Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices

Kun Qian*, Chenshu Wu†, Fu Xiao‡, Yue Zheng*, Yi Zhang*, Zheng Yang*, Yunhao Liu*

*School of Software and TNLIST, Tsinghua University
†Department of Electrical and Computer Engineering, University of Maryland, College Park
‡College of Computer, Nanjing University of Posts and Telecommunications
{qiank10, wucs32, cczhengy, zhangyithss, hmilyyz, yunhaoliu}@gmail.com, {xiaof}@njupt.edu.cn

Abstract—Vital signs such as heart rate and heartbeat interval are currently measured by electrocardiograms (ECG) or wearable physiological monitors. These techniques either require contact with the patient's skin or are usually uncomfortable to wear, rendering them too expensive and user-unfriendly for daily monitoring. In this paper, we propose a new non-invasive technology to generate an *Acousticcardiogram (ACG)* that precisely monitors heartbeats using inaudible acoustic signals. ACG uses only commodity microphones and speakers commonly equipped on ubiquitous off-the-shelf devices, such as smartphones and laptops. By transmitting an acoustic signal and analyzing its reflections off human body, *ACG* is capable of recognizing the heart rate as well as heartbeat rhythm. We employ frequency-modulated sound signals to separate reflection of heart from that of background motions and breath, and continuously track the phase changes of the acoustic data. To translate these acoustic data into heart and breath rates, we leverage the dual microphone design on COTS mobile devices to suppress direct echo from speaker to microphones, identify heart rate in frequency domain, and adopt an advanced algorithm to extract individual heartbeats. We implement *ACG* on commercial devices and validate its performance in real environments. Experimental results demonstrate *ACG* monitors user's heartbeat accurately, with median heart rate estimation error of 0.6 beat per minute (bpm), and median heartbeat interval estimation error of 19 ms.

I. INTRODUCTION

Mobile health sensing is an emerging area that has attracted significant interests from both the industry and the research sides [1], [2], [3]. An attractive vision is imagined to monitor main vital signs (e.g., pulse rate, respiration rate, blood pressure, etc.) using daily mobile and smart devices. Vital signs are routinely monitored by medical professionals using dedicated equipment usually at a very limited frequency of only once or twice a year. To live a healthy lifestyle, however, it is critical to manage personal health data with a much finer granularity and in particular keep tabs on vital signs more regularly and frequently, or even daily for best. Mobile health sensing, thanks to the rapid development of Internet of Things and mobile sensing, provides a pervasive, user-friendly, sustainable and affordable chance to achieve the above goals. If we can monitor vital signs with everyday smartphones, we are able to track measurements of these otherwise elusive signs anywhere and anytime and keep health records day-by-day.

Significant efforts have been devoted to promote mobile vital sign monitoring, such as pulse rate [4], blood pressure [5], temperature [6], respiratory rate [7], etc. Different from these indicators, monitoring of *cardiac rhythm*, an even critical
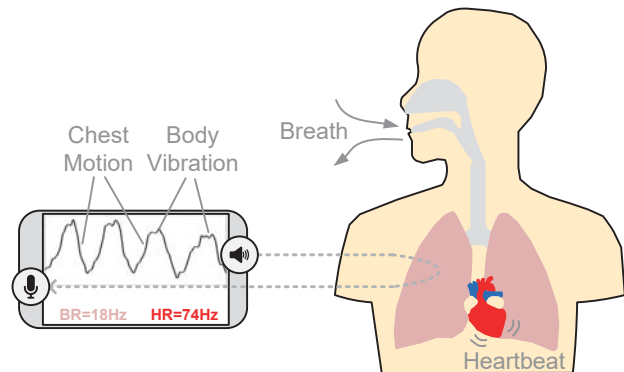

Fig. 1. Illustration of ACG.

vital sign among others, however, is non-trivial and far from being reality on smart devices. Daily cardiac monitoring is especially important to diagnose heart rhythm disorders because such disorders are usually sporadic and might not present themselves during one single or several doctor's visits. Although heart rhythm disorders are not life threatening most of the time, they seriously risk the sufferers life when they are accompanied by dizziness, fainting, palpitations, and otherwise unexplained strokes, etc. Conceptually, heart rhythm disorders are irregularities in the rhythm of the heart that can cause the heart to flutter, skip a beat, or skip between the two for a short time. Thus to detect them, one needs to monitor continuous heartbeats precisely.

Besides medical inspections, heartbeats are typically measured by wearable or implantable cardiac monitors that record the electrical signals of the heart and produce the well-known electrocardiograms (ECG). Despite of the precise and reliable reports, the use of wearable or implantable devices is cumbersome and usually expensive, making them suitable for confirmed patients but less-than-ideal for regular usage. Existing approaches for mobile heartbeat monitoring resort to radio frequency (RF) signals [8] or smartphone cameras [9]. While RF-based approach like EQ-Radio [8] provides an attractive contactless manner, it is vulnerable to surrounding environments and is available only in limited areas with prior deployed RF devices. Image-based methods like Cardiio [2] and HemaApp [10] employ smartphone cameras to sense the patterns of blood changes and to further infer heartbeats. However, they require the user to put a finger tightly on the

camera for a sufficient period.

In this paper, we present a non-invasive technology for heartbeat monitoring. Our method uses only the commodity off-the-shelf smartphones and does not require the user to carry any on/in-body devices. Our design uses inaudible acoustic signals to sense heartbeat motion and generates an *Acousticcardiogram* (ACG) of heart rhythms, as illustrated in Figure 1. The key insight is to turn the phone into an active sonar that transmits high frequency inaudible signals and captures the signals reflected off the human chest. These reflections are encoded by chest motions caused by breath as well as heartbeat and can be decoded for vital sign monitoring.

Technically, *ACG* addresses three critical challenges. First, the speakers and microphones on commodity smart devices are mainly designed for general purpose, and thus lack hardware solution to cancel strong power leakage from the speaker directly to the microphone, which would heavily overwhelm the reflection signals of interests. To overcome this, *ACG* leverages two microphones available on common smart devices to achieve pseudo self-interference cancellation. Since the direct power leakage from the speaker to the two microphones are much stronger than reflections from the human chest, it is able to align, scale the power leakages and cancel it by calculating the difference of the signals of two microphones.

Second, the human chest motion is minute and the heartbeat motion is even weaker. Previous work [7] employs frequency modulated continuous wave (FMCW) to separate reflections from human chest from other environmental reflections, and track the frequency shifts caused by the chest motion. However, it still cannot spot heartbeat signal from the overall reflections, as both the amplitude and the frequency shift caused by the heartbeat is usually beyond the spatial resolution of the FMCW sonar. Instead, we track signal phase within target spatial range, which contains the motion of target chest, to obtain heart rhythms measurement. To obtain the signal phase, we introduce FMCW front-end that down-converts the received passband signal to baseband complex signal.

Third, chest motions are induced by not only heartbeat but also human breathing. However, the heartbeat signal is orders of magnitude weaker than the breath signal, and thus submerged in breath signal as well as other irrelevant noises. By observing that heart signal is quasi-periodic and that the heart rate is usually higher than breath rate, we find the heart rate can be accurately estimated in frequency domain with enough data. To achieve this, we propose a novel comb notch filter that is adaptive to real-time estimation of the heart rate, which separates the fundamental frequency and harmonics of the heart signal from out-of-band noises and breath signal. Then we further apply an EM algorithm [8] in RF sensing field to obtain accurate measurement of heartbeats.

To evaluate the performance of *ACG*, we implement it as a third-party APP on commercial smartphones and recruit a group of 10 participants for testing. Experimental results demonstrate elegant performance of *ACG*, which monitors users heartbeat accurately, with a median heart rate error of 0.6 bpm, and a median heartbeat period error of 19 ms. As comparison, *ACG* achieves comparable accuracy with the
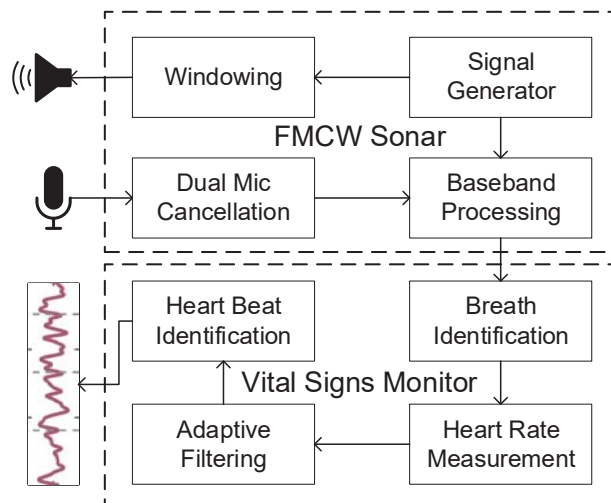

Fig. 2. Work flow of ACG.

vision-based contact approach (by framing finger), which has a median heart rate error of 0.4 bpm and a median heartbeat period error of 21 ms, and outperforms the non-contact approach (by framing face), which has a median heart rate error of 1 bpm and a median heartbeat period error of 30 ms.

In summary, our core contributions are as follows.

- We design a general FMCW sonar front-end that fits for commercial audio systems on smart devices, and enable fine-grained baseband signal processing. A novel algorithm is proposed to leverage dual microphones available to cancel self-interference of the sonar.

- We model the relations between the FMCW baseband signal phase information and the chest movements due to breath and heartbeat, and propose a novel contactless technique that monitors vital signs including heart rate and heartbeat interval by tracking fine-grained baseband signal phase information. As far as we are aware of, it is the first acoustic-based work that obtains accurate vital signs contactlessly on smart devices.

- We implement *ACG* on commodity smart devices and validate its effectiveness with various parameters and scenarios settings. Experimental results demonstrate that *ACG* achieves comparable accuracy with popular vision-based contact approach and significantly outperforms the vision-based contactless approach on smart devices.

In the rest of the paper, we first provide the overview of *ACG* in Section II, followed by the design and implementation of FMCW sonar in Section III and vital signs monitor in Section IV. Then, performance evaluation and parameter study of *ACG* are provided in Section V. Finally, related works are reviewed in Section VI and conclusion is drawn in Section VII.

II. ACG OVERVIEW

ACG is a passive vital signs monitoring system using off-the-shelf audio system on smart devices (e.g. phones, laptops). Figure 2 illustrates the logic process of *ACG*. *ACG* mainly consists of two parts, FMCW sonar and vital signs monitor. First,
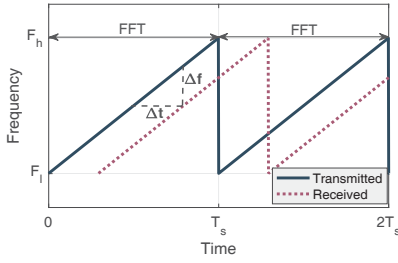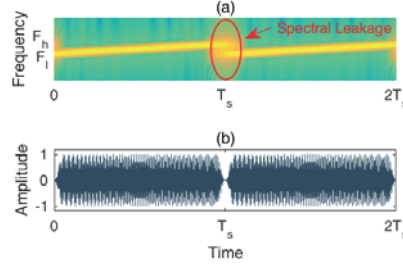
Fig. 3. FMCW sonar processing.

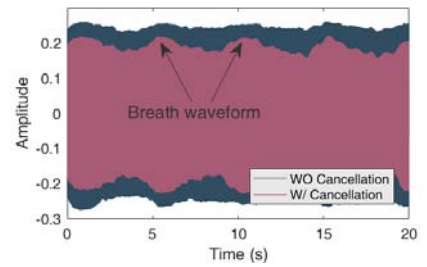Fig. 4. FMCW Signal design. (a) Spectrogram of chirp signal. (b) Windowing chirp signal.

Fig. 5. Illustration of dual microphone cancellation.

ACG simulates an FMCW sonar [3] that transmits inaudible chirp signals through the speaker and capture reflections with the microphones. The reflected signal is down-converted to complex baseband signal. Then, *ACG* detects the signal phase that contains user's chest motion, removes the breath signal and extracts heartbeat signal. The heart rates and heartbeat intervals are estimated.

The main technical challenge for *ACG* is the extremely low signal-to-interference ratio (SIR) and signal-to-noise ratio (SNR) for heartbeat identification. Specifically, these interferences and noises obfuscate the harmonic components of the periodic heartbeat signal, which is necessary for identifying individual heartbeats. On one hand, to mitigate random environmental and internal hardware noises, *ACG* tracks down-converted baseband signal phase, which is more robust against noises than the frequency shift [7]. On the other hand, the interferences stem from various sources including power leakage directly from speaker to microphone, large chest motion and other unconscious body motions. To cancel power leakage, *ACG* leverages dual microphone design on modern smart devices to achieve pseudo self-interference cancellation. To reduce impacts of irrelevant motions, *ACG* adopts filters adaptive to vital signs parameters to amplify heartbeat signal.

III. FMCW SONAR

ACG implements an FMCW sonar to capture signals reflected by moving chest and vibrating body. This section provides the technical preliminaries, fundamental model and practical issues of the sonar design.

A. FMCW Sonar Principle

An FMCW sonar transmits a chirp signal, whose instantaneous frequency increases linearly during the predefined sweeping period T_s, as shown in Figure 3. The transmitted signals reflect off the reflectors in the environment and arrive back at the sonar after some time delay. Since the transmitted frequency increases linearly, the sonar can determine the time delay by comparing the received and transmitted signal frequency. For example, in Figure 3, the time delay Δ_t can be calculated as $\Delta_t = \frac{\Delta_f}{k}$, where Δ_f is the frequency shift between the transmitted and received signal and $k = \frac{F_h - F_l}{T_s}$ is the slope of linear frequency sweeping. The range of the target reflector is thus $d = \frac{v\Delta_t}{2}$, where v is the speed of the signal.

When multiple reflectors locate at different distances from the receiver, their reflections have different frequency shifts, which can be distinguished by performing Fourier transform over a sweeping period of the signal, as in Figure 3. Essentially, FMCW sonar acts as a spatial filter that separates spatially different reflections, which is the fundamental principle for monitoring vital signs in multipath-rich daily environment.

B. FMCW Signal Design

In time domain, the chirp signal transmitted by FMCW sonar is:

$$s(t) = A\cos(2\pi(f_c t + \frac{B(t - NT_s)^2}{2T_s}))$$ (1)

where $t \in (NT_s - \frac{T_s}{2}, NT_s + \frac{T_s}{2}]$, $N \in \mathcal{Z}$. The parameters of a chirp signal include: A, the amplitude of the signal; $f_c = \frac{F_h + F_l}{2}$, the carrier frequency; $B = F_h - F_l$, the bandwidth; and T_s, the sweep time; To make the chirp signal inaudible, the upper and lower frequency F_h, F_l have to be larger than $16kHz$ [11]. A common parameter setting is $F_h = 19kHz$, $F_l = 17kHz$, and $T_s = 10.7ms$, corresponding to 512 samples under the typical $48kHz$ sampling rate of phone speakers. However, as Figure 4a shows the spectrogram of the chirp signal, strong power leakage can be spotted due to frequency hopping from F_h to F_l between successive sweeps, making the chirp signal audible. Thus, a tapered cosine window [12] is applied to the chirp signal, in order to eliminate the audible noises caused by spectral leakage with most signal samples unaltered. The windowing factor r determines the length of cosine parts of the window. The final FMCW signal is shown in Figure 4b.

C. Dual Microphone Cancellation

As speakers and microphones on commercial smart devices are separated for general use, the direct power leakage from speaker to microphone cannot be removed by audio hardware, which causes severe self-interference and may lead to mistaken identification of target reflection. Figure 5 shows an example of raw sound segment (blue) captured by single microphone. With the existence of strong direct path between the speaker and microphone, it is hard to spot the weak breath signal.

Existing art [13] in RF sensing field uses two transmit antennas to cancel direct power leakage at the receive antenna. However, this approach does not work for *ACG*. The reasons are two folds. First, smart phone commonly has two
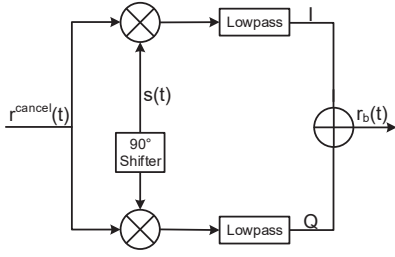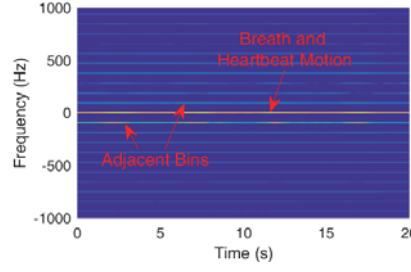
Fig. 6. *ACG* front end.
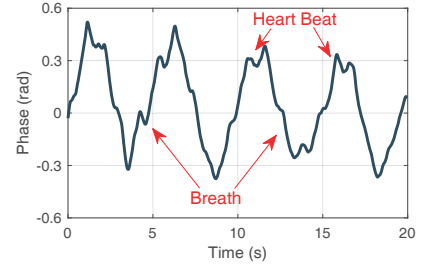

Fig. 7. Baseband spectrogram.

Fig. 8. Signal phase of the first PCA component.

pairs of co-located speakers and microphones. Playing sounds with any speaker may saturate the corresponding microphone with power leakage, making it infeasible to sense reflections. Second, the speakers on smart phone are designed for different uses (e.g. communication, playing sound) and thus highly heterogeneous. It is hard to perform equalization on commercial audio system for FMCW chirp signals.

Instead, we leverage two microphones available on smart phones to achieve interference cancellation. Specifically, suppose one speaker plays FMCW signal and two microphones receive $r_1(t)$ and $r_2(t)$ respectively, *ACG* estimates sub-sample delay δ_t with the phase slope changing in frequency domain, and further calculates correlation c between two aligned signals [14]:

$$\delta_t = \min_\delta \|\angle(\mathcal{F}[r_1(t)]\mathcal{F}^*[r_2(t)]) + 2\pi f \delta\|$$
$$r_2^{\text{shift}}(t) = \mathcal{F}^{-1}[\mathcal{F}[r_2(t)] \cdot e^{-j2\pi f \delta_t}] \qquad (2)$$
$$c = \frac{r_1(t) \cdot r_2^{\text{shift}}(t)}{\|r_1(t)\|\|r_2^{\text{shift}}(t)\|}$$

where $\mathcal{F}[\cdot]$ denotes Fourier transform. Since direct power leakages from the speaker to the microphones are the strongest components, we can approximate the estimation above as the delay and amplitude ratio of the power leakages of the two microphones. Thus, *ACG* scales $r_2^{\text{shift}}(t)$ with c, and subtracts it from $r_1(t)$:

$$r(t)^{\text{cancel}} = r_1(t) - c r_2^{\text{shift}}(t). \qquad (3)$$

Note that the parameters only need to be calculated once at the beginning of the monitoring, and each time after the device is moved. *ACG* carries out cancellation discretely chunk by chunk (e.g. 2s periods), and splices the outputs to form the cancellation signal. Figure 5 shows the output of cancellation (red), where breath waveform can be clearly identified.

D. Baseband Signal Processing

The act of breath and heartbeat cause vital motions such as minute chest motion and body vibration that modulate FMCW reflections. However, such minute motion is dramatically smaller than the ranging resolution of FMCW sonar:

$$\delta_d = \frac{v}{2B} \qquad (4)$$

where v is the sound wave speed, and B is the bandwidth of chirp signal. Specifically, with a typical bandwidth of $2kHz$, the ranging resolution of FMCW sonar is $\frac{343m/s}{2 \times 2000Hz} = 8.6cm$.

In contrast, breath and heartbeat motions have sub-centimetre level amplitudes.

Due to limited bandwidth of inaudible sound available on commercial audio system, it is impossible to directly range breath and heartbeat motion with frequency shift. To address this problem, *ACG* down-converts the FMCW signal to baseband, and continuously tracks the signal phase of the spatial bin that contains vital motions. Figure 6 shows the front end used by *ACG*.

The rationale is that while vital motions are hidden within the spatial bin, they change the range of reflection and thus modulate the signal phase of the bin. Specifically, with the propagation delay τ_t ($< \frac{T_s}{2}$) and attenuation $\alpha < 1$, the reflection of vital motion $r(t)$ is:

$$r(t) = \alpha\cos(2\pi(f_c(t - \tau(t)) + \frac{B(t - \tau(t) - N'T_s)^2}{2T_s})) \quad (5)$$

where $N' = N - 1$ when $t \in P_1 = (NT_s - \frac{T_s}{2}, NT_s - \frac{T_s}{2} + \tau(t)]$, and $N' = N$ when $t \in P_2 = (NT_s - \frac{T_s}{2} + \tau(t), NT_s + \frac{T_s}{2}]$. Based on the system structure shown in Figure 6, the baseband signal $b(t)$ is derived as:

$$r_b(t) = \begin{cases} \alpha e^{j2\pi(f_c\tau(t)+\frac{B(t-NT_s)(\tau(t)-T_s)}{T_s}-\frac{B(\tau(t)-T_s)^2}{2T_s})} & t \in P_1 \\ \alpha e^{j2\pi(f_c\tau(t)+\frac{B(t-NT_s)\tau(t)}{T_s}-\frac{B\tau(t)^2}{2T_s})} & t \in P_2 \end{cases}$$
$$(6)$$

ACG calculates the spectrogram of $r_b(t)$ and selects the signal $b(t)$ in spatial bin with frequency shift $f = \frac{\tau(t)B}{T_s}$:

$$b(t) = 2\alpha\frac{T_s - \tau(t)}{T_s}e^{j2\pi(f_c\tau(t)-\frac{B\tau(t)^2}{2T_s})}$$
$$+ 2\alpha\frac{\sin(2\pi B\tau(t))}{2\pi B}\frac{\tau(t)}{T_s}e^{j2\pi(f_c\tau(t)-\frac{B(\tau(t)^2-T_s^2)}{2T_s})} \qquad (7)$$

On one hand, the second term in Equation 7 is due to frequency hopping at the end of each sweep, where the frequency difference ambiguously changes from Δf to $B - \Delta f$ (see Figure 3). As the strongest direct signal is eliminated by dual microphone cancellation, the main signal remained is the reflection of user's chest. Thus, *ACG* further aligns the received signal with the reference signal using cross correlation (recall Equation 2), and thus removes the second term. On the other hand, both amplitude and phase of the first term encode the vital motions. As signal amplitude is more vulnerable to random noises, *ACG* tracks vital motions with
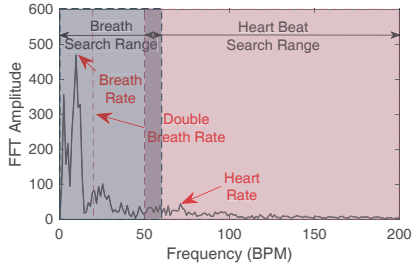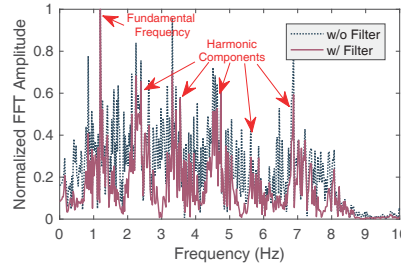
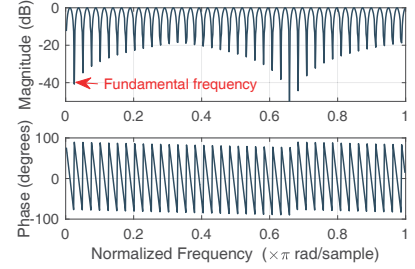Fig. 9. Spectrum of signal phase.　Fig. 10. Spectrum of the output of differentiator.　Fig. 11. Response of adaptive comb notch filter.

signal phase $\phi(t) = 2\pi(f_c\tau(t) - \frac{B\tau(t)^2}{2T_s})$. In typical settings, $f_c >> \frac{B\tau(t)}{2T_s}$, *ACG* omits the quadratic term as:

$$\phi(t) \approx 2\pi f_c\tau(t) \tag{8}$$

For implementation, *ACG* calculates the spectrogram of baseband signal, as shown in Figure 7. Specifically, multiple sweeps (e.g. 10) of the chirp signal are grouped for Fourier transform to mitigate high frequency random noises. The spatial bin with highest time variation and its adjacent bins are selected. Time variations of bins are estimated by calculating the average difference of bin values that are spaced at some constant time (e.g. 0.4s). Then *ACG* performs PCA analysis on the selected bins and tracks the phase of the first PCA components, which mainly contains the target reflections. Figure 8 shows the signal phase of the first PCA component, where breath-induced chest motion and heartbeat-induced body vibration can be spotted. To eliminate the impact of displacement and unconscious body motion, *ACG* smooths the signal, and subtracts the smoothed output from the signal. The span of the smooth filter is set to 5s.

IV. VITAL SIGNS MONITOR

This section details the process of extracting vital signs including breath rate, heart rate and individual heartbeat from acoustic signal phase.

A. Breath Identification

As breath signal is periodic and can be clearly spotted in signal phase (Figure 8), *ACG* extracts the breath rate by performing Fourier transform on the signal phase, and the spectrum is shown in Figure 9. To estimate breath rate, *ACG* searches the most prominent peak in a predefined frequency range for normal breath (e.g. $0 \sim 60$ BPM). The frequency of the peak yields a coarse estimation of breath rate.

However, the accuracy of coarse estimation is limited by the resolution of Fourier transform, which is determined by the time length of the signal. To obtain more accurate measurements, *ACG* performs inverse Fourier transform on the dominant peak and its adjacent two bins to obtain a complex time-domain signal $\phi'(t)$. As $\phi'(t)$ only contains single dominant breath frequency, its phase is quasi-linear and its slope corresponds to the breath rate. Thus, a finer estimation can be obtained as:

$$\text{BR} = 60 \cdot \frac{\text{slope}(\angle\phi'(t))}{2\pi} \tag{9}$$
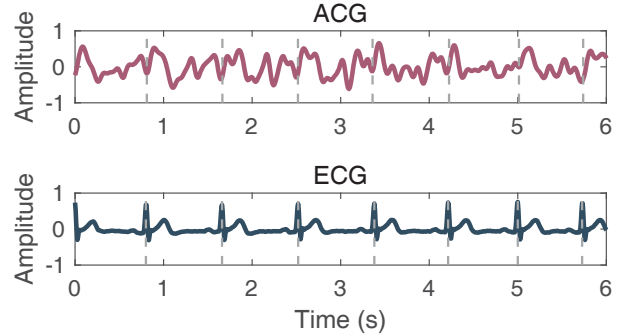

Fig. 12. Comparison of heartbeats in ACG and ECG.

B. Heart Rate Measurement

As heartbeat signal is also periodic, the same process can be applied to estimate heart rate. However, since breath signal is orders of magnitudes stronger than heartbeat signal, the search range of heart rate should be carefully defined to avoid obfuscations by breath signal. The major concern is that the breath signal is not perfect sine wave, and may contain strong harmonic components. By trading off the impact of breath signal and detection range of heartbeat, *ACG* sets the lower bound of the search range as $\max(2 \cdot \text{BR}, 50)$ BPM.

Similar to breath, two-step estimation is carried out to obtain accurate heart rate. Figure 9 shows the peak corresponding to the heart rate.

C. Heartbeat Extraction

Extracting individual heartbeat incurs more challenges than estimating breath and heart rates. First, the heartbeat signal is obfuscated by the orders of magnitude stronger breath signal. Second, the heartbeat waveform lack sharp peaks as in ECG signal, making it harder to accurately identify heartbeat intervals. Existing art [8] in RF sensing field achieves accurate heartbeat extraction by two-step process. First, a second order noise robust differentiator [15] is applied to the signal phase to mitigate low frequency breath signal, and amplify high frequency heartbeat signal. Second, it treats the heartbeat signal as successive multiple copies of the heartbeat waveform with different scale, and uses expectation maximisation (EM) algorithm to jointly estimate heartbeat intervals and the template of heartbeat waveform.
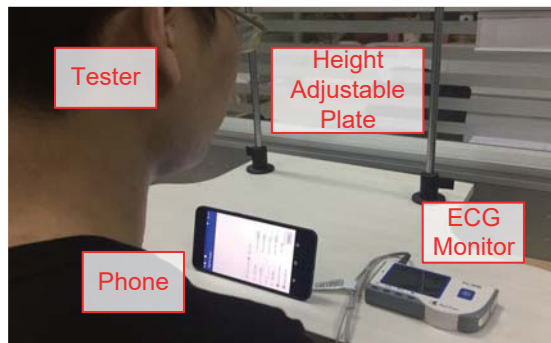
Fig. 13. Experiment setup.

However, the RF-based heartbeat extraction algorithm cannot be directly applied to the acoustic signal due to its dramatically low SNR. Specifically, the strong noises may obfuscate the harmonic components of the heartbeat signal, and drive the EM algorithm to oscillate or converge at wrong local minima. Figure 10 shows the normalized spectrum of the output of the differentiator (blue). The differentiator mitigates the breath signal at 0.17 Hz, and amplifies the heartbeat signal at 1.19 Hz. However, the differentiator also amplifies the high frequency noises, which submerges the harmonic components of heartbeat signal.

Since heartbeat signal is quasi-periodic, its power should concentrate on the fundamental frequency and harmonic components. Thus, *ACG* adopts IIR comb notch filter that is adaptive to heart rate to extract heartbeat signal from noise background. As shown in Figure 11, the order of the filter is set as $\frac{\text{SamplingRate}}{\text{HeartRate}}$. Denote the filter as $h(t)$, then the heart signal, $\phi_h(t)$, is calculated as:

$$\phi_h(t) = \phi(t) - h(t) \otimes \phi(t) \tag{10}$$

where $\otimes$ is the convolution operation. *ACG* then filters the heart signal $\phi_h(t)$ with the second order noise robust differentiator, and applies the EM algorithm [8] to extract individual heartbeats. Figure 12 shows examples of heartbeat extraction results from both *ACG* and ECG. Heartbeats are accurately estimated with low SNR acoustic signal.

V. EVALUATION

This section presents the experimental settings and the detailed performance of *ACG*.

A. Experimental Methodology

Implementation. We implement *ACG* as a third-party APP on recent Android platform, e.g. Google Nexus 6P with Android 6.0 OS. *ACG* plays FMCW signal samples on the communication speaker and continuously records the raw samples from the two microphones. The communication speaker and its co-located microphone are at the top of the phone, and the other microphone is at the bottom of the phone. When the microphone plays at full volume, the SNR of the acoustic signal at 5cm away is about 30dB. We implement baseband signal processing, breath identification and heart rate measurement as C functions using Android NDK to achieve better efficiency. The heart signal is exported to the PC for heartbeat extraction, which is implemented in MATLAB. Meanwhile, since the iOS platform only authorizes single channel recording to third-party developers, so we do not implement *ACG* on the iOS platform.

Evaluation Setup. The experiments are conducted in standard office environments, where exist various sounds caused by server running, human talking and tapping keyboards, etc. As illustrated in Figure 13, we place the phone vertically on a height-adjustable plate, facing towards the user. A commercial 3-lead ECG monitor, Heal Force PC-80B, is used to obtain the ground truth. The sampling rate of the ECG monitor is 150 Hz, which converts to the time resolution of 6.7 ms. Experiments are carried out in groups, each group lasts 1 minute. During the experiment, testers wear single coats, such as shirts and dresses, sit in front of the phone and breathe normally. Both audio signal and ECG signal are recorded for further performance evaluation. Various factors including FMCW signal parameters, user and device placement diversity, and key processes of *ACG* are considered in the experiment.

B. Overall Performance

Performance of vital signs measurement. We first report overall performance of *ACG* in measurement of heart rate and heartbeat interval. For better clarification, we compare *ACG* with two state-of-the-art vision-based heart monitor approaches on smart devices, which track light absorbing of the finger (contact) and the face (non-contact). A commercial heart monitor APP, Cardiio [2], is used to obtain measurements of these two approaches. For brevity, we denote the two approaches as FINGER and FACE.

Figure 14a plots the CDF of heart rate measurement errors of three approaches over all experiments. As illustrated, *ACG* achieves a median error of 0.6 bpm, which is slightly worse than FINGER, yet superior to FACE. The explanation of evaluation results are two folds. First, both *ACG* and FACE are non-contact approaches, which severely suffer from user's unconscious body motions. In contrast, FINGER requires contact of user's finger with flash light and phone camera, and thus obtain stable measurements. Second, in addition to user's motion, FACE is further limited by ambient lighting conditions. For example, the approach even fails in extreme weak or strong lighting conditions.

The CDF of heartbeat interval error in Figure 14b demonstrates similar evaluation results. Specifically, *ACG* achieves a median error of 18.7 ms, which is better than FINGER (median error of 21.7 ms) and FACE (30.0 ms). The high average accuracy of *ACG* is attributed to the high sampling rate of the acoustic signal on smart devices. Specifically, by setting the sampling rate of microphone to 48 kHz and the number of samples per sweep of FMCW signal to 512, *ACG* obtains $\frac{48000}{512} = 93.75$ Hz heart signal. In contrast, the vision method is limited by the frame rate of the camera, which is no more than 60 fps on common smart devices. However, the error tail of *ACG* is larger than FINGER and comparable to FACE, which is due to unconscious body motions of the user.

Relation between heart rate and heartbeat interval. Figure 15 shows the relation between heart rate error and
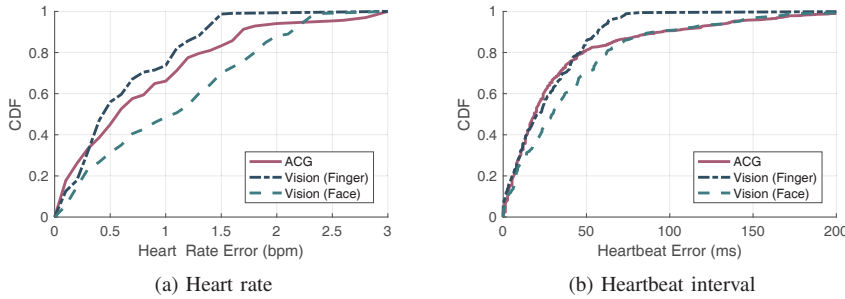
(a) Heart rate

(b) Heartbeat interval

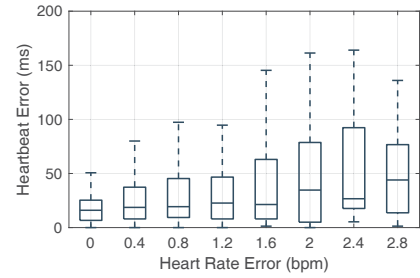Fig. 14. Overall performance of *ACG*.

Fig. 15. Relation between heart rate and heartbeat interval.

heartbeat interval error. Clearly, the heartbeat error increases with the heart rate error. The experimental result is consistent with the system design. Specifically, as *ACG* uses comb filter whose order is determined by the estimation of heart rate, a deviation of the estimation may distort the spectrum of heart signal and retain some irrelevant out-of-band noises. Since both errors are positively correlated, we only demonstrate the heartbeat error of the rest experiments, due to space limitation.

C. Use Issue Study

This section studies the impact of various issues for practical use of *ACG*, including the user-device distances, user orientations, body parts and user diversity.

User-device distance. We evaluate the performance of *ACG* as users sit at various distances away from the phone. Figure 16 shows the impact of the distance. *ACG* achieves the highest accuracy when the user is 5 cm in front of the phone, and slightly lower accuracy with larger distance less than 30 cm. As the user moves away from the phone, the area illuminated by the acoustic signal becomes larger, which to some extent compensates the power loss due to longer travel distance. Since *ACG* targets near-field vital signs monitoring with portable smart devices, a monitor range of 30 cm is sufficient for practical use.

User orientation. To evaluate the impact of user orientation, we ask users to sit and face towards various directions, including left, front-left, front, front-right, and right. As shown in Figure 17, *ACG* achieves highest accuracy when the user faces towards the phone (front), and the accuracy degrades as the angle between the orientations of the user and the device. As the user orientation deviates from the phone orientation, the reflection surface changes from the front chest to side chest, causing the decrease of reflection area and weaker body vibration captured by reflections. It is also noted that the performance of right directions are better than that of left directions. It is because that the human heart is at the left side of the chest, causing stronger vibration on the left side than the right side.

Body part. Since acoustic signals played by phone speaker is much weak, only partial body part can be illuminated during monitoring. So we study the performance of *ACG* when monitoring different body parts, including heart, chest, head and abdomen. Specifically, for heart, we place the phone to directly face the heart of the user. As shown in Figure 18,

ACG achieves highest accuracy for chest, since it is close to the vibration source, i.e. heart. The performance for heart is slightly worse, which is due to the smaller area of the heart in comparison with chest. The performance for head is worse than the first two cases. We find that in this case, *ACG* cannot accurately track the user's breath and suffer from irrelevant and significant head motions. The performance for abdomen is the worst and unacceptable. The body vibration caused by heartbeat is the weakest at abdomen, and is easily obfuscated by harmonics of abdomen motions caused by breath. As a result, it is recommended to use *ACG* to monitor heart or chest for the best performance.

User diversity. To evaluate the robustness of *ACG* for various users, we recruit 10 participants (5 males, 5 females) to test *ACG*. The testers have various ages (20~40) and body conditions. Figure 19 shows the performance of *ACG* with different participants. The accuracy of *ACG* for males are statistically lower than that for females, in that females have weaker chest motions. Except for the users 7 and 9, *ACG* achieves median measurement errors of about 20 ms for the rest testers. We carefully check both acoustic and ECG signals recorded for the users 7 and 9. The user 7 has much abnormal movements during the test, which obfuscates weak body vibration caused by heartbeat. The user 9 has significant cardiac arrhythmia. Her heart signal has a wider frequency band and thus suffers more from random noises. The further study of these special cases is left as future work.

D. Key Processes Study

This section provides study on the performance of key processes used in *ACG*, including dual microphone cancellation, baseband signal phase processing and application of comb filter.

Dual microphone cancellation. *ACG* leverages dual microphone cancellation to eliminate power leakage directly from the speaker to the microphone and amplify reflections from user's chest. To evaluate the impact of the the dual microphone cancellation, we compare the performance of *ACG* using dual microphones and using single microphone. Specifically, in the latter scheme, *ACG* uses the microphone that is not co-located with the speaker to avoid overwhelmed power leakage from the speaker. As shown in Figure 20, with dual microphone cancellation, the median error of *ACG* decreases from 27.9 ms to 18.7 ms, and the error tail is shortened from 400 ms
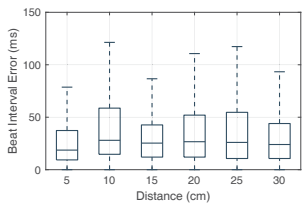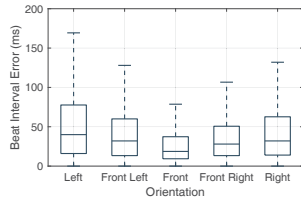
Fig. 16. Impact of user-device distance.

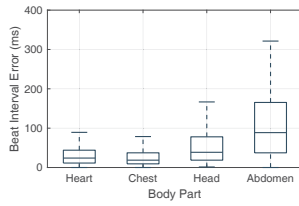Fig. 17. Impact of user orientation.

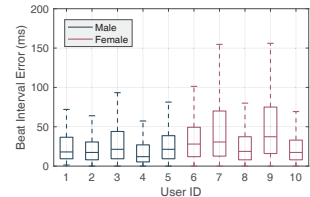Fig. 18. Impact of illuminating body part.
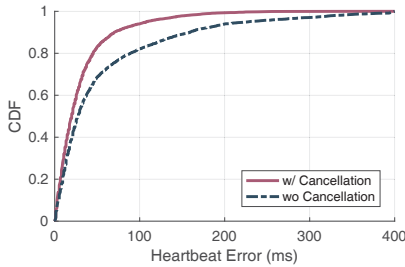
Fig. 19. Impact of user diversity.

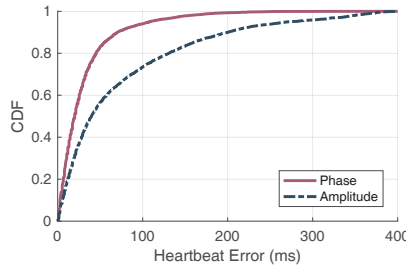Fig. 20. Impact of dual microphone cancellation.
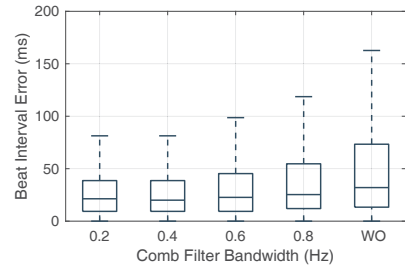
Fig. 21. Impact of signal phase processing.

Fig. 22. Impact of adaptive filter.

to 200 ms, demonstrating the effectiveness of the cancellation process.

Baseband signal phase processing. While Equation 7 shows that both amplitude and phase of the baseband signal encode vital motions, the signal amplitude is more vulnerable to random noises and *ACG* uses signal phase to track the vital signs. To evaluate the impact of usage of signal phase, we compare the performance of *ACG* with both signal amplitude and phase. As shown in Figure 21, by tracking baseband signal phase, the median error of *ACG* decreases from 40.7 ms to 18.7 ms, and the error tail is shortened from 400 ms to 200 ms.

Adaptive filter. *ACG* uses comb notch filter to extract quasi-periodic heart signal from out-of-band noises. The -3 dB bandwidth of the filter notch determines the extent that the noise is filtered. Specifically, the smaller the bandwidth is, the more the noise is removed. To evaluate the impact of adaptive comb filter, we compare the performance of *ACG* that uses comb filters with different notch bandwidth, as well as does not use the filter. As shown in Figure 22, the median error of *ACG* without the adaptive filter is 32.0 ms. In contrast, with comb filter, the median error decreases as the notch bandwidth of the filter decreases, and finally converges to 18.7 ms. In practice, *ACG* sets the filter notch bandwidth to 0.4 Hz, in order to avoid opportunistically significant heart rate variation.

VI. RELATED WORK

We categorize the related work into two parts, those related to vital sign monitoring, and those related to acoustic activity sensing.

Vital signs monitoring. Vital signs such as heart rate and heartbeat intervals are important indicators for human health condition, mental status and stress level. Traditional medical approaches use dedicated equipments such as Spirometer [16], electrocardiography (ECG) and echocardiography (ECHO) [17]. While promising, these devices incur high cost, require cumbersome installation and are vulnerable to electromagnetic interference. To enable daily monitoring of these information, various techniques have been developed. Prosperous wearable devices such as smart phones [18], watches [1] and bands [19] turn on-board cameras or photo-detectors into pulse oximeters. By attaching some body part (e.g. finger, wrist) and sensing the light absorbing variations of the blood, they can track pulsating nature of the arterial blood flow as well as the heart [20]. Other works leverage attached inertial sensors to detect tiny heartbeat motions [21], [22], [23], [24], [25]. These techniques require contact with the body or infrastructure (e.g. bed), making them uncomfortable and inconvenient to use.

Recent advance in computer vision and wireless sensing enables non-contact monitoring. Vision-based techniques either measure light absorbing of distant body part (e.g. face) as oximeters [9], [2], or track the tiny body vibration caused by heartbeat [4], [26]. However, they require strict illumination conditions, and have privacy concerns. In contrast, RF-based techniques track the impact of body motion caused by respiratory and heartbeats on RF signal. As COTS RF devices [27], [28], [29], [30] transmit single or multiple carrier continuous wave signal, they cannot separate the reflection containing breath and heartbeat motion, and thus can only roughly estimate respiratory and heart rate. While accurate measurement of vital signs can be obtained via specialized Radar system [3], [8], *ACG* leverages commercial audio system on nowadays smart phones to achieve accurate non-contact monitoring of vital signs.

Acoustic activity sensing. Passive acoustic sensing recognizes various activities directly from sound recording [31], [32], [33]. Sleep Hunter [34] tracks sleeping stage transitions from body movement and acoustic signal. Fine-grained sleep

monitoring is further achieved by capturing breath sound with earphones [35]. In contrast, active acoustic sensing uses active sonar that transmits inaudible signals for finer tracking [36], activity sensing [37] and vital signs monitoring [7]. LLAP [11] uses signal phase of single carrier continuous wave to track centimetre-level finger motions. Due to multipath effect, it requires target finger to move with distance of multiple wavelength to remove irrelevant reflections, which cannot be fulfilled in heartbeat monitoring. ApneaApp [7] leverages FMCW sonar to capture minute breath and identify sleep apnea event. In contrast, by carefully processing FMCW baseband signal phase information, *ACG* pushes the sensing limit to millimetre-level heartbeat.

VII. CONCLUSION

In this paper, we propose an acoustic-based contactless vital signs monitoring system *ACG*, that measures user's heart rhythms. First, we design a general FMCW sonar that fits for commercial speakers and microphones on smart devices, enabling fine-grained baseband signal processing. Then, we model the relations between the FMCW baseband signal phase information and the chest movements due to breath and heartbeat, amplify the heartbeat signal with adaptive filter and achieve accurate heart rhythms measurement. We implement *ACG* on commodity smart devices and evaluate it in real environments. Experimental results show that *ACG* achieves high accuracy superior to state-of-the-art vision-based contactless approach, with a median error of 0.6 bpm for heart rate and a median error of 19 ms for heartbeat interval. Further work focuses on robustly monitoring vital signs during user motion, extending monitoring range and further pushing accuracy towards contact approaches.

ACKNOWLEDGEMENT

REFERENCES

[1] "Apple watch," https://www.apple.com/watch/.
[2] "Cardiio," https://www.cardiio.com/.
[3] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proceedings of ACM CHI*, 2015.
[4] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of IEEE CVPR*, 2013.
[5] I. C. Jeong and J. Finkelstein, "Introducing contactless blood pressure assessment using a high speed video camera," *Journal of medical systems*, vol. 40, no. 4, pp. 1–10, 2016.
[6] "Shecare," http://www.ikangtai.com/.
[7] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proceedings of ACM MobiSys*, 2015.
[8] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016.
[9] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," 2012.
[10] E. J. Wang, W. Li, D. Hawkins, T. Gernsheimer, C. Norby-Slycord, and S. N. Patel, "Hemaapp: noninvasive blood screening of hemoglobin using smartphone cameras," in *Proceedings of ACM Ubicomp*, 2016.
[11] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of ACM MobiCom*, 2016.
[12] P. Bloomfield, *Fourier analysis of time series: an introduction.* John Wiley & Sons, 2004.
[13] F. Adib and D. Katabi, *See through walls with WiFi!*, 2013.
[14] "Delay estimation by fft," https://www.dsprelated.com/showarticle/26.php.
[15] "Noise robust differentiators for second derivativeestimation," http://www.holoborodko.com/pavel/downloads/NoiseRobustSecondDerivative.
[16] A. Lanatà, E. P. Scilingo, E. Nardini, G. Loriga, R. Paradiso, and D. De-Rossi, "Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 378–386, 2010.
[17] N. A. Goldstein, N. Sculerati, J. A. Walsleben, N. Bhatia, D. M. Friedman, and D. M. Rapoport, "Clinical diagnosis of pediatric obstructive sleep apnea validated by polysomnography," *OtolaryngologyHead and Neck Surgery*, vol. 111, no. 5, pp. 611–617, 1994.
[18] T. Han, X. Xiao, L. Shi, J. Canny, and J. Wang, "Balancing accuracy and fun: designing camera based mobile games for implicit heart rate monitoring," in *Proceedings of ACM CHI*, 2015.
[19] "Microsoft band," https://www.microsoft.com/microsoft-band/.
[20] L. J. Mengelkoch, D. Martin, and J. Lawler, "A review of the principles of pulse oximetry and accuracy of pulse oximeter estimates during exercise," *Physical therapy*, vol. 74, no. 1, pp. 40–49, 1994.
[21] S. Nirjon, R. F. Dickerson, Q. Li, P. Asare, J. A. Stankovic, D. Hong, B. Zhang, X. Jiang, G. Shen, and F. Zhao, "Musicalheart: A hearty way of listening to music," in *Proceedings of ACM Sensys*, 2012.
[22] H. Aly and M. Youssef, "Zephyr: Ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones," in *Proceedings of IEEE INFOCOM*, 2016.
[23] J. Hernandez, D. J. McDuff, and R. W. Picard, "Biophone: Physiology monitoring from peripheral smartphone motions," in *Proceedings of IEEE EMBC*, 2015.
[24] R. Mohamed and M. Youssef, "Heartsense: Ubiquitous accurate multimodal fusion-based heart rate estimation using smartphones," *Proceedings of ACM IMWUT*, 2017.
[25] Z. Jia, M. Alaziz, X. Chi, R. E. Howard, Y. Zhang, P. Zhang, W. Trappe, A. Sivasubramaniam, and N. An, "Hb-phone: a bed-mounted geophone-based heartbeat monitoring system," in *Proceedings of IEEE IPSN*, 2016.
[26] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proceedings of IEEE CVPR*, 2016.
[27] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-sleep: Contactless sleep monitoring via wifi signals," in *Proceedings of IEEE RTSS*, 2014.
[28] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: do user location and body orientation matter?" in *Proceedings of ACM Ubicomp*, 2016.
[29] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi," in *Proceedings of ACM Mobihoc*, 2015.
[30] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. N. Patel, J. A. Kientz, and T. Choudhury, "Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar," in *Proceedings of ACM Ubicomp*, 2015.
[31] J. Wang, K. Zhao, X. Zhang, and C. Peng, "Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization," in *Proceedings of ACM MobiSys*, 2014.
[32] M. Mirtchouk, C. Merck, and S. Kleinberg, "Automated estimation of food type and amount consumed from body-worn audio and motion sensors," in *Proceedings of ACM Ubicomp*, 2016.
[33] T. Yu, H. Jin, and K. Nahrstedt, "Writinghacker: audio based eavesdropping of handwriting via mobile devices," in *Proceedings of ACM Ubicomp*, 2016.
[34] W. Gu, Z. Yang, L. Shangguan, W. Sun, K. Jin, and Y. Liu, "Intelligent sleep stage mining service with smartphones," in *Proceedings of ACM Ubicomp*, 2014.
[35] Y. Ren, C. Wang, J. Yang, and Y. Chen, "Fine-grained sleep monitoring: Hearing your breathing with smartphones," in *Proceedings of IEEE INFOCOM*, 2015.
[36] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of ACM CHI*, 2016.
[37] 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 *Proceedings of ACM Ubicomp*, 2016.