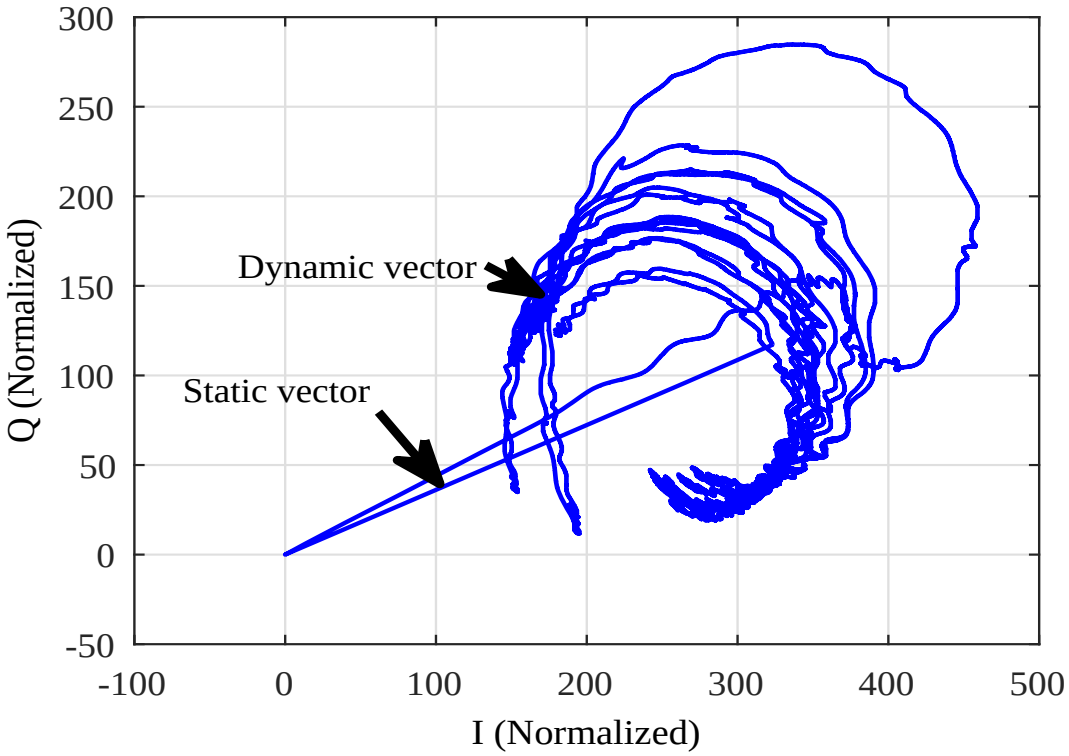

Figure 3.2: Complex I/Q traces of the received audio signal.

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 $Q = \frac{A_r A}{2} \sin(-2\pi f d(t) - \theta)$. We then demodulate the phase of the inaudible signal data as

$$\begin{aligned} \varphi(t) &= \arctan(Q/I) = -2\pi f d(t) - \theta \\ &= -2\pi f (D_0 + D\cos(2\pi f_b t))/c - \theta. \end{aligned} \tag{3.2}$$

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, Fig. 3.2 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. Fig. 3.3 shows the complex I/Q traces of the received audio signal after removing static vector. It is noticed that the demodulated phase is a good indicator of the breathing caused chest movements.

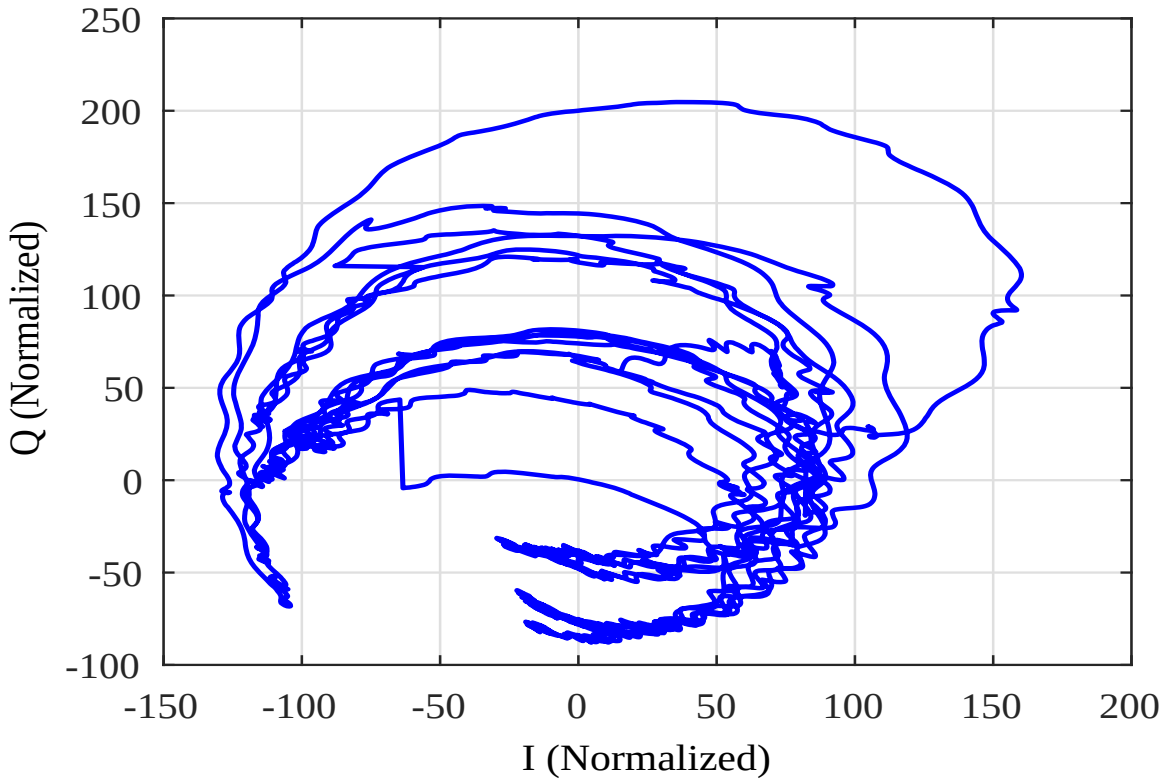

Figure 3.3: Complex I/Q traces of the received audio signal after removing the static vector effect.

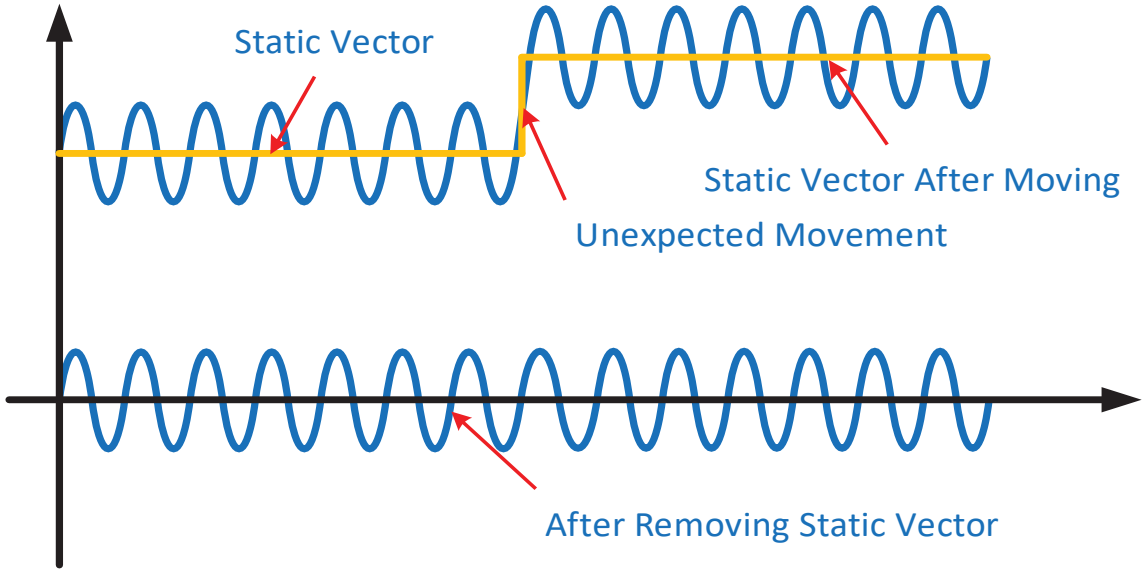

Figure 3.4: Illustration of adapting to body movements by eliminating the static vector.

3.2 Teacnical Challenges

Mitigate the Static Vector Effect

The main challenge for respiration monitoring with phase modulated data is to mitigate 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. In [46], the authors 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 paper, we propose an adaptive median filter method to deal with this challenge, which is shown effectively for removing the stationary component in different scenarios.

Adapt to Body Movements and Environment Noise

The second challenge for tracking breathing signals is adapt to body movements and environment noise. Body movement is unavoidable in the monitoring phase, e.g., during sleep, and its impact should be mitigated. The FMCW based scheme in [45] requires to estimate 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 on adapting to body movements. For example, Fig. 3.4 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 Fig. 3.4.

Furthermore, consider the case of multiple persons in the testing environment. Not only 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 environmental noise from external audio sources.

Realtime Monitoring with Lower Delay

For a vital sign monitoring system to be really useful, it should work in realtime, with good interactions with the user. Realtime 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 base band. 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 realtime 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.

Chapter 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 a 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 [16], the phase based approach has a lower latency and complexity. Finally, the sonar phase data is robust to different orientations, different distances, different cloth thickness, 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.

Fig. 4.1 presents the SonarBeat system architecture, which includes four basic modules: (i) Signal Generation, (ii) Data Extraction, (iii) Received Signal Preprocessing, and (iv) 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 (i) I/Q demodulation, which implements downsampling followed by a coherent phase detector; (ii) static vector effect reduction, which implements the proposed adaptive median filter method to remove the static vector of the In-phase and Quadrature signals; (iii) phase extraction, where the sonar phase information

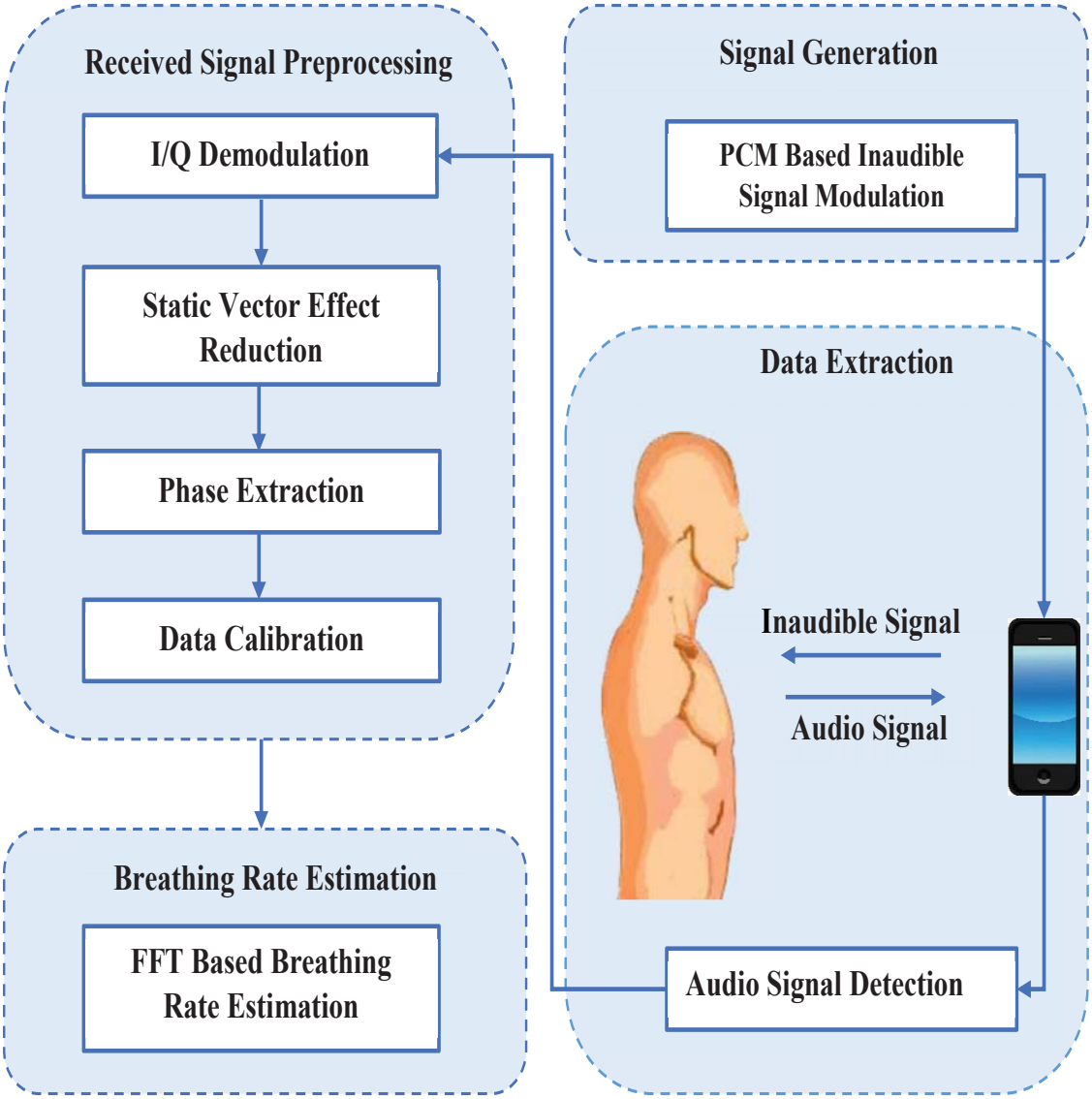

Figure 4.1: The SonarBeat system architecture.

is extracted and calibrated; and (iv) 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 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 18–22 KHz in the form of a CW signal, i.e., $C(t) = A\cos(2\pi ft)$. 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 at uniform intervals, where each sampled value is quantized. The PCM based inaudible signal modulation is implemented with the AudioTrack class.

4.3 Signal Generation

We use the microphone of the smartphone to receive the inaudible signal reflected from the chest with a sampling rate of 48 KHz. The microphone will record other sound signals with different frequencies from the surrounding environment as well. 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 of the signal, there will be a drastic of power at the carrier frequency. A threshold based method is used to detect the beginning of the inaudible signal. Fig. 4.2 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.

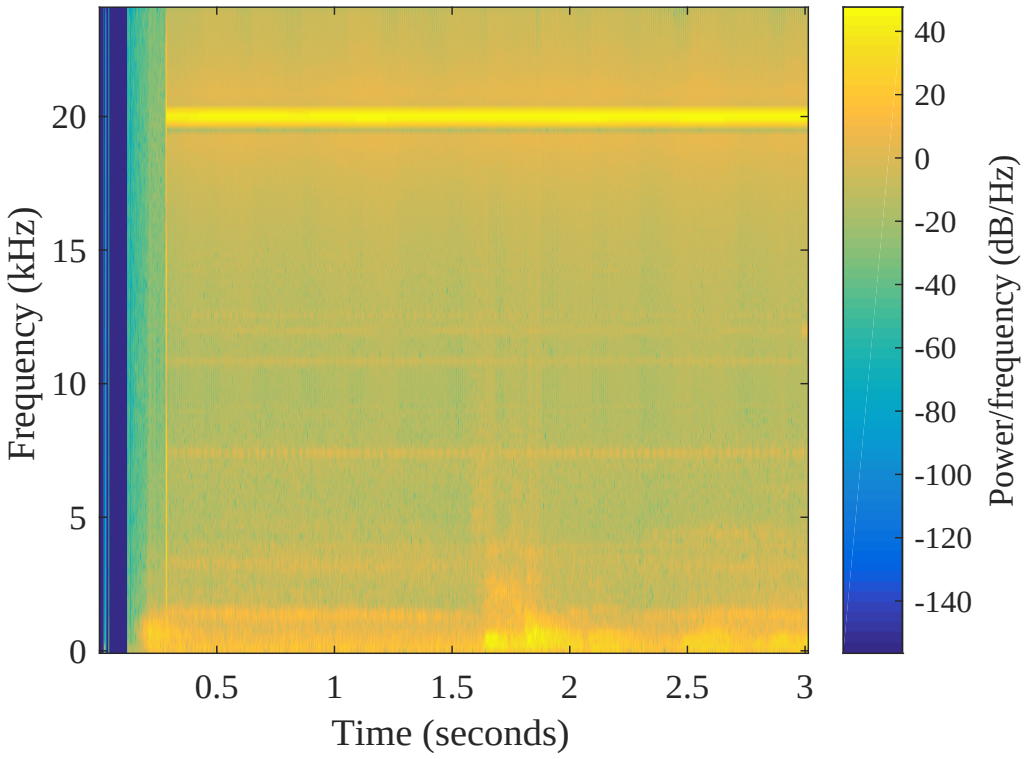

Figure 4.2: STFT based method for audio signal detection.

I/Q Demodulation

Before I/Q Demodulation, we need to down-sample the received signal $R(t) = A_r \cos(2\pi ft - 2\pi fd(t) - \theta)$ for reducing computation complexity, which is necessary for realtime monitoring. The original system with sampling frequency of 48 KHz is reduce to 480 Hz with a down-sampling ratio 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 reduce the number of samples of the amplitude of breathing signal, rather than the phase information.

Finally, a 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 Figs. 4.3 and 4.4, we plot the raw I-component and Q-component of the

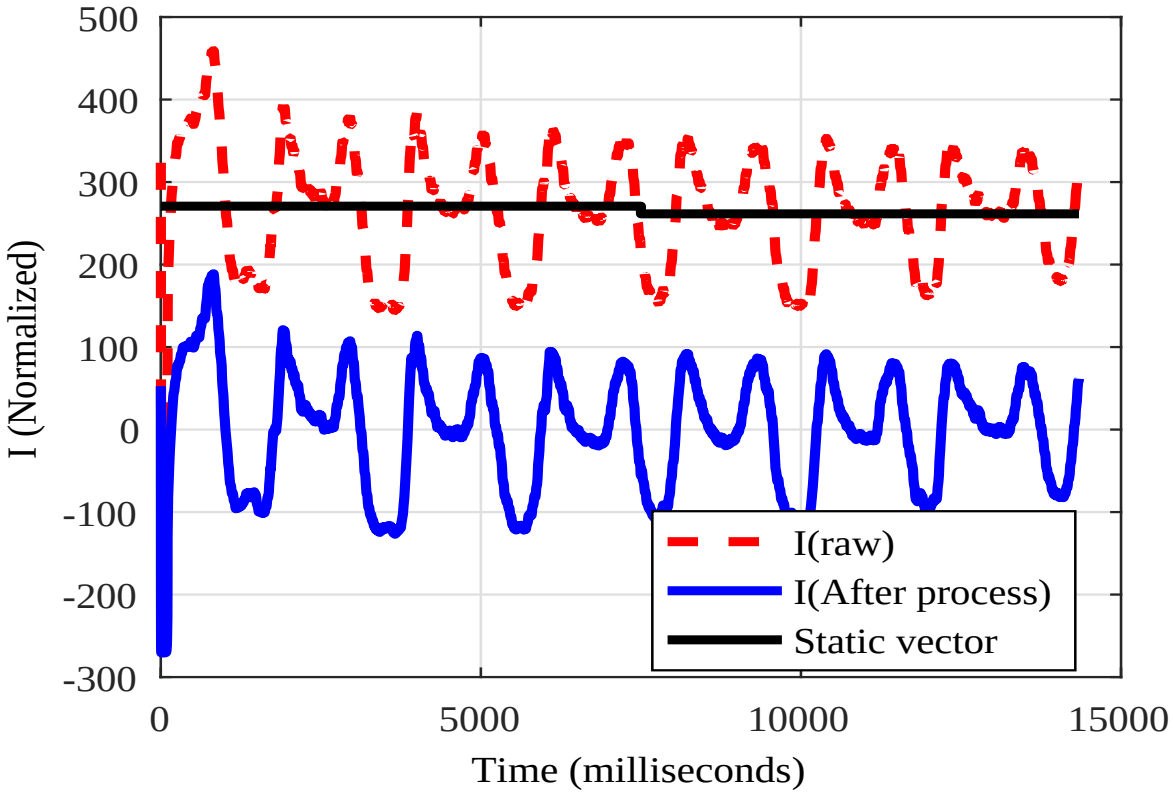

Figure 4.3: The adaptive median filter for removing the static vector in the baseband I-component.

baseband signal, respectively, after the LPF (the dashed curves), which, however, still include their static vectors.

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 representing the small chest movements. It is difficult to detect the weak breathing signal if we directly use the received sound signal. Recently, there are two methods proposed for static vector effect mitigation. The Dual-Differential Background Removal approach is used for hand tracking with 60 GHz mmWave signals [64]. The method is susceptible to environment noise and has large latency, which is not effective for realtime respiration monitoring. The second scheme, termed LEVD [46], is 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.

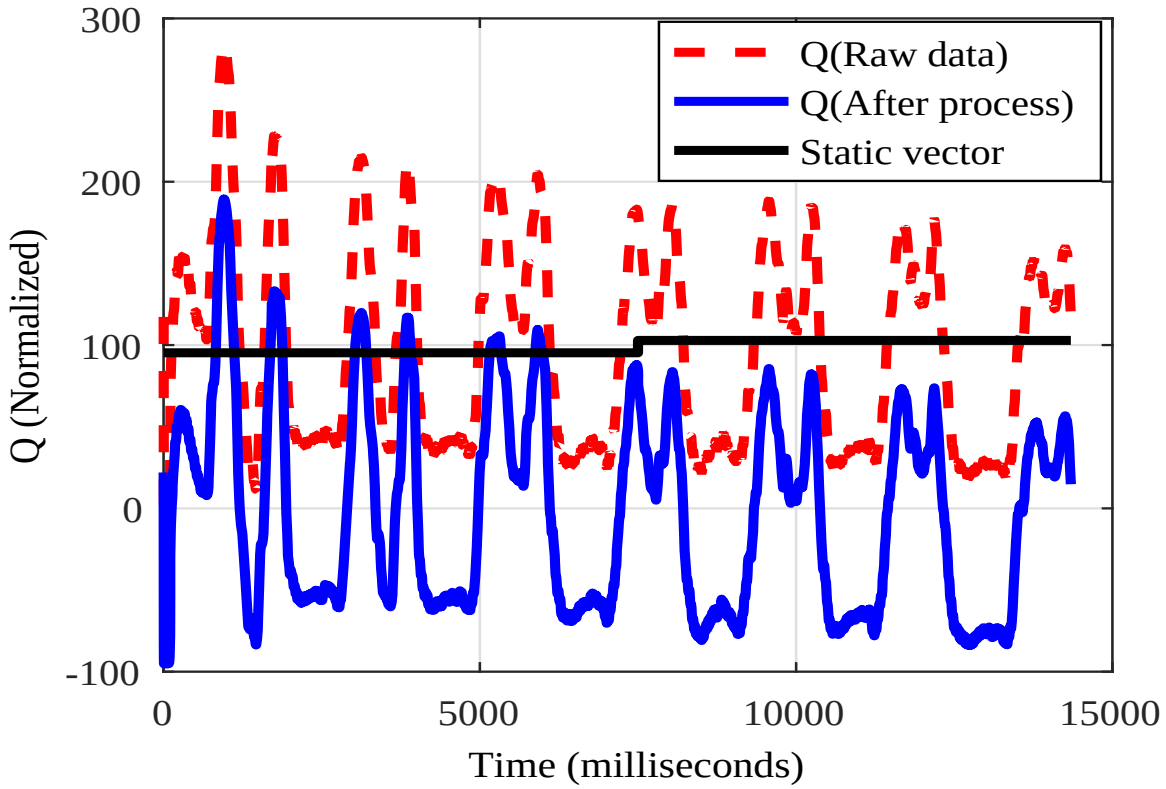

Figure 4.4: The adaptive median filter for removing the static vector in the baseband Q-component.

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 $Q(n)$, $n = 0, 1, ..., N-1$, we partition it into multiple non-overlapping sublists, each denoted by $W[1, 2, ..., w]$) with window size w, and a single sublist $R[1, 2, ..., r]$ 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 $W[1, 2, ..., w]$ and sublist $R[1, 2, ..., r]$ 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 $O(n)$, $n = 0, 1, ..., N-1$, can be obtained as in Steps 14 and 21 by removing the static vector. The proposed method is simple and robust for realtime processing of received data in different environments with a low delay.

Fig. 4.3 and Fig. 4.4 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

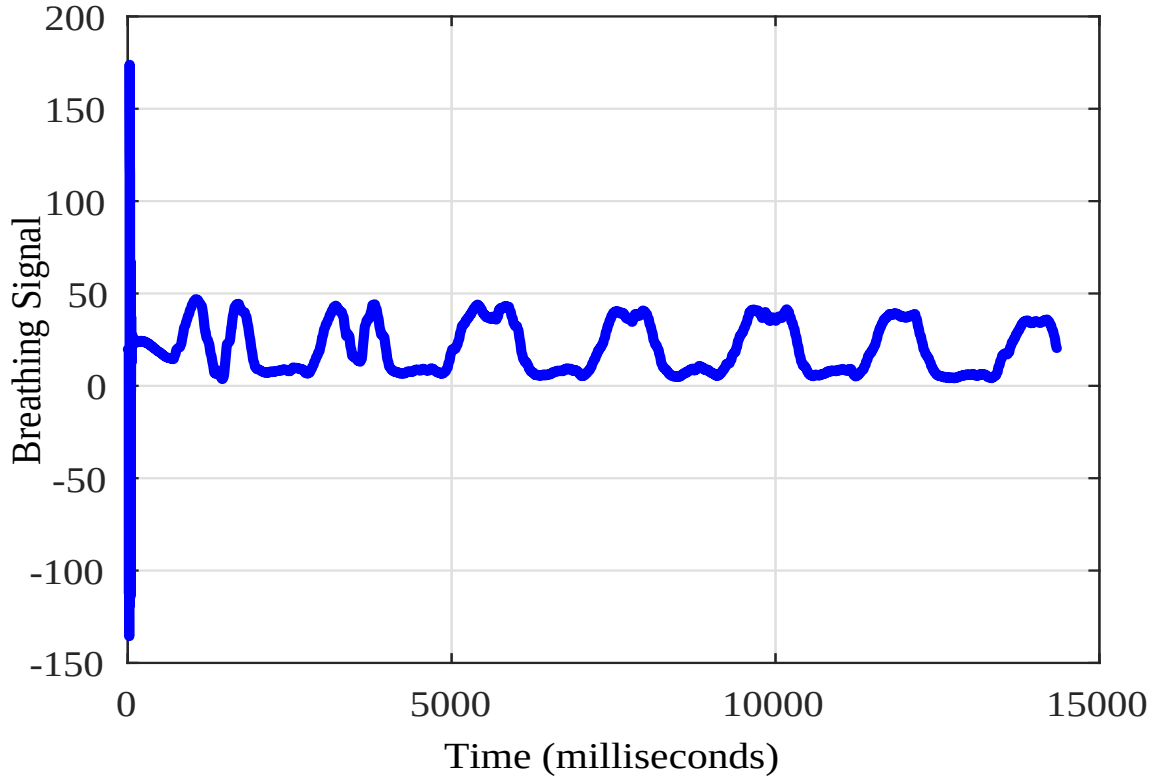

Figure 4.5: Respiration curve for phase data without removing static vector effect.

represent well the average amplitude 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 signal they carry.

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 (3.2), that is

$$\varphi(t) = \arctan\left(\frac{O_Q(t)}{O_I(t)}\right). \tag{4.1}$$

With (3.2) and (4.1), 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

Algorithm 1: The Adaptive Median Filter Method

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

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.

Fig. 4.6 presents the respiration curve obtained from the phase data with removing static vectors. 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.

4.4.1 Data Calibration

-

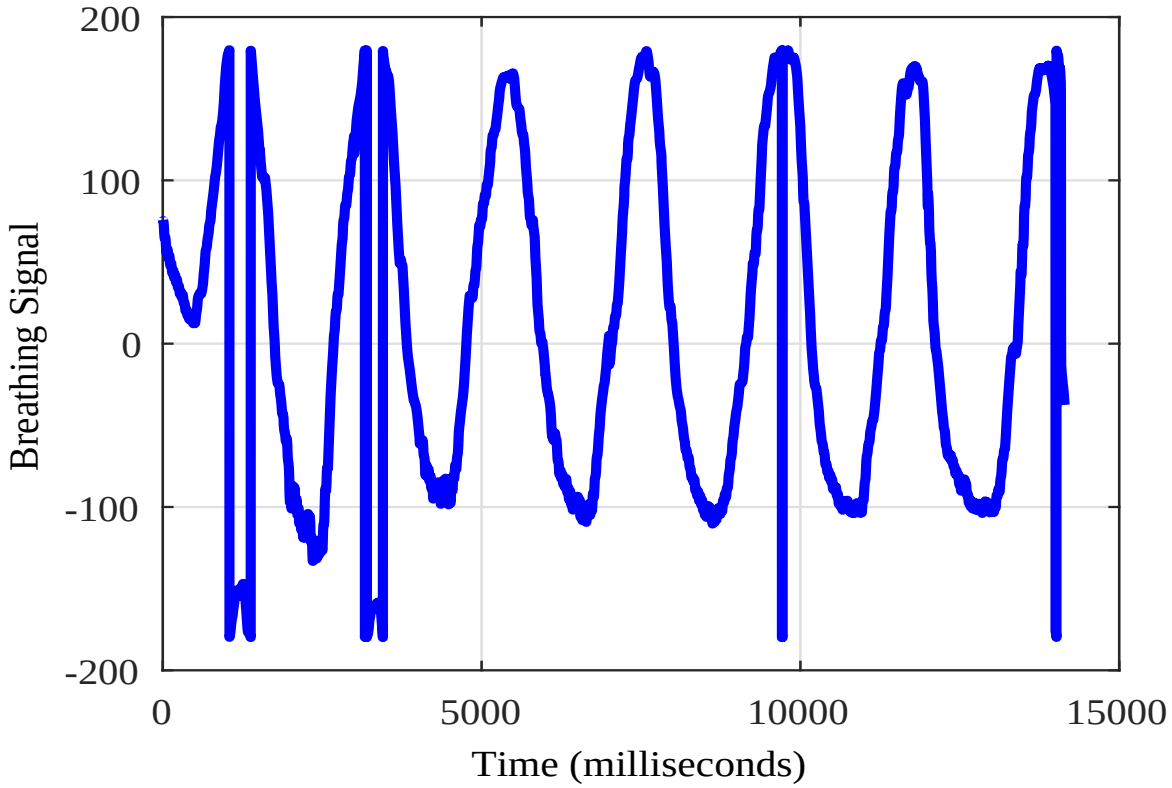

Figure 4.6: Respiration curve for phase data with removing static vector effect.

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.

Fig. 4.7 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. Fig. 4.8 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 signal, because the phase data in a window can continually

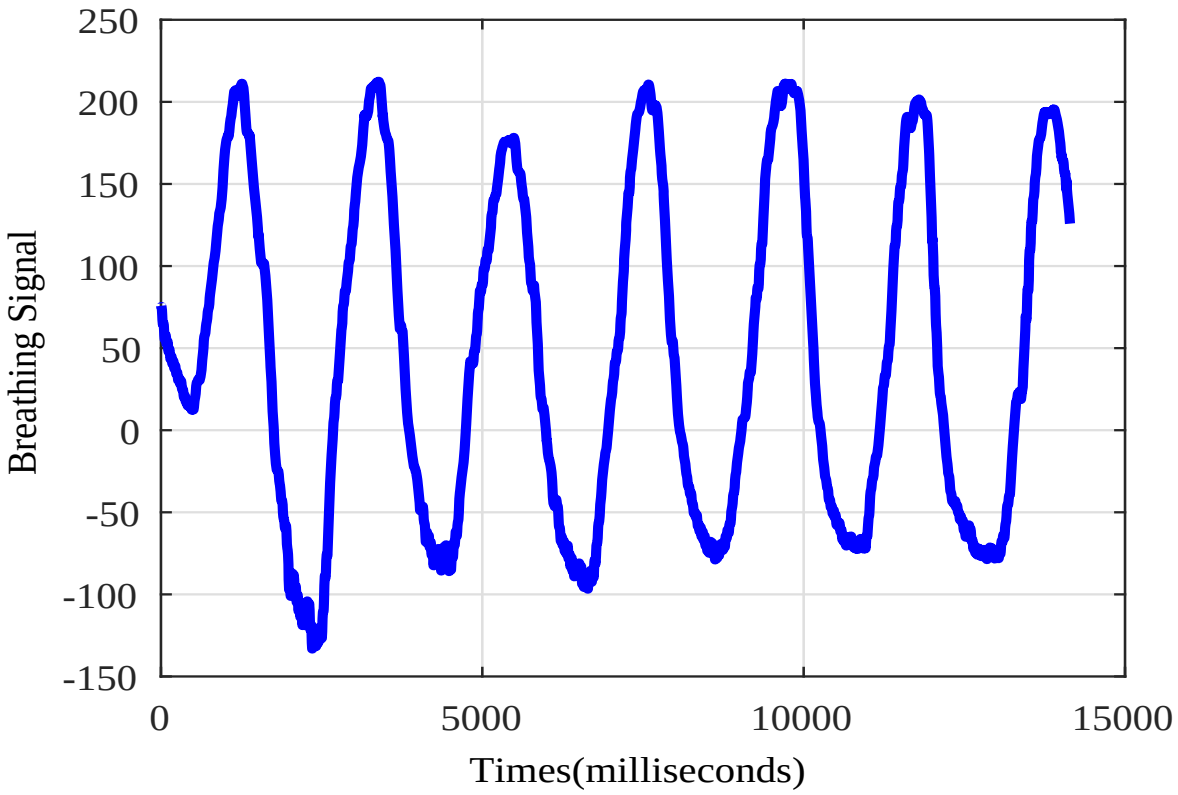

Figure 4.7: Respiration curve for phase data after unwrapping the phase data.

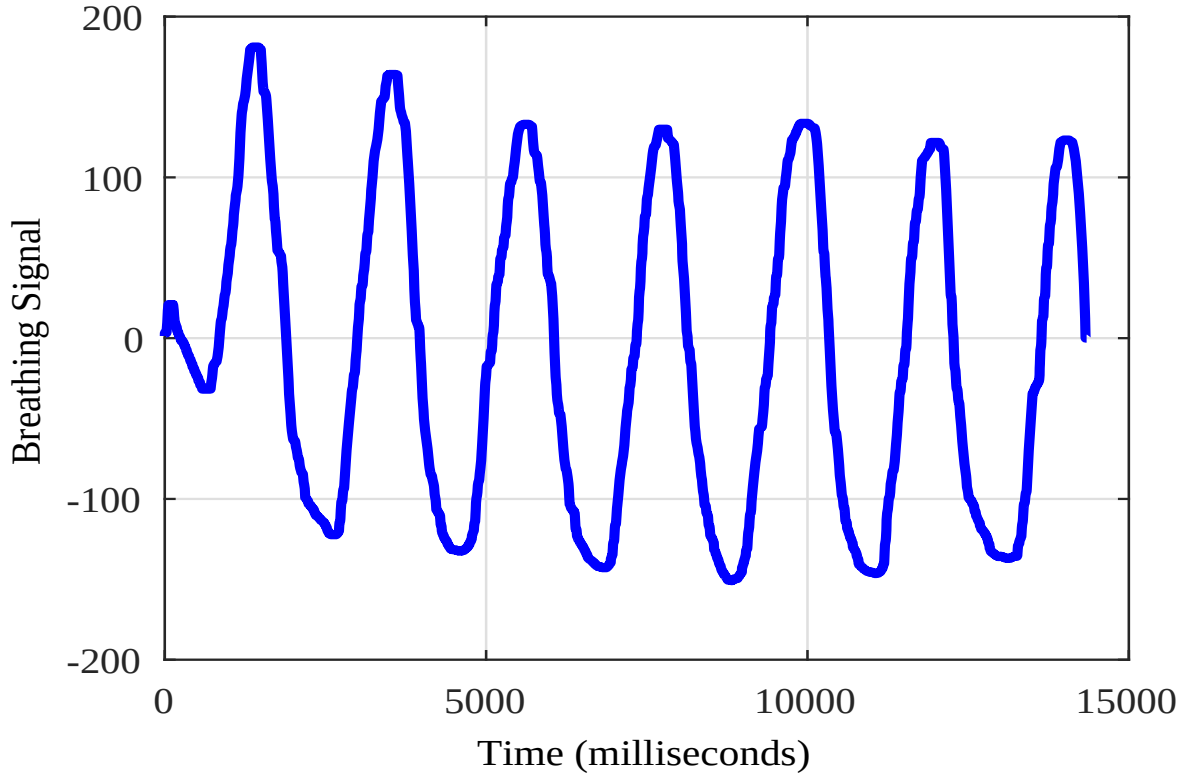

Figure 4.8: Respiration curve for unwrapped phase data after the median filter method.

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

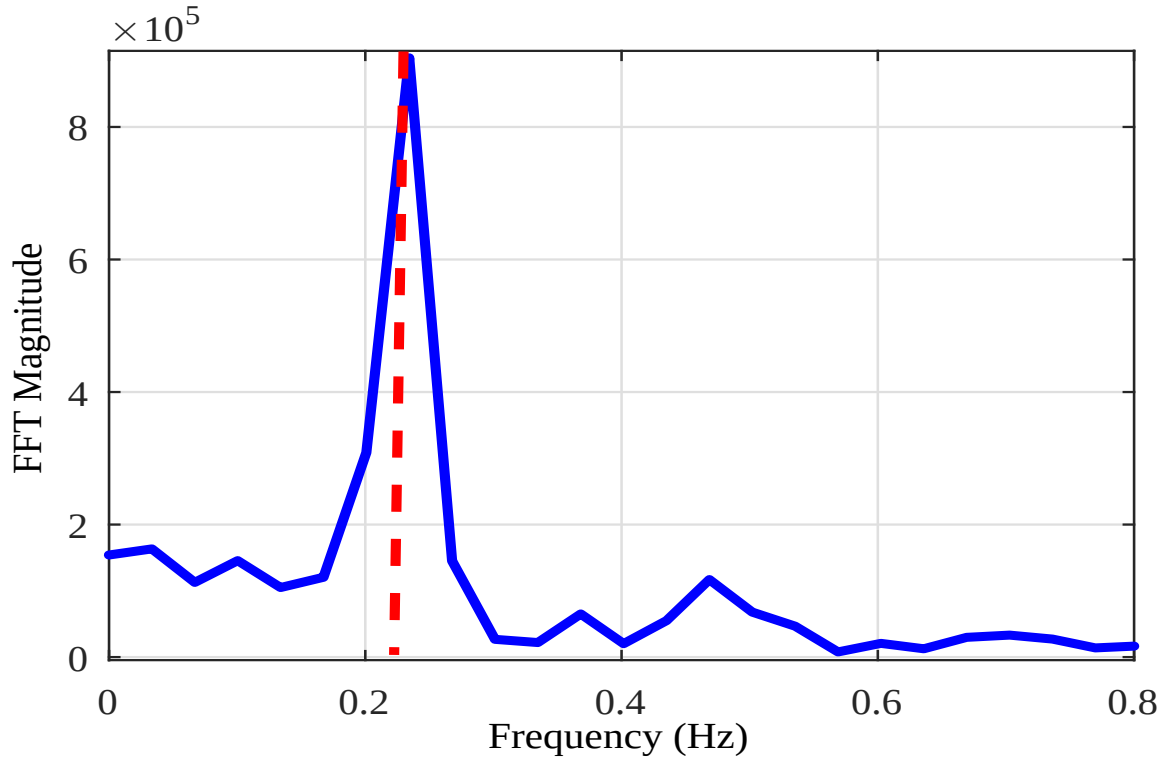

Figure 4.9: Respiration rate estimation based on FFT.

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.

Fig. 4.9 illustrates the FFT based respiration rate estimation. We can see that the estimated 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.

Chapter 5

Experimental Study

5.1 Experiment Configuration

We prototype the SonarBeat system on the Android platform in Java and the Android SDK, as an 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 1920 points with a sampling rate of 48 KHz. Therefore, we set the realtime signal processing unit to 1920 points, which is about 40 ms.

We conduct extensive experiments with SonarBeat with five persons over three months. The test scenarios include an office, a bedroom and a movie theater. The *office* is a 4.5×8.8 m^2 room. The room is crowed with tables and PCs, which form a complex propagation environment. In this office environment, we test SonarBeat under different parameters 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 peoples 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 that is tested in one hour for watching a movie.

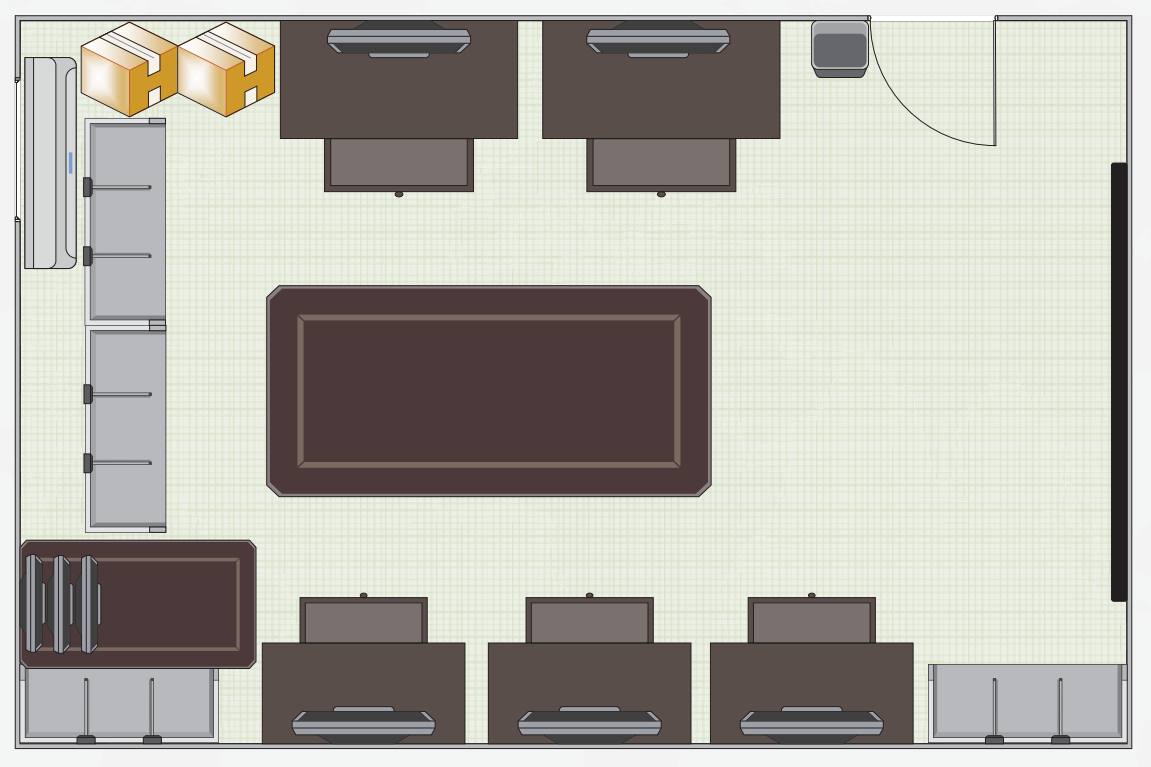

Figure 5.1: Experimental setup in the *office* scenario.

Moreover, the proposed system mainly focuses on estimating the breathing rate for a person, where we consider other persons have above 60 cm distance away from the smartphone. For comparison purpose, we use the NEULOG Respiration Monitor Belt Logger Sensor to record the ground truth of the breathing rate (see Fig. 5.4).

For breathing rate estimation each time, we use all the collected data for breathing rate estimation with FFT for 30 seconds. Moreover, we use cumulative distribution function (CDF) of breathing errors as the measurement metric, which can be employed to evaluate the total performance of the proposed system. Moreover, 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 parameter.

5.2 Performance of Breathing Rate Estimation

Fig. 5.5 presents the CDFs of estimation errors in breathing rate estimation with SonarBeat. For comparison purpose, we also develop an LEVD based system [46], where the LEVD method is used for estimating the static vector and all other signal processing methods are the same as

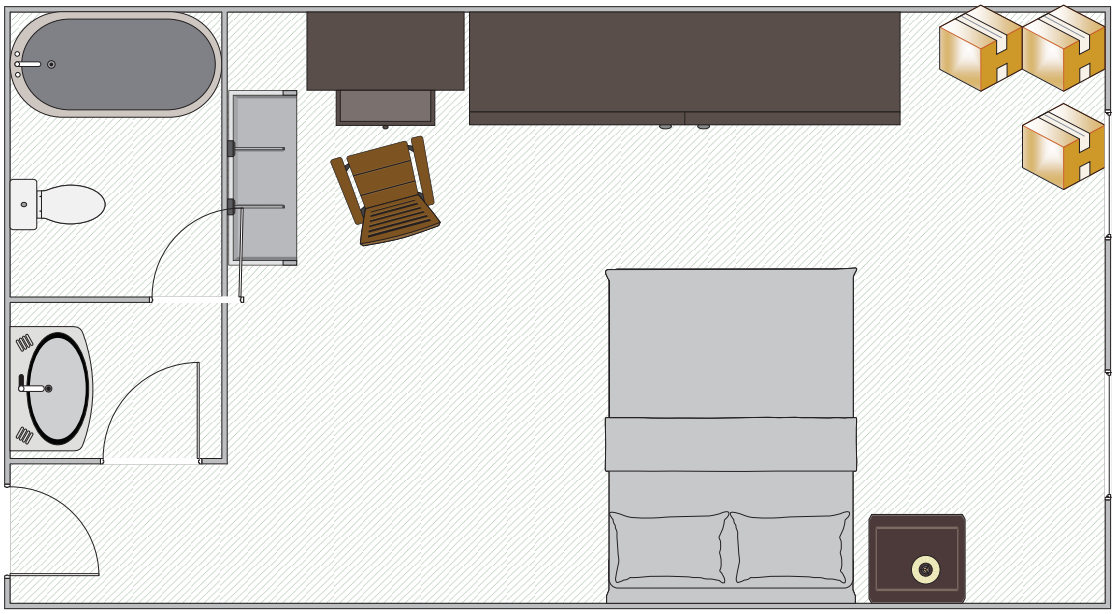

Figure 5.2: Experimental setup in the *bedroom* scenario.

in SonarBeat. We find that SonarBeat and the LEVD based method achieve a median error of 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 have an estimated error under 0.5 bpm, while only 60% of the test results with the LEVD based method have an estimated error 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 robust in varying environments where the same threshold will not work. However, SonarBeat leverages the adaptive median filter method, and is thus more robust to changes in the environment. Thus SonarBeat can achieve a higher and more stable breathing rate estimation accuracy than LEVD.

Fig. 5.6 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.

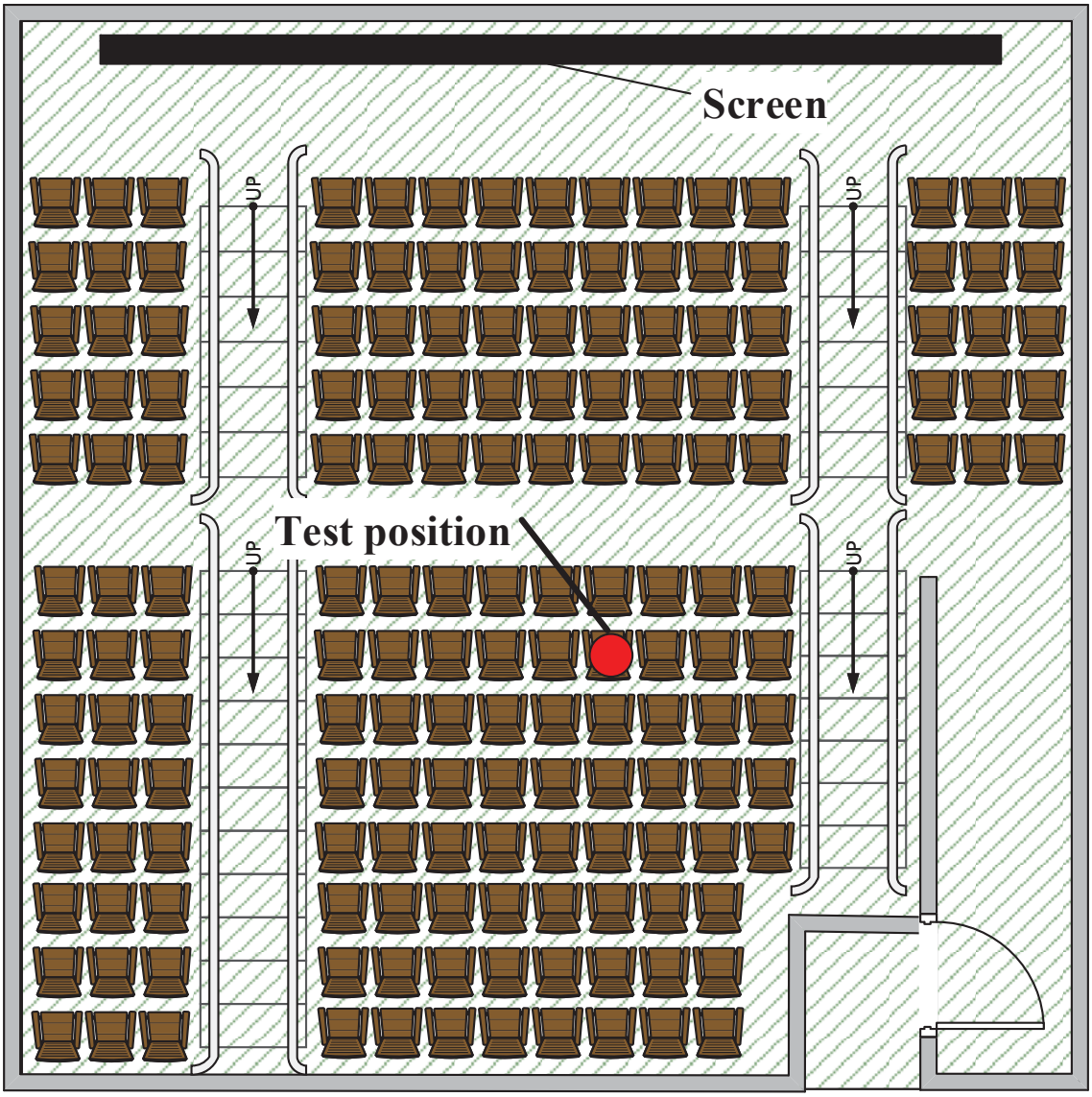

Figure 5.3: Experimental setup in the *movie theater* scenario.

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 conditioner, 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.

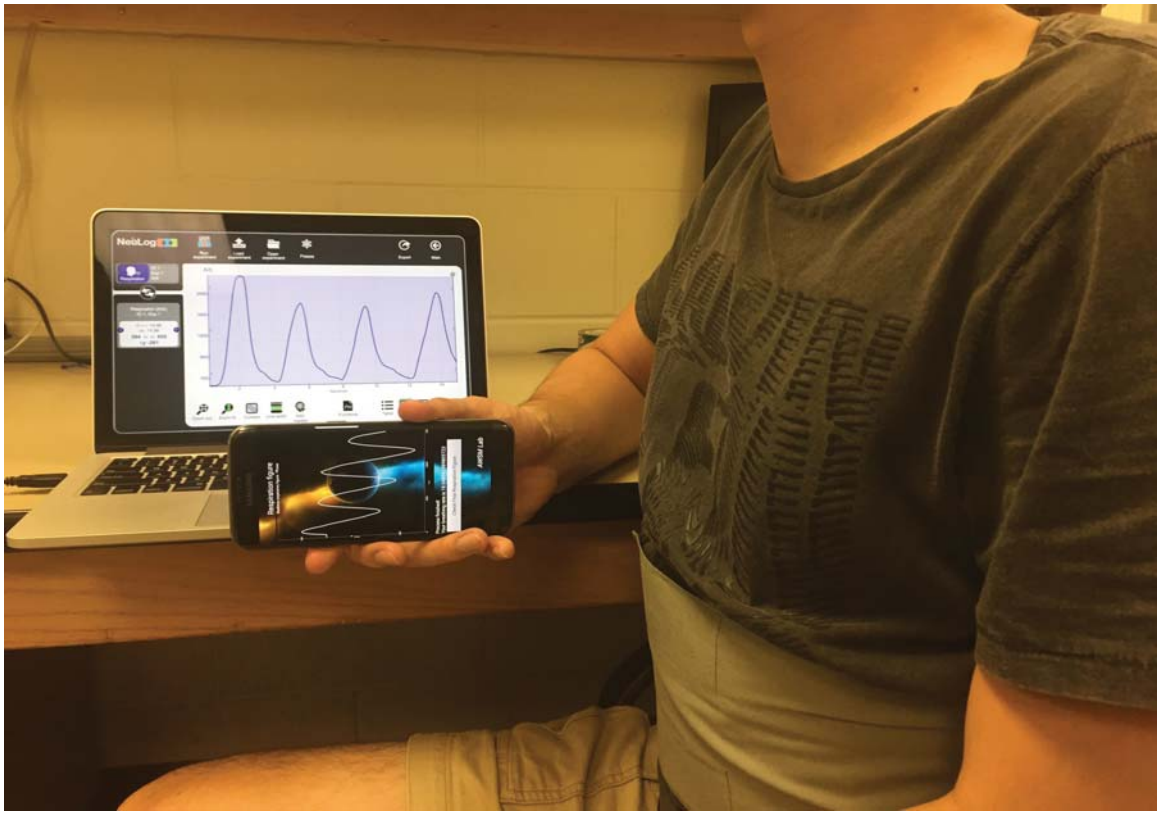

Figure 5.4: The office experiment, where the NEULOG Respiration Monitor Belt Logger Sensor records the ground truth (shown on the laptop screen).

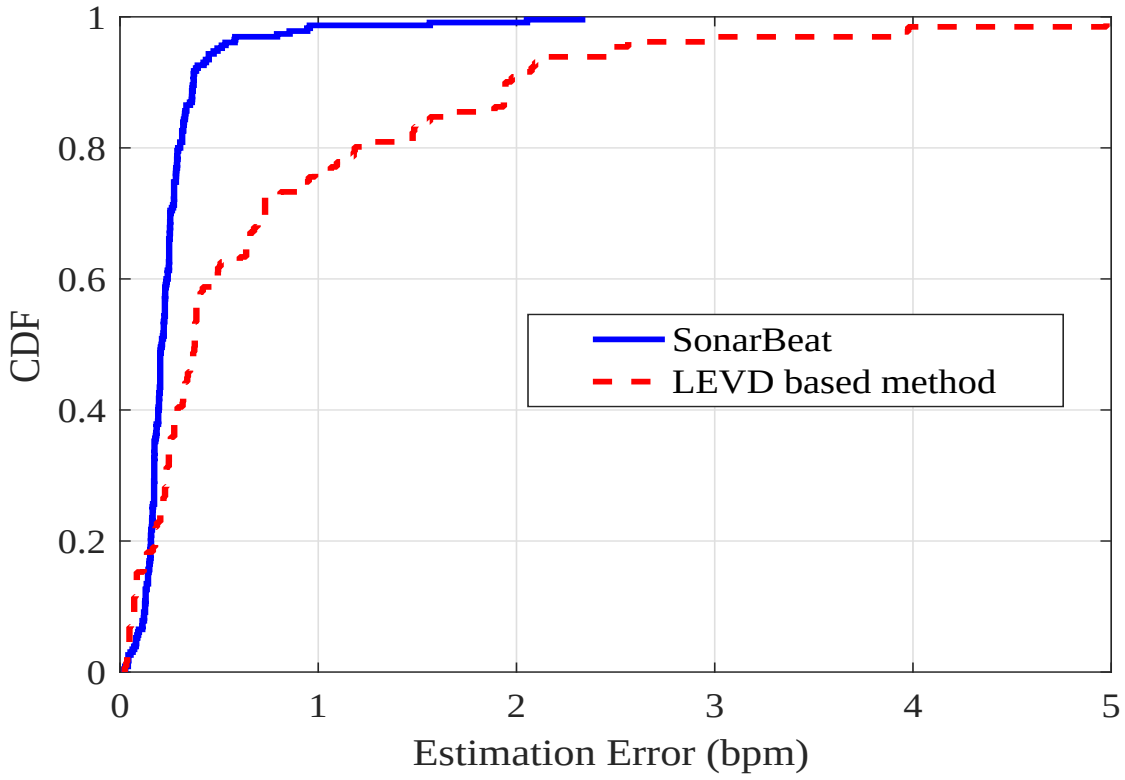

Figure 5.5: CDFs of estimation errors in breathing rate estimation.

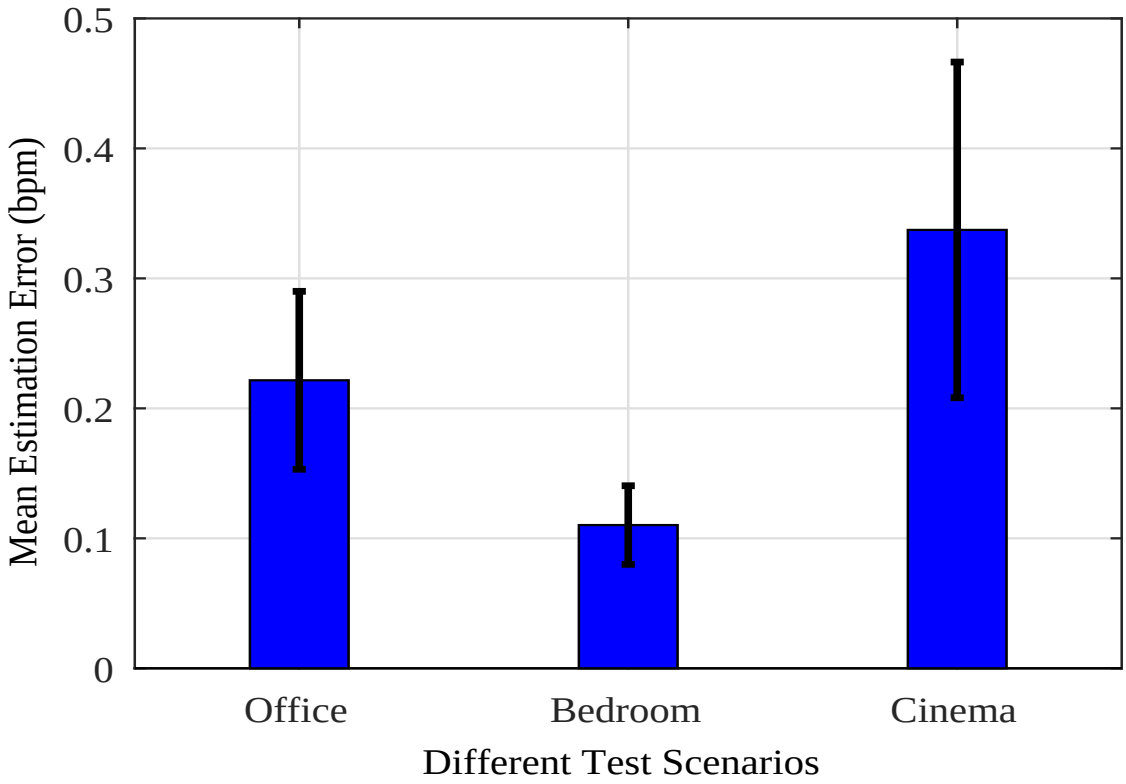

Figure 5.6: Mean estimation error for three different scenarios: office, bedroom, and movie theater.

5.3 Impact of Various Environmental Factors

Fig. 5.7 shows the impact of different persons in the office scenario. In the experiment, we test five persons including three men and two women. Every volunteer wears the NEULOG Respiration unit to record the ground truth for breathing data. From Fig. 5.7, 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. 5.8 shows the impact of different breathing rates in the office scenario, where the test subject controls his/her breathing at a slow, normal, and fast breathing rates, which are in the rangs 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 higher breathing rate, the stability of the breathing signal becomes weaker. In other words, with fast breathing, the chest movements are more irregular, thus leading to

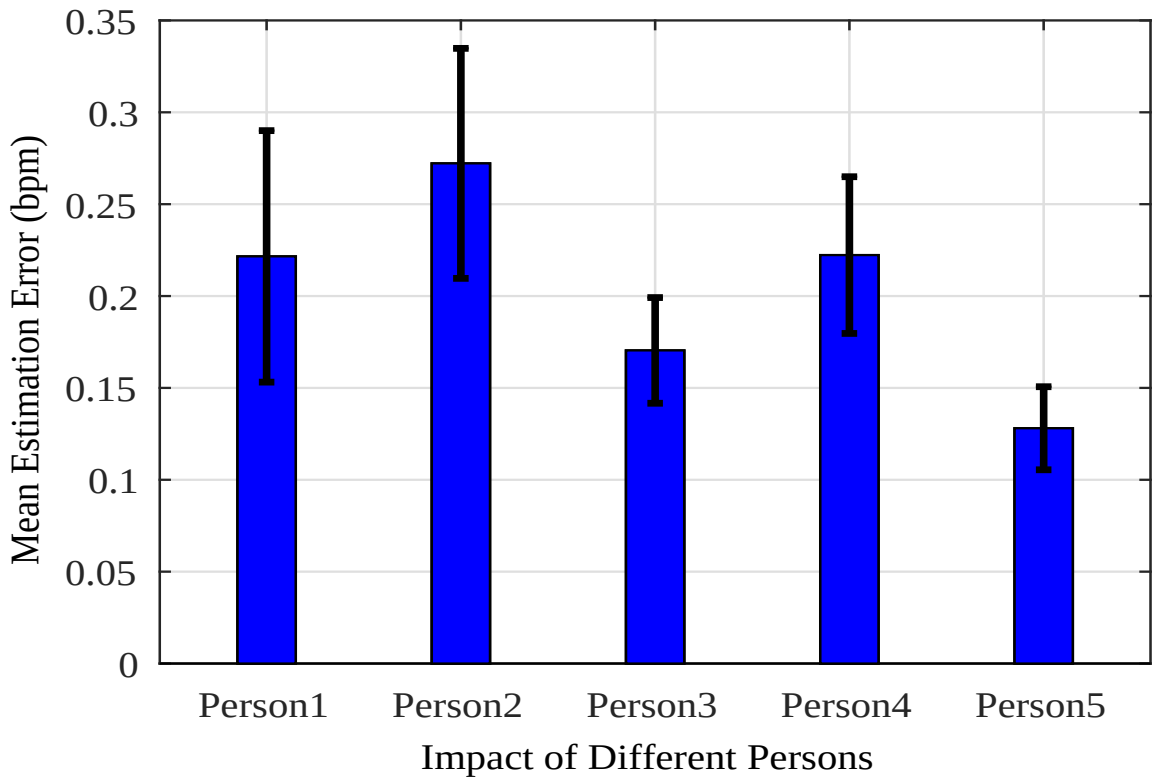

Figure 5.7: Breathing rate results for five different persons.

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 produced good frequency estimation. Nevertheless, SonarBeat can still effectively capture the different breathing rates with a mean error a little over 0.4 bpm in the fast breathing case.

Fig. 5.9 shows 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 a lower power from the chest reflection if the distance is increased. Moreover, breathing rate estimation with SonarBeat depends on I/Q demodulation. The magnitude of the I/Q components becomes weaker when the distance is increased, leading to higher errors. To improve the measurement distance, 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,

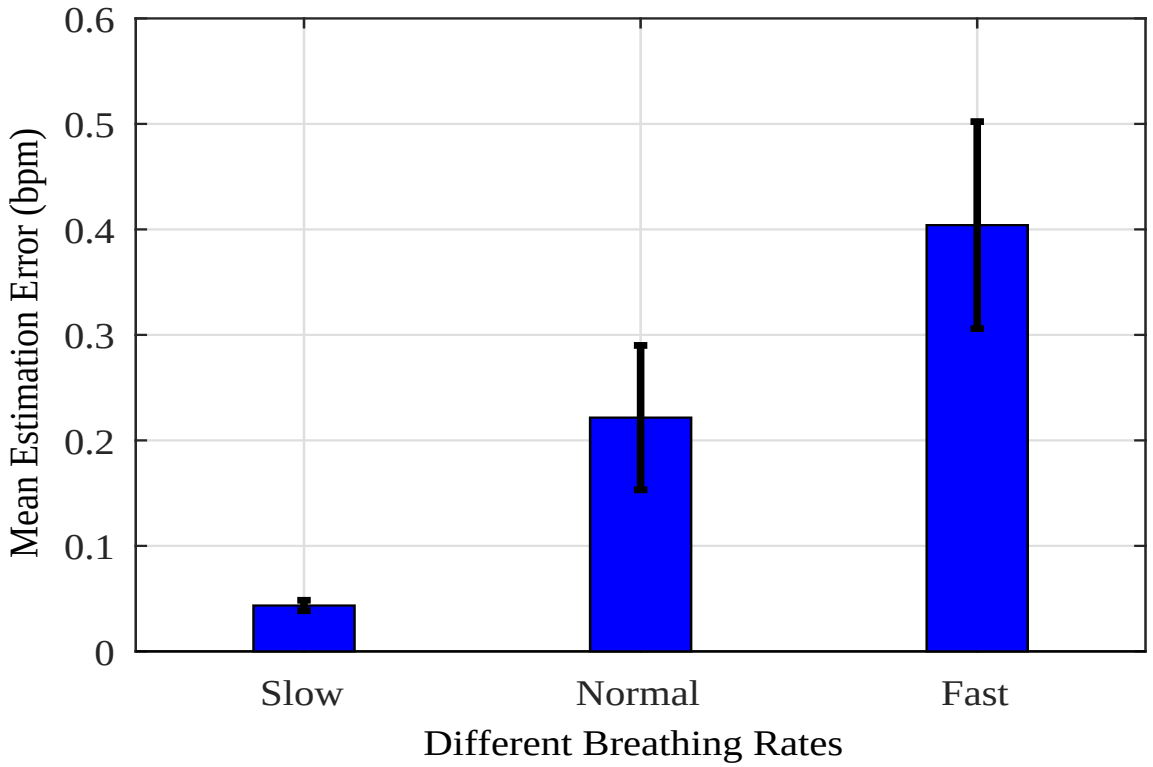

Figure 5.8: Impact of different breathing rates.

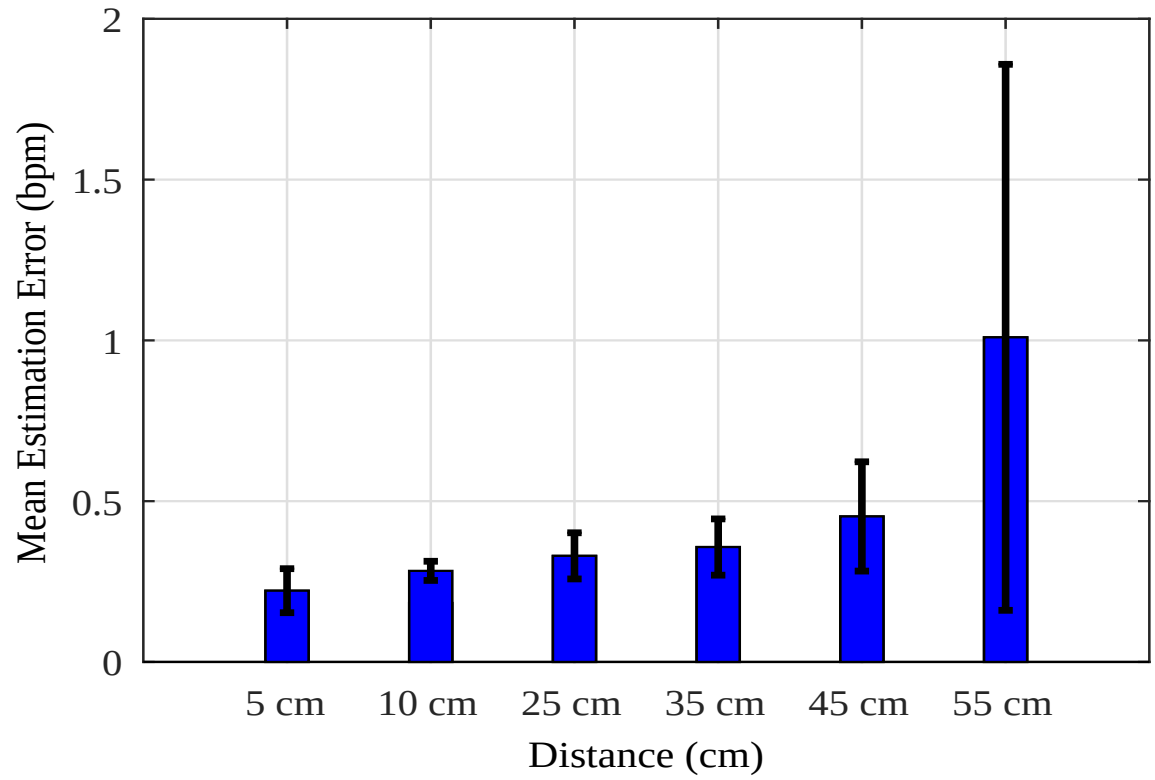

Figure 5.9: Impact of the distance between the test subject and the smartphone.

we set the cutoff frequency to 40 Hz for the sampling rate of 48 KHz. We can see that under 50 cm, the proposed system can achieve very good accuracy.

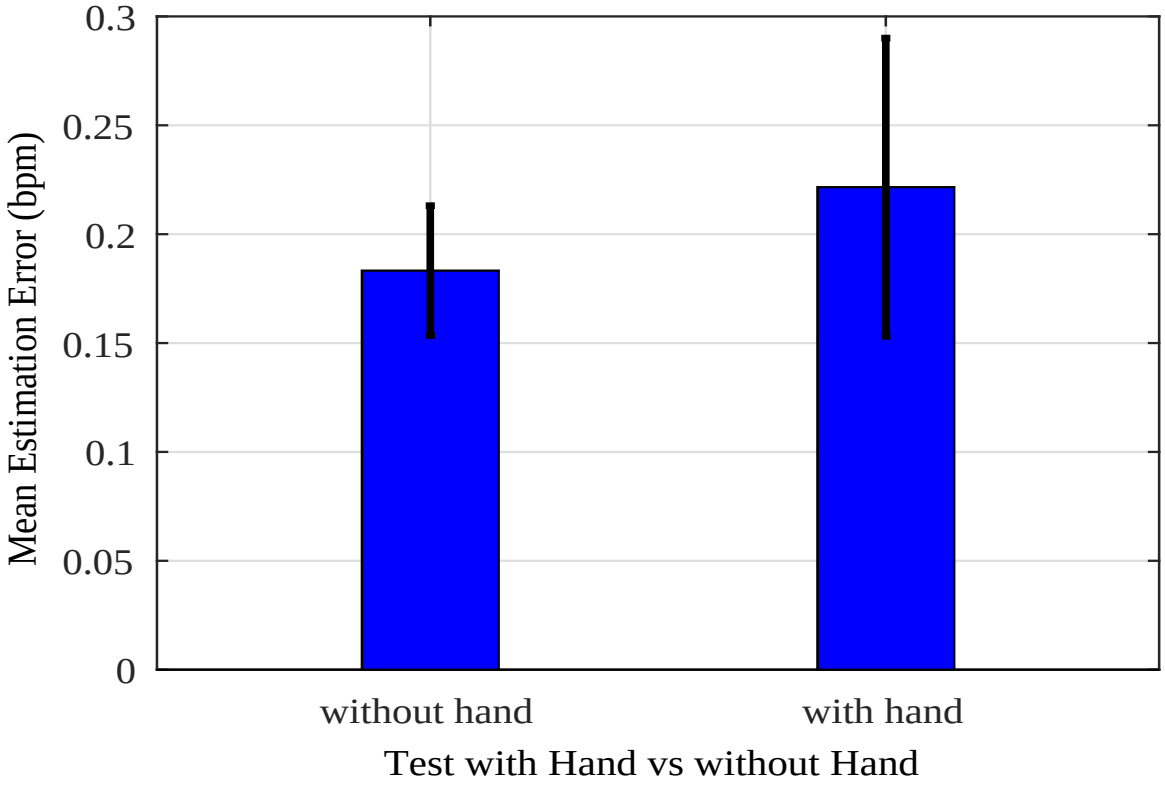

Figure 5.10: Breathing rate results when the smartphone is held in hand or put on a desk.

Fig. 5.10 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 are 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 hand held 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. 5.11 shows the impact of cloth thickness. In the experiment, the test subject wears clothes of different types and thickness. The distance between the user and the smartphone is kept between 10 cm to 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 gets thicker, which leads to 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.

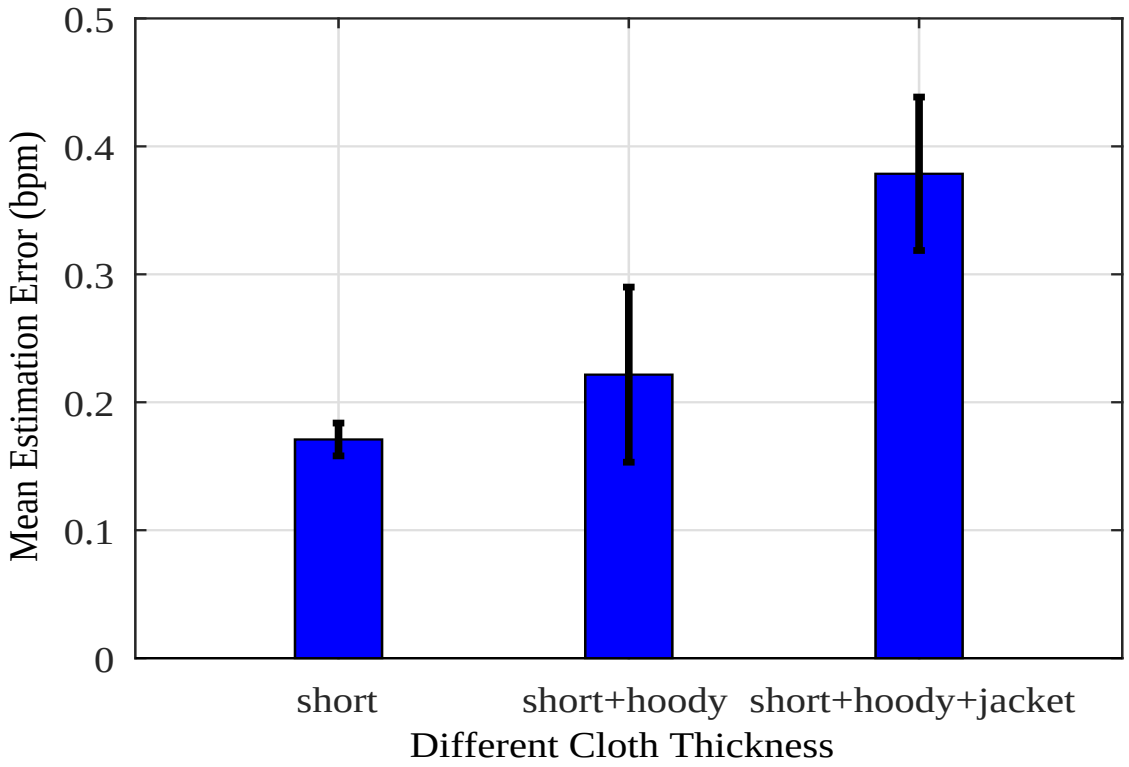

Figure 5.11: Impact of cloth thickness.

Fig. 5.12 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.

Fig. 5.13 shows the impact of different poses including sitting, standing, and sleeping in the bedroom scenario. For sitting case, the smartphone is put on a desk and the distance between the smartphone and the person is under 15 cm. For standing case, the person holds the smartphone with the same distance. For sleeping case, the person wearing night-clothes is laying in the bed, who faces the smartphone on the desk with the same distance. We can see that for poses with the sitting and standing, the mean estimation errors are smaller than that with the sleeping. This is because the strength of the received signal in the sleeping scenario is smaller than the other two cases.

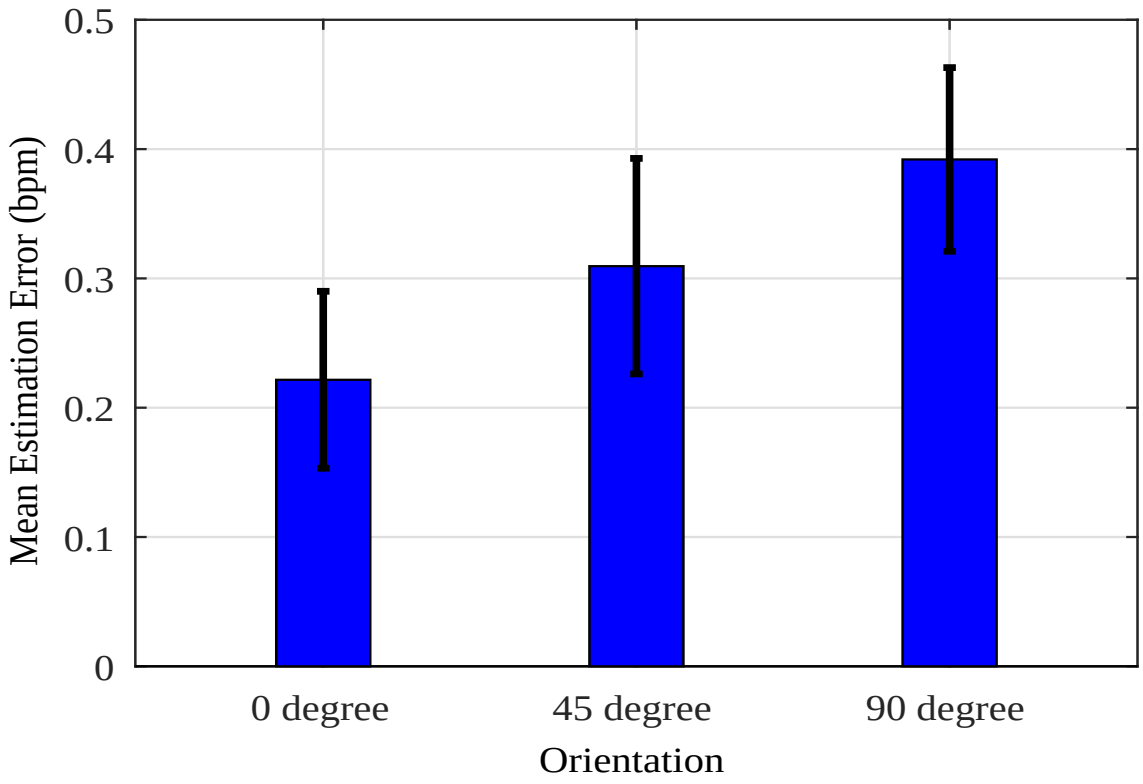

Figure 5.12: Impact of user orientation relative to the smartphone.

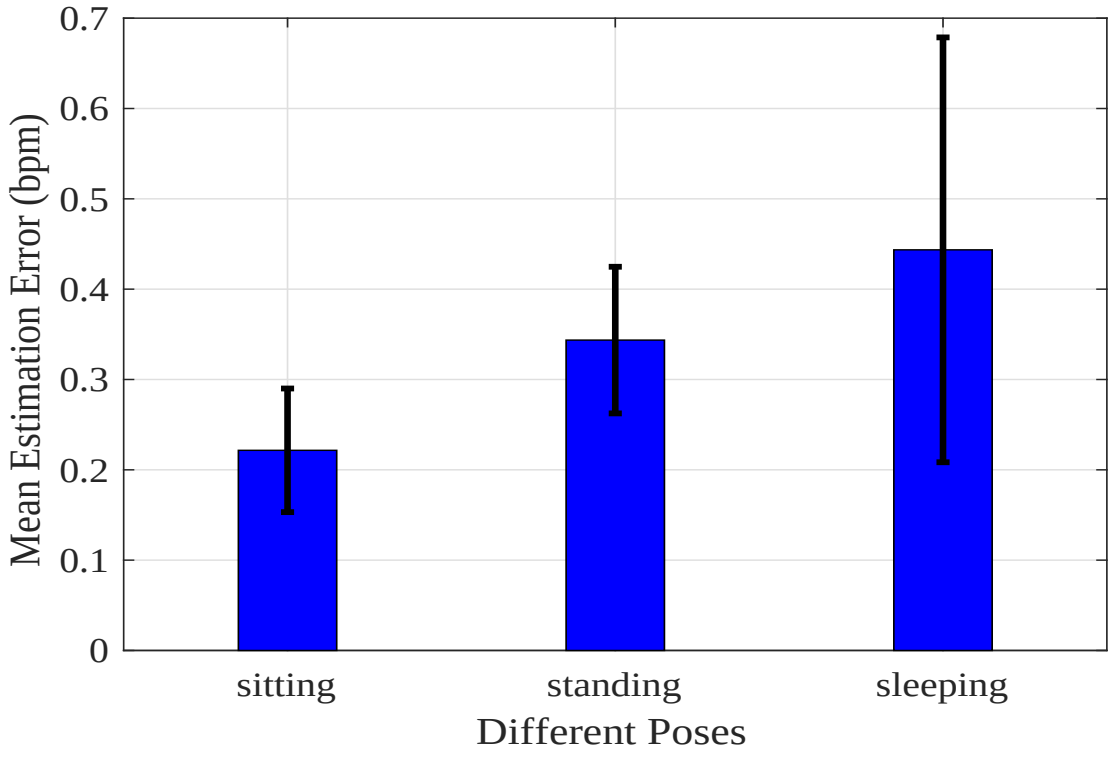

Figure 5.13: Impact of different poses.

5.4 Impact of Various System Parameters

We evaluate the impact of various system parameters in this sections. Fig. 5.14 presents the estimation errors with a Samsung Galaxy S6 with Android 6.0 and a Samsung Galaxy S7 Edge

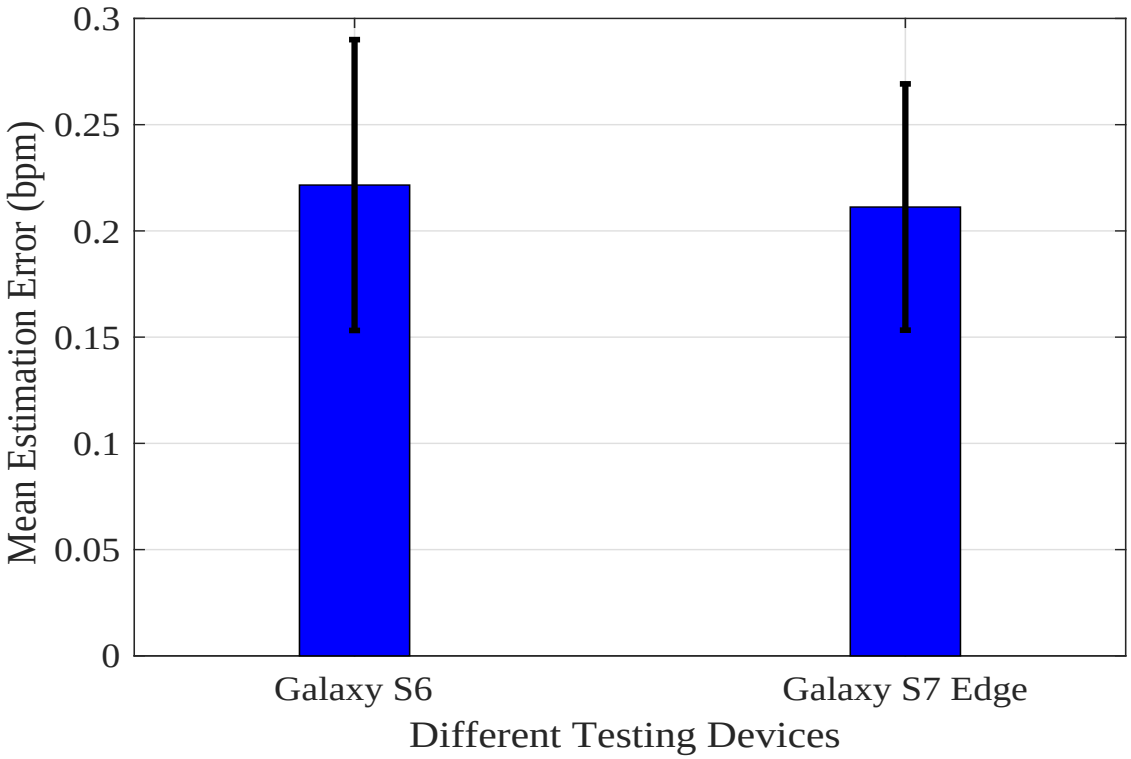

Figure 5.14: Estimation error results with two different smartphone platforms.

working with the lasted Version Android 7.0. The speaker and microphone are on the bottom of both smartphones. We can see that Samsung Galaxy S7 Edge has a similar performance as Samsung Galaxy S6, with mean error of 0.21 bpm and 0.22 bpm, respectively. We find that Samsung Galaxy S7 Edge has stronger processing power, and thus it can obtain better realtime performance than Samsung Galaxy S6.

Fig. 5.15 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 smaller SNR for the received breathing signal. Thus, the higher the frequency, the larger the mean breathing rate estimation error.

Fig. 5.16 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

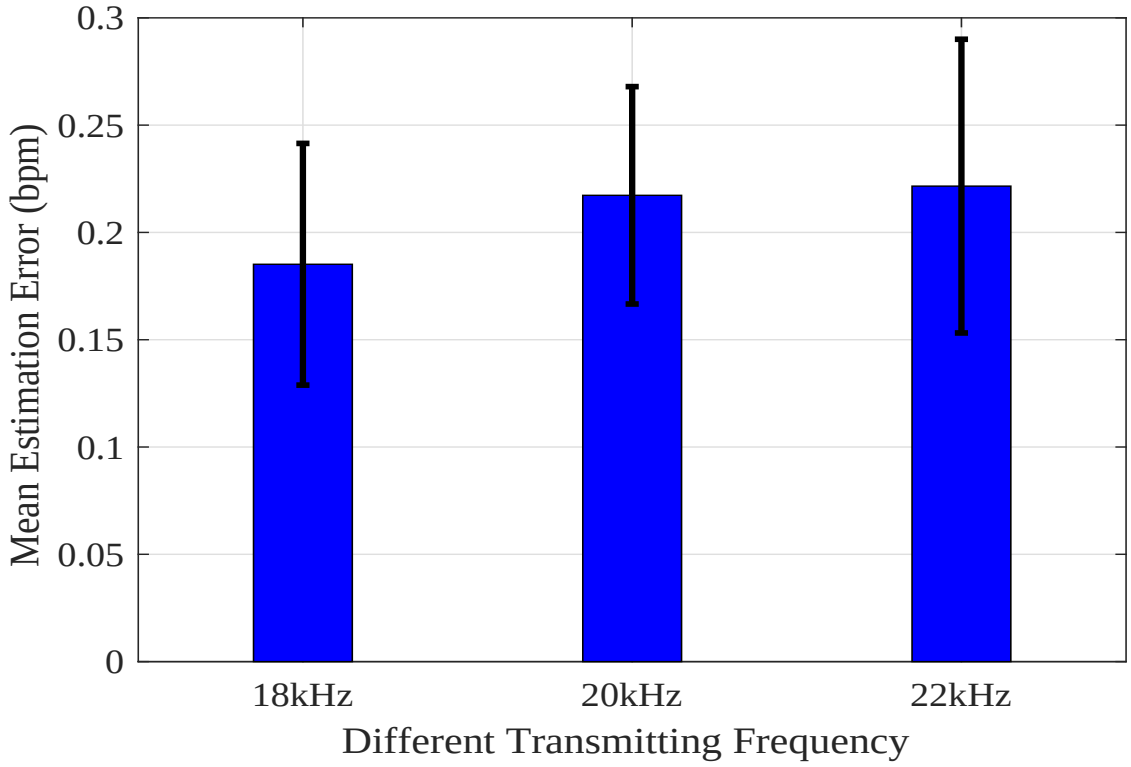

Figure 5.15: Impact of the frequency of the inaudible acoustic signal.

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 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 realtime breathing monitoring with the smartphones, but also achieve a high accuracy.

Fig. 5.17 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 Fig. 5.17, we can see that, when the window size is increased from 7400 to 7700, the mean breathing rate estimation error is only increased slightly from 0.21 bpm to 0.26 bpm. Moreover, the larger window sizes from 7400 to 7700 are almost half of the signal points, which is effective for removing the static vector. Thus, we choose a window size of 7500 for SonarBeat, which can achieve the best breathing estimation accuracy.

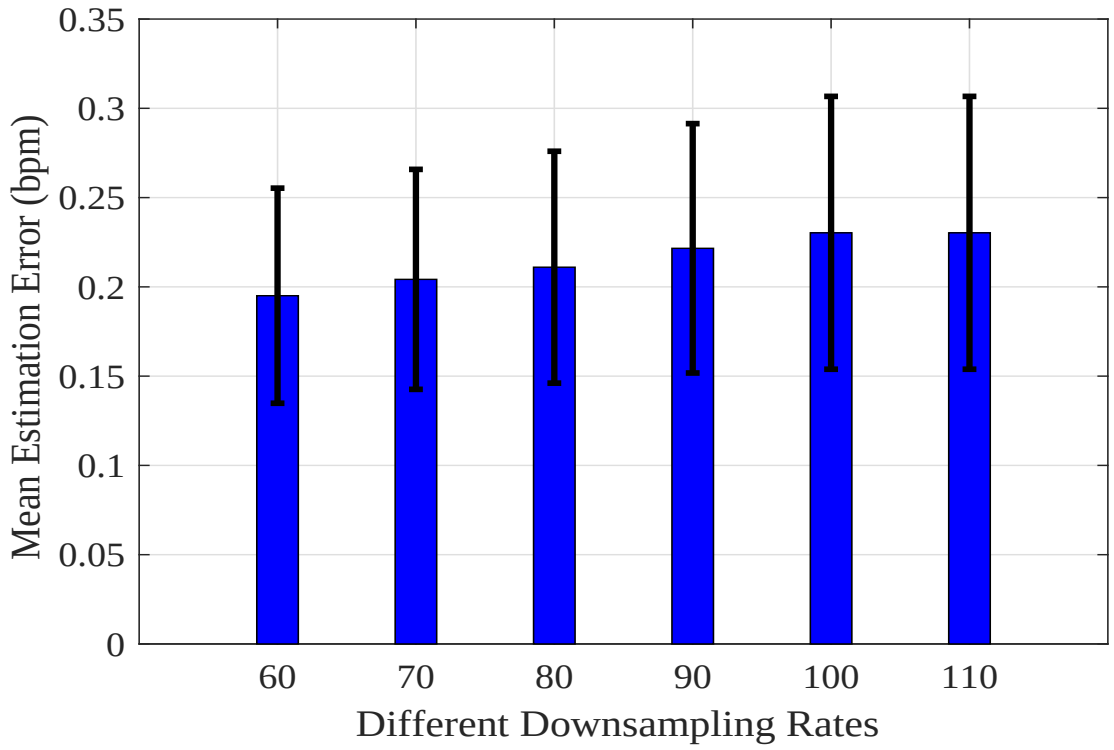

Figure 5.16: Impact of different downsampling rates.

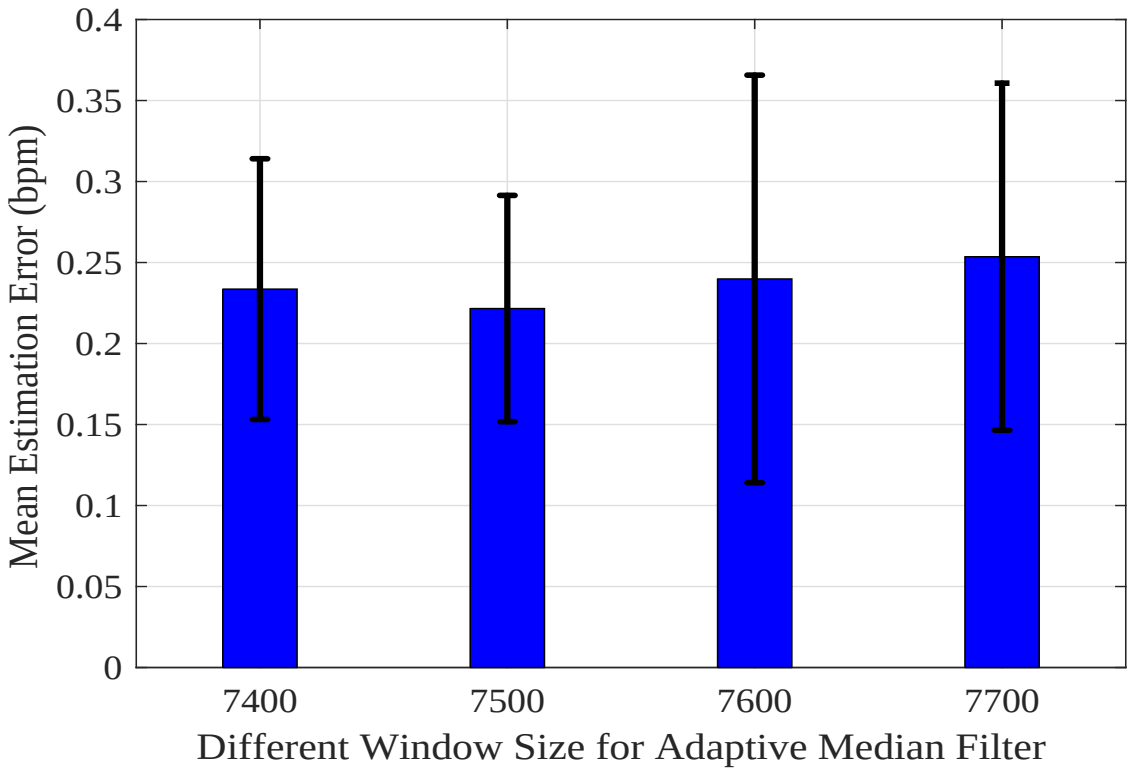

Figure 5.17: Impact of different window sizes of the adaptive median filter.

Fig. 5.18 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

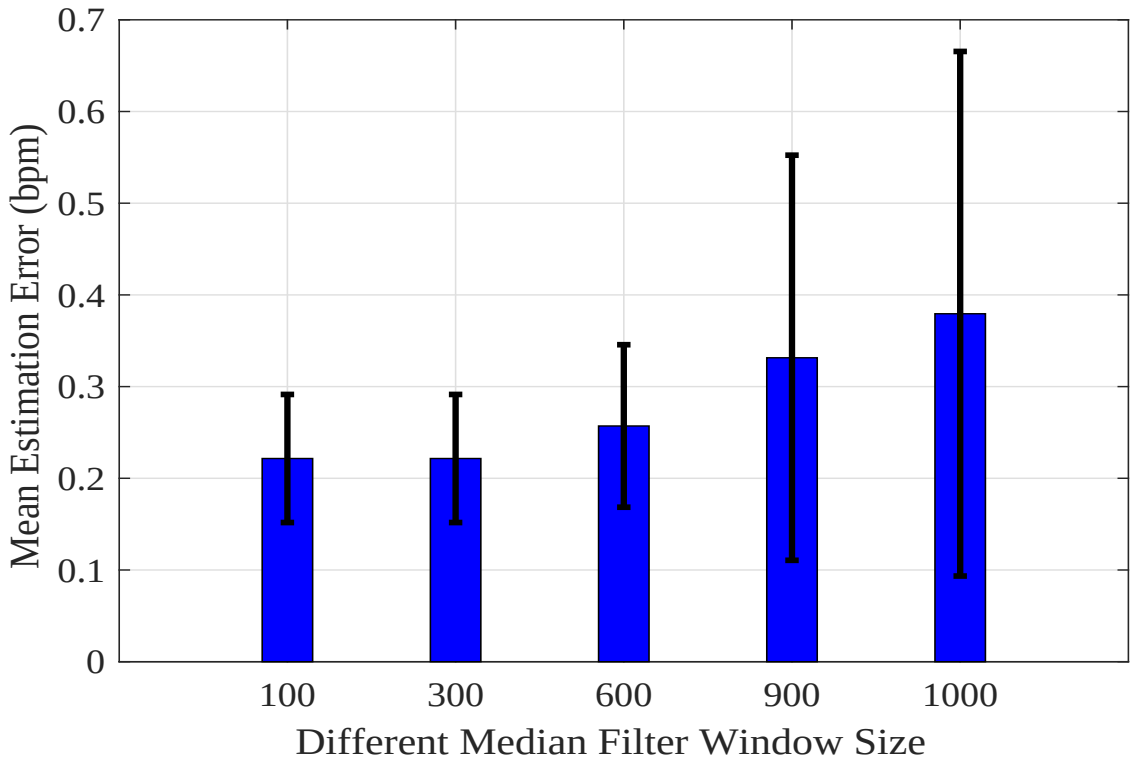

Figure 5.18: Impact of different sizes of the median filter window.

window size is set to 100 or 300, similar mean error of 0.21 bpm are achieved. On the other hand, when the median filter window size is increased from 600 to 1000, 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, thus 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 realtime breathing monitoring.

Chapter 6

Conclusions and Future Work

6.1 Conclusion

In this paper, 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 on-line 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 breathing rate estimation for different factors and parameters.

6.2 Future work

In the future, we will focus on three research directions as the following .

6.2.1 Apnea detection

With development of artificial intelligent, such as deep learning, we will provide smartphones such ability as human brain to distinguish which is normal respiration and what is respiration disease. Long-Short Term Memory (LSTM) neural network and Convolutional Neural Networks (CNN) are two candidates for mobile and wireless applications [65, 66, 67, 68, 69, 70,

71, 72, 33, 73]. Basically, we will collect respiration data with smartphone and training it with deep learning methods, and leverage the trained network in smartphone for apnea detection when a person is sleeping.

6.2.2 Indoor localization

Deep learning related research and development are pervasive research hot point. Also, there is still lacking of a good solution for indoor localization. Considering those prerequisite, we consider to use smartphone sensors for collecting WiFi signals and strengths of magnetic field for different points as training data, and feed it to LSTM network for training. This fusion method can improve localization accuracy for online test phase.

6.2.3 Gesture tracking

For sensing system, we will consider computer vision technology combined with audio signal for reaching 3-D gesture tracking. With the experiments of SonarBeat, acoustic signal has the ability to measure the variation of distance. Combine it with computer vision using smartphone camera, we can get the position of an object in a 3-D space.

References

[1] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[2] X. Wang, X. Wang, and S. Mao, "RF sensing for Internet of Things: A general deep learning framework," *IEEE Communications*, Aug. 2018.

[3] X. Wang, S. Mao, and M. X. Gong, "An overview of 3gpp cellular vehicle-to-everything standards," *GetMobile: Mobile Computing and Communications*, vol. 21, no. 3, pp. 19–25, 2017.

[4] S. Cicalo, M. Mazzotti, S. Moretti, V. Tralli, and M. Chiani, "Multiple video delivery in m-health emergency applications," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 1988–2001, 2016.

[5] S.-E. Moon and J.-S. Lee, "Implicit analysis of perceptual multimedia experience based on physiological response: A review," *IEEE Transactions on Multimedia*, vol. 19, no. 2, pp. 340–353, 2017.

[6] H. Müeller and D. Unay, "Retrieval from and understanding of large-scale multi-modal medical datasets: A review," *IEEE Transactions on Multimedia*, 2017.

[7] I. M. Baytas, K. Lin, F. Wang, A. K. Jain, and J. Zhou, "Phenotree: Interactive visual analytics for hierarchical phenotyping from large-scale electronic health records," *IEEE Transactions on Multimedia*, vol. 18, no. 11, pp. 2257–2270, 2016.

[8] P. Henriquez, B. J. Matuszewski, Y. Andreu-Cabedo, L. Bastiani, S. Colantonio, G. Coppini, M. D'Acunto, R. Favilla, D. Germanese, D. Giorgi *et al.*, "Mirror mirror on the

wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization," *IEEE Transactions on Multimedia*, 2017.

[9] X. Yuan, X. Wang, C. Wang, J. Weng, and K. Ren, "Enabling secure and fast indexing for privacy-assured healthcare monitoring via compressive sensing," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2002–2014, 2016.

[10] C. Hunt and F. Hauck, "Sudden infant death syndrome," *Can. Med. Assoc. J*, vol. 174, no. 13, pp. 1309–1310, Apr. 2006.

[11] N. H. Shariati and E. Zahedi, "Comparison of selected parametric models for analysis of the photoplethysmographic signal," in *Proc. 1st IEEE Conf. Comput., Commun. Signal Process.*, Kuala Lumpur, Malaysia, Nov. 2005, pp. 169–172.

[12] M. L. R. Mogue and B. Rantala, "Capnometers," *Journal of clinical monitoring*, vol. 4, no. 2, pp. 115–121, Apr. 1988.

[13] A. Droitcour, O. Boric-Lubecke, and G. Kovacs, "Signal-to-noise ratio in Doppler radar system for heart and respiratory rate measurements," *IEEE Trans. Microw. Theory Technol.*, vol. 57, no. 10, pp. 2498–2507, Oct. 2009.

[14] P. Nguyen, X. Zhang, A. Halbower, and T. Vu, "Continuous and fine-grained breathing volume monitoring from afar using wireless signals," in *Proc. IEEE INFOCOM'16*, San Francisco, CA, Apr. 2016, pp. 1–9.

[15] J. Salmi and A. F. Molisch, "Propagation parameter estimation, modeling and measurements for ultrawideband mimo radar," *IEEE Trans. Microw. Theory Technol.*, vol. 59, no. 11, pp. 4257–4267, Nov. 2011.

[16] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. ACM CHI'15*, Seoul, Korea, April 2015, pp. 837–846.

[17] D. Hu, S. Mao, and J. H. Reed, "On video multicast in cognitive radio networks," in *INFOCOM 2009, IEEE*. IEEE, 2009, pp. 2222–2230.

[18] M. X. Gong, B. Hart, and S. Mao, "Advanced wireless lan technologies: Ieee 802.11 ac and beyond," *GetMobile: mobile computing and communications*, vol. 18, no. 4, pp. 48–52, 2015.

[19] Y. Xu and S. Mao, "User association in massive mimo hetnets," *IEEE Systems Journal*, vol. 11, no. 1, pp. 7–19, 2017.

[20] M. Feng, S. Mao, and T. Jiang, "Base station ON-OFF switching in 5G wireless systems: Approaches and challenges," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 46–54, Aug. 2017.

[21] ——, "Joint duplex mode selection, channel allocation, and power control for full-duplex cognitive femtocell networks," *Elsevier Digital Communications and Networks Journal*, vol. 1, no. 1, pp. 30–44, Feb. 2015.

[22] M. Feng, T. Jiang, D. Chen, and S. Mao, "Cooperative small cell networks: High capacity for hotspots with interference mitigation," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 108–116, Dec. 2014.

[23] M. Feng, S. Mao, and T. Jiang, "Boost: Base station on-off switching strategy for energy efficient massive mimo hetnets," in *Proc. IEEE INFOCOM 2016*, San Francisco, CA, Apr. 2016, pp. 1395–1403.

[24] X. Wang, "Deployment of high altitude platforms in heterogeneous wireless sensor network via mrf-map and potential games," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*. IEEE, 2013, pp. 1446–1451.

[25] K. Xiao, S. Mao, and J. Tugnait, "MAQ: A multiple model predictive congestion control scheme for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2614–2626, Apr. 2017.

[26] ——, "Congestion control for infrastructure-based CRNs: A multiple model predictive control approach," in *Proc. IEEE GLOBECOM 2016*, Washington DC, Dec. 2016, pp. 1–6.

[27] ——, “QoE-driven resource allocation for DASH over OFDMA networks,” in *Proc. IEEE GLOBECOM 2016*, Washington, DC, Dec. 2016, pp. 1–6.

[28] Z. He, S. Mao, and S. Kompella, “Quality of experience driven multi-user video streaming in cellular cognitive radio networks with single channel access,” *IEEE Transactions on Multimedia*, vol. 18, no. 7, pp. 1401–1413, July 2016.

[29] Z. He, S. Mao, and T. Rappaport, “On link scheduling under blockage and interference in 60 ghz ad hoc networks,” *IEEE Access Journal*, vol. 3, pp. 1437–1449, Sept. 2015.

[30] Y. Wang and S. Mao, “On distributed power control in full duplex wireless networks,” *Elsevier Digital Communications and Networks Journal*, vol. 3, no. 1, pp. 1–10, Feb. 2017.

[31] N. Tang, S. Mao, Y. Wang, and R. Nelms, “LASSO-based single index model for solar power generation forecasting,” in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017.

[32] Y. Wang, S. Mao, and T. Rappaport, “On directional neighbor discovery in mmwave networks,” in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 1704–1713.

[33] X. Wang, H. Zhou, S. Mao, S. Pandey, P. Agrawal, and D. Bevly, “Mobility improves LMI-based cooperative indoor localization,” in *Proc. IEEE WCNC 2015*, New Orlean, LA, Mar. 2015, pp. 2215–2220.

[34] Z. Jiang and S. Mao, “Opportunistic spectrum sharing in lte-unlicensed with lyapunov optimization based auction,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5217–5228, June 2017.

[35] Y. Wang, S. Mao, and R. M. Nelms, “Distributed online algorithm for optimal real-time energy distribution in the smart grid,” *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 70–80, 2014.

[36] S. Mao and S. S. Panwar, “A survey of envelope processes and their applications in quality of service provisioning,” *IEEE Communications Surveys & Tutorials*, vol. 8, no. 3, pp. 2–20.

[37] H. Abdelnasser, K. A. Harras, and M. Youssef, “Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator,” in *Proc. IEEE MobiHoc’15*, Hangzhou, China, June 2015, pp. 277–286.

[38] Z. Yang, P. Pathak, Y. Zeng, X. Liran, and P. Mohapatra, “Monitoring vital signs using millimeter wave,” in *Proc. IEEE MobiHoc’16*, Paderborn, Germany, July 2016, pp. 211–220.

[39] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, “Tracking vital signs during sleep leveraging off-the-shelf WiFi,” in *Proc. ACM Mobihoc’15*, Hangzhou, China, June 2015, pp. 267–276.

[40] X. Wang, C. Yang, and S. Mao, “PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices,” in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 1–10.

[41] ——, “Tensorbeat: Tensor decomposition for monitoring multi-person breathing beats with commodity WiFi,” *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 1, pp. 8:1–8:27, Sept. 2017.

[42] ——, “ResBeat: Resilient breathing beats monitoring with online bimodal CSI data,” in *Proc. IEEE GLOBECOM 2017*, Singapore, Dec. 2017.

[43] H. Aly and M. Youssef, “Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones,” in *Proc. IEEE INFOCOM’16*, San Francisco, CA, Apr. 2016, pp. 1–9.

[44] Y. Ren, C. Wang, J. Yang, and Y. Chen, “Fine-grained sleep monitoring: Hearing your breathing with smartphones,” in *Proc. IEEE INFOCOM’15*, Hong Kong, China, Apr. 2015, pp. 1194–1202.

[45] R. Nandakumar, S. Gollakota, and N. Watson, “Contactless sleep apnea detection on smartphones,” in *Proc. ACM MobiSys’15*. Florence, Italy: ACM, May 2015, pp. 45–57.

[46] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. ACM MobiCom'16*. New York City, NY: ACM, Oct. 2016, pp. 82–94.

[47] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proc. ACM CHI'16*. Santa Clara, CA: ACM, June 2016, pp. 1515–1525.

[48] F. Li, H. Chen, X. Song, Q. Zhang, Y. Li, and Y. Wang, "CondioSense: High-quality context-aware service for audio sensing system via active sonar," *Personal and Ubiquitous Computing J.*, vol. 22, no. 1, pp. 17–29, Feb. 2017.

[49] D. Stowell, D. Giannoulis, E. Benetos, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events," *IEEE Transactions on Multimedia*, vol. 17, no. 10, pp. 1733–1746, 2015.

[50] B. Fu, J. Karolus, T. Grosse-Puppendahl, J. Hermann, and A. Kuijper, "Opportunities for activity recognition using ultrasound doppler sensing on unmodified mobile phones," in *Proc. 2nd international Workshop on Sensor-based Activity Recognition and Interaction*. Rostock, Germany: ACM, June 2015, p. 8.

[51] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proc. ACM MobiSys'15*. ACM, May 2015, pp. 15–29.

[52] W. Mao, J. He, and L. Qiu, "CAT: High-precision acoustic motion tracking," in *Proc. ACM Mobicom'16*. New York City, NY: ACM, Oct. 2016, pp. 491–492.

[53] J. Liu, Y. Wang, G. Kar, Y. Chen, J. Yang, and M. Gruteser, "Snooping keystrokes with mm-level audio ranging on a single phone," in *Proc. ACM Mobicom'15*. Paris, France: ACM, Sept. 2015, pp. 142–154.

[54] J. Wang, K. Zhao, X. Zhang, and C. Peng, "Ubiquitous keyboard for small mobile devices: Harnessing multipath fading for fine-grained keystroke localization," in *Proc. ACM Mobisys'14*. Bretton Woods, NH: ACM, June 2014, pp. 14–27.

[55] C. Qiu and M. W. Mutka, "Silent whistle: Effective indoor positioning with assistance from acoustic sensing on smartphones," in *A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2017 IEEE 18th International Symposium on*. IEEE, 2017, pp. 1–6.

[56] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan, "Dhwani: secure peer-to-peer acoustic NFC," vol. 43, no. 4, pp. 63–74, 2013.

[57] R. Nandakumar and S. Gollakota, "Unleashing the power of active sonar," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 11–15, 2017.

[58] 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*. ACM, 2017, pp. 42–55.

[59] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, "AudioGest: Enabling fine-grained hand gesture detection by decoding echo signal," in *Proc. ACM UBICOMP'16*. Heidelberg, Germany: ACM, Sept. 2016, pp. 474–485.

[60] C. A. Dimoulas, "Audiovisual spatial-audio analysis by means of sound localization and imaging: A multimedia healthcare framework in abdominal sound mapping," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 1969–1976, 2016.

[61] E. A. Bernal, X. Yang, Q. Li, J. Kumar, S. Madhvanath, P. Ramesh, and R. Bala, "Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors," *IEEE Transactions on Multimedia*, 2017.

[62] F. Mokaya, R. Lucas, H. Y. Noh, and P. Zhang, "Burnout: A wearable system for unobtrusive skeletal muscle fatigue estimation," in *Proc. IEEE/ACM ISPN'16*. Kobe, Japan: IEEE, Oct. 2016, pp. 1–12.

[63] C. Yang, G. Cheung, and V. Stankovic, "Estimating heart rate and rhythm via 3d motion tracking in depth video," *IEEE Transactions on Multimedia*, 2017.

[64] T. Wei and X. Zhang, "mTrack: High-precision passive tracking using millimeter wave radios," in *Proc. ACM Mobicom'15*. Paris, France: ACM, Sept. 2015, pp. 117–129.

[65] X. Wang, Z. Yu, and S. Mao, "DeepML: Deep LSTM for indoor localization with smartphone magnetic and light sensors," in *Proc. IEEE ICC 2017*, Kansas City, MO, May 2018.

[66] W. Yang, X. Wang, S. Cao, H. Wang, and S. Mao, "Multi-class wheat moisture detection with 5GHz Wi-Fi: A deep LSTM approach," in *Proc. ICCCN 2018*, Hangzhou, China, July/Aug. 2018.

[67] X. Wang, L. Gao, and S. Mao, "Phasefi: Phase fingerprinting for indoor localization with a deep learning approach," in *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[68] ——, "Csi phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, 2016.

[69] X. Wang, L. Gao, S. Mao, and S. Pandey, "Deepfi: Deep learning for indoor fingerprinting using channel state information," in *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*. IEEE, 2015, pp. 1666–1671.

[70] ——, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[71] X. Wang, X. Wang, and S. Mao, "Cifi: Deep convolutional neural networks for indoor localization with 5 ghz wi-fi," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.

[72] X. Wang, L. Gao, and S. Mao, "Biloc: Bi-modal deep learning for indoor localization with commodity 5ghz wifi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[73] X. Wang, X. Wang, and S. Mao, "Resloc: Deep residual sharing learning for indoor localization with csi tensors," in *Proc. IEEE PIMRC 2017*, Montreal, Canada, Oct. 2017.