Your Smart Speaker Can ”Hear” Your Heartbeat!

Fusang Zhang, Zhi Wang, Jin Beihong, Jie Xiong, Daqing Zhang

To cite this version:

Fusang Zhang, Zhi Wang, Jin Beihong, Jie Xiong, Daqing Zhang. Your Smart Speaker Can ”Hear” Your Heartbeat!. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies , 2020, 4 (4), pp.1-24. 10.1145/3432237 . hal-03363346

HAL Id: hal-03363346
https://hal.science/hal-03363346

Submitted on 30 Jan 2022

Your Smart Speaker Can "Hear" Your Heartbeat!

FUSANG ZHANG*, Institute of Software, Chinese Academy of Sciences and University of Chinese Academy of Sciences, China
ZHI WANG*, Institute of Software, Chinese Academy of Sciences and University of Chinese Academy of Sciences, China
BEIHONG JIN, Institute of Software, Chinese Academy of Sciences and University of Chinese Academy of Sciences, China
JIE XIONG, University of Massachusetts Amherst, USA
DAQING ZHANG, Peking University, China and Institut Polytechnique de Paris, France

Vital sign monitoring is a common practice amongst medical professionals, and plays a key role in patient care and clinical diagnosis. Traditionally, dedicated equipment is employed to monitor these vital signs. For example, electrocardiograms (ECG) with 3 -12 electrodes are attached to the target chest for heartbeat monitoring. In the last few years, wireless sensing becomes a hot research topic and wireless signal itself is utilized for sensing purposes without requiring the target to wear any sensors. The contact-free nature of wireless sensing makes it particularly appealing in current COVID-19 pandemic. Recently, promising progress has been achieved and the sensing granularity has been pushed to millimeter level, fine enough to monitor respiration which causes a chest displacement of 5 mm. While a great success with respiration monitoring, it is still very challenging to monitor heartbeat due to the extremely subtle chest displacement (0.1 - 0.5 mm) – smaller than 10% of that caused by respiration. What makes it worse is that the tiny heartbeat-caused chest displacement is buried inside the respiration-caused displacement. In this paper, we show the feasibility of employing the popular smart speakers (e.g., Amazon Echo) to monitor an individual's heartbeats in a contact-free manner. To extract the submillimeter heartbeat motion in the presence of other interference movements, a series of novel signal processing schemes are employed. We successfully prototype the first real-time heartbeat monitoring system using a commodity smart speaker. Experiment results show that the proposed system can monitor a target's heartbeat accurately, achieving a median heart rate estimation error of 0.75 beat per minute (bpm), and a median heartbeat interval estimation error of 13.28 ms (less than 1.8%), outperforming even some popular commodity products available on the market.

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

Additional Key Words and Phrases: Contactless sensing; Acoustic signal; Vital sign; Heart rate;

ACM Reference Format:
Fusang Zhang*, Zhi Wang*, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can "Hear" Your Heartbeat!. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 161 (December 2020), 24 pages. https://doi.org/10.1145/3432237

Authors' addresses: F. Zhang, Z. Wang, B. Jin, State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China; E-mail: {zhangfusang, wangzhi20, jbh}@otcaix.iscas.ac.cn. J. Xiong, College of Information and Computer Sciences, University of Massachusetts Amherst, USA; E-mail:jxiong@cs.umass.edu. D. Zhang, Key Laboratory of High Confidence Software Technologies (Ministryof Education), School of Electronics Engineering and Computer Science, Peking University, Beijing, China, Telecom SudParis, InstitutPolytechnique de Paris, Evry, France, dqzhang@sei.pku.edu.cn. ∗Both authors contributed equally to this work. Corresponding Author: Beihong Jin, Daqing Zhang; E-mail: Beihong@iscas.ac.cn, dqzhang@sei.pku.edu.cn.

2474-9567/2020/12-ART161 $15.00
https://doi.org/10.1145/3432237

1 INTRODUCTION

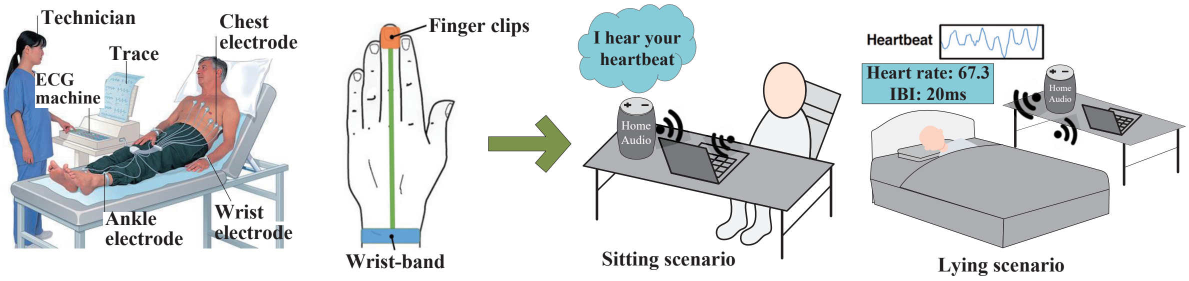
Vital sign monitoring is a common practice amongst medical professionals and plays a critical role in assisting clinical diagnosis, assessing the overall health of a patient, and indicating if a patient's physical condition is recovering or worsening. Traditionally, dedicated equipment operated by professionals is used to monitor the vital signs. For example, respiration is usually monitored by impedance photoplethysmography (PPG) [22] and heartbeat is monitored by electrocardiogram (ECG) [19]. While the achieved accuracy is high, these devices are usually expensive and need to be operated by professionals. To facilitate vital sign monitoring at home, portable devices and even wearables are developed [4, 6]. While the portability and cost are the advantages, the accuracy is slightly decreased compared to dedicated equipment.

In the last few years, wireless sensing becomes a hot research area and a lot of wireless sensing-based vital sign monitoring systems are proposed. The key difference between wireless sensing and traditional sensor-based approaches is that it does not require any dedicated equipment or sensor/wearable for sensing. The pervasive wireless signals are utilized to sense the information of the human target. The contact-free nature of wireless sensing makes it particularly appealing in challenging scenarios such as current COVID-19 pandemic. The key rational of wireless sensing is that the propagation of wireless signal in the air gets affected by target movements. By analyzing the variations of signal reflected from the target, rich target information can be obtained such as the movement speed and displacement. With the latest advance of wireless sensing, researches have successfully exploited WiFi [42], RFID [10], LoRa [41], acoustic [34] and 60 GHz [28] signals to accurately monitor the fine-grained respiration. However, one interesting observation is that while there are a lot of studies on respiration monitoring already, there is very little work [26] on heartbeat monitoring. After a thorough study on this, we find that the main reason is because heartbeat is much more difficult to be sensed in a contact-free manner compared to respiration. For human respiration, the chest displacement is around 5 mm. Taking acoustic signal as the example, this 5 mm displacement will cause a signal phase variation of larger than 100° when a 16 kHz signal is employed and this large phase variation can be easily detected. However, the chest displacement caused by heartbeat is merely 0.1 - 0.5 mm. Therefore, the induced phase variation is very tiny and much more difficult to be sensed. What makes it worse is that even a human target is stationary, the respiration-induced chest movement can easily submerge the displacement caused by heartbeat.

In this work, we target to monitor the important vital sign —heartbeat— with a smartspeaker for the first time. Heartbeat monitoring provides critical information regarding to the efficiency and the functionality of the cardiovascular system. For instance, heart rate variability is proven to be a factor closely related to Sudden Cardiac Deaths (SCDs). Cardiovascular disease (CVD) is also the number one cause of death globally, taking 17.9 million lives each year (31% of all deaths worldwide) [5]. Alone in the U.S., there are approximately 370,000 deaths from CVD annually [3]. We believe heartbeat monitoring is the important missing piece in the current wireless sensing-based vital sign monitoring research.

In this work, we propose to employ the commodity off-the-shelf smart speaker for heartbeat monitoring without requiring any dedicated sensors. Smart speakers such as Amazon Echo and Google Home are becoming more and more popular in home environment. It was estimated that more than 200 million smart speakers have been sold until the end of 2019 [8] and this number continues to increase every year. Samsung and Huawei are also following this trend to launch their smart speaker products. We believe smart speaker is an appropriate platform to host this home heartbeat monitoring application.

As shown in Figure 1(c), no matter the user sits or lies, the smart speaker can be utilized to accurately monitor the user's heartbeat rate (HR) and the more detailed interbeat interval (IBI) in a contact-free manner. The smart speaker transmits ultrasound signals and captures the signal reflected off the human chest. The reflected signals contain the chest motions caused by both human breath and heartbeat. With novel signal processing, we can



(a) ECG based approach (b) PPG based approach (c) Home audio based approach

Fig. 1. The evolution from traditional ECG and PPG based approach to our proposed smart speaker based approach.

extract not just the respiration but also the much finer heartbeat information, moving the state-of-the-art one step forward.

Though promising, it is non-trivial to utilize a commodity smart speaker to extract the very tiny heartbeat motion in the presence of interference. To achieve the objective, several challenges need to be addressed:

(a) *Random buffer delay.* During our experiments with the smart speaker, we observe a random time delay before the acoustic signal is transmitted out. This brings in a surprising result: during the signal processing process, the reflection signal sometimes shows up before the Line of Sight (LoS) direct path signal in the time domain. This is counterintuitive because the direct path signal travels the shortest distance and it should arrive first in the time domain. We deeply study this issue and discover the underlying reason for this interesting phenomenon. We then propose a novel scheme to construct a “virtual” transmission signal to help remove this random time delay to achieve accurate distance measurements.

(b) *Extraction of the tiny heartbeat motion.* The second challenge is to detect and extract the very tiny heartbeat motion (only about $0.1 - 0.5$ *mm*) in the presence of strong interference such as human respiration motion (about 5 *mm*) and ambient noise. Although the frequencies of respiration and heartbeat are quite different, because the heartbeat motion is much smaller and filters have leakages, removing the respiration motions in frequency domain does not work well and even a small leakage can greatly interfere with heartbeat sensing. We thus model the relationship between the signal changes and the superimposed respiration and heartbeat motions to understand the underlying mechanisms. To separate the superimposed respiration and heartbeat motions, we employ the *complete ensemble empirical mode decomposition* method by keeping adding white noise to the signal and perform an iterative extraction. The signal is then decomposed into multiple separated intrinsic modes, corresponding to waveforms of heartbeats, respiration and other noises accordingly.

(c) *A robust system which can work in real world.* The last challenge is to build a system which is robust against multiple real-world issues and work properly in different scenarios. We would like our system to work when the user is at different locations, facing different orientations and having different postures (lying or sitting). We target to achieve comparable or even better sensing accuracy than the wearable-based commodity products on the market.

The main contributions of this work are summarized as follows.

- To the best of our knowledge, we are the first to exploit commodity smart speaker to achieve accurate monitoring of heartbeats.
- We study and analyze the random system delay issue associated with commodity smart speaker. We show that this delay can cause the reflection path to be ahead of direct path, confusing the distance estimation. We propose to construct a “virtual” signal and design a delay-removal scheme to address this issue.

- We model the relationship between the signal variation and the heartbeat and respiration motions. We propose a series of signal processing methods to separate the signal variation caused by heartbeat from that caused by respiration to extract the heartbeat information.
- We are the first to build a robust real-time heartbeat monitoring system with smart speaker, and demonstrate that the proposed system can accurately monitor the subtle human heartbeats in the presence of noise and inference. Please find our demo on heartbeat monitoring at: https://youtu.be/b5So4tN6UEc.

2 RELATED WORK

Human Heartbeat Monitoring. Heartbeat monitoring has been extensively investigated in literature and can be categorized into contact-based and contact-free approaches. The contact-based solutions leverage wearable devices and sensors that are attached to human body for monitoring. Traditional clinical ECG (electrocardiogram) is a typical contact-based method [17]. An electrocardiograph is recorded by electric potential changes occurring between electrodes placed on a patient's torso to monitor the cardiac activity. In practice, the electrodes are attached to different positions on the body such as wrists and ankles. These methods require trained professionals to operate the devices, which prevent these systems from daily use at ordinary homes. The contact-based methods that can be used at home include finger-clamp pulse meter [24], smartphone [7] and smartwatch [4]/wristband [6]. These devices adopt PPG sensor (photoplethysmogram), which contains a light emitting diode (LED) and a photosensitive sensor. The LED emits green light on the skin. The photosensitive sensor then monitors changes in the arterial blood volume upon systolic/diastolic fluctuations of light wave and thus derives the heart rate. Besides, the built-in accelerometer inside the smartphone can also be used to monitor the heart rate by placing the phone on the chest and sense the heartbeat-induced tiny body movements [29] [32]. However, these methods still require the wearable/sensor to be in direct contact with the target and thus they are not suitable for long-term monitoring.

With recent advance of wireless sensing, contact-free approaches are emerged, such as vision-based [36] [13] and radio frequency (RF)- based [27] [9]. Taking the vision-based approach [13] as an example, the facial video is captured by the camera. The heartbeat causes the change of blood oxygen saturation, leading to face color changes. By measuring the face color changes from the video, the target's heartbeat information can be extracted. However, it is obvious that this method requires good lighting conditions, and also raises privacy concerns. On the other hand, RF-based approaches have been exploited, including adopting the large bandwidth FMCW radar [9] [12] or UWB radar [27] [18]. The radar based approaches require dedicated hardware and usually incur high hardware cost. On the other hand, commodity hardware such as WiFi and RFID-based approaches can be applied for respiration sensing but are still not able to monitor the extremely tiny heartbeat motions.

Acoustic-based sensing techniques. Acoustic signals have been widely employed for a large variety of applications, ranging from coarse-grained localization [31] [16], gait recognition [39], driver behavior monitoring [38], gesture sensing [37] to fine-grained respiration monitoring [34], finger drawing tracking [11] and lip-reading recognition [20]. Wang et al. [33] utilize the influence of the airflow changes caused by breathing on the sound wave to extract the respiration information. AcousticID [39] utilizes the doppler effect of various body parts on acoustic signals to recognize gait. Further, the gait features are used for user authentication. CAT [21] employs a mobile phone carried by a user to track the drone's relative location. The authors develop a distributed FMCW system to obtain the relative distance change, achieving a median tracking error of 4 mm. LLAP [37] tracks fine-grained hand moments by measuring the phase change of the signals from two microphones of a smartphone. ApneaApp [23] leverages FMCW chirp signal to capture human breath and infer sleep apnea event. While a lot of applications have been enabled by acoustic sensing, monitoring the tiny heartbeat is still a challenging task. The only work which enables acoustic-based heartbeat sensing is a recent work [26]. This work utilizes dual microphones on a smartphone to eliminate direct power leakage and obtain the reflection signal from user's chest.

The signal is then processed with an IIR comb notch filter [15] to extract the heart rate information. However, the achieved accuracy is not fine enough and the working range is limited. We re-implement the proposed system and find the sensing performance severely degrades when the sensing range is larger than 20 cm which means the device still needs to be placed very close to the target. In this work, for the first time, we employ the smart speaker to achieve accurate and stable heartbeat sensing. The sensing range can be up to 1.2 m and a prototype which can monitor the target's heartbeat in real time is demonstrated.

3 HEARTBEAT MONITORING USING ACOUSTIC SIGNAL

In this section, we first introduce the basics of FMCW signal and then present the practical issues associated with smart speaker-based FMCW signal processing.

3.1 Primer of FMCW Signal

Frequency-modulated continuous-wave (FMCW) signals come in the form of a sinusoid with time-varying frequencies. The signal frequency linearly increases over time and FMCW signal is also called chirp signal. As shown in Figure 2, the blue line denotes the transmitted signal with a predefined sweep time T. [1] The instantaneous frequency at time t can be expressed as $f(t) = f_0 + kt$, where f_0 is the starting frequency, $k = \frac{B}{T}$ is the slope of the frequency change and B is bandwidth. The instantaneous phase of the transmitted signal is denoted as $\phi(t) = 2\pi \int_0^t f(t)dt = 2\pi(f_0 t + \frac{kt^2}{2})$. Suppose the amplitude of the signal is A, the transmitted signal can be represented as:

$$x_{tx}(t) = Acos(\phi(t)) = Acos(2\pi(f_0 t + \frac{kt^2}{2})). \tag{1}$$

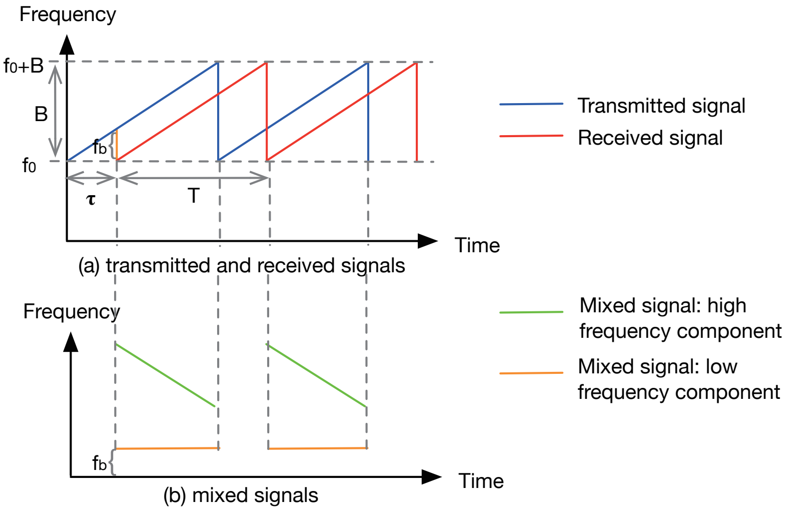

Frequency
f0+B
B
fb
f0
τ
T
Time
Transmitted signal
Received signal
(a) transmitted and received signals
Frequency
fb
Time
Mixed signal: high frequency component
Mixed signal: low frequency component
(b) mixed signals

Fig. 2. Signals in FMCW. (a) transmitted and received signals (b) mixed signal

FMCW signal is widely utilized to measure the distance from the target to the transceivers. When the signal transmitter and receiver are located at the same position and the target is located at a distance of R from the

[1] T is also called chirp time.

sensing devices, the signal arrives at the target and then gets reflected back from the target to the receiver after a time period of $\tau = \frac{2R}{c}$, where c is the signal propagation speed in the air. Therefore, there is a time delay of τ between the signal is transmitted out and the signal is reflected back. The reflected signal is thus represented as:

$$x_{rx}(t) = A'cos(\phi(t-\tau)) = A'cos(2\pi(f_0(t-\tau) + \frac{k(t-\tau)^2}{2})), \tag{2}$$

where A' is the received signal amplitude. After the reflected signal is received, the received signal is multiplied by the transmitted signal $x_{tx}(t)$ to perform the signal mixing operation. The mixed signal is $x_m(t) = x_{tx}(t) \cdot x_{rx}(t)$. By applying the product-to-sum conversion $cos\alpha \cdot cos\beta = \frac{cos(\alpha-\beta)+cos(\alpha+\beta)}{2}$, the mixed signal can be represented as:

$$x_m(t) = \frac{AA'}{2}\Bigg(\underbrace{cos\Big(2\pi(f_0\tau - \frac{k(\tau^2-2t\tau)}{2})\Big)}_{\textbf{Low frequency term}} + \underbrace{cos\Big(2\pi(f_0(2t-\tau) + \frac{k(2t^2-2t\tau+\tau^2)}{2})\Big)}_{\textbf{High frequency term}}\Bigg). \tag{3}$$

As shown in Figure 2, the mixed signal contains the low frequency component (i.e., $cos(\alpha-\beta)$) and the high frequency component (i.e., $cos(\alpha+\beta)$). The low frequency component has a constant frequency, corresponding to frequency differences of $x_{tx}(t)$ and $x_{rx}(t)$. The high frequency component contains the t^2 term and thus the frequency still changes with time. Here, the mixed signal is passed through a low pass filter to remove the high frequency component. We thus obtain $x_m(t) = \frac{AA'}{2}cos\Big(2\pi(f_0\tau - \frac{k(\tau^2-2t\tau)}{2})\Big)$ with only the low frequency part left. The frequency of the low frequency component is a constant $f_b = k\tau = \frac{2kR}{c}$. Thus, the distance R from the target to the transceiver can be calculated as:

$$R = \frac{c \cdot f_b}{2k}. \tag{4}$$

The resolution of the distance measurements depends on the frequency bandwidth of the signal. The larger the bandwidth, the finer the range resolution. If the frequency of the acoustic signal sweeps from 16 kHz to 21 kHz, the signal bandwidth is B=5 kHz. The speed of sound is 343 m/s. We can thus obtain the range resolution $\delta R = \frac{343}{2\times5000} = 0.0343\ m = 3.43\ cm$. This resolution is much larger than the displacement of heartbeat motion (i.e., 0.1 - 0.5 mm). This implies that we can not accurately measure the heartbeat-induced motion displacement using the absolute distance estimates.

3.2 Overview of the Proposed Heartbeat Sensing System

In this section, we present the overview of the proposed heartbeat monitoring system. Figure 3 illustrates the framework of the system. Our system mainly consists of two main modules: FMCW signal processing and heartbeat extraction. First, the speaker transmits FMCW chirp signals. These signals hit the target, get reflected back and received by the microphone. Following the basic FMCW signal processing strategy presented in Section 3.1, the reflection signal is mixed with the transmitted signal and then passed through a low-pass filter. However, in a real hardware system, the acoustic signal is first put in a buffer before it is sent out and there is a random time delay from the signal is triggered to be sent until the signal is actually sent out. This time delay is random so it can not be measured beforehand and removed. Theoretically, this random delay happens to both direct path and reflection path and thus the direct path signal will still arrive first before the longer reflection path signal. However, one interesting observation is that with the FMCW signal adopted, this random delay can cause the reflection path to appear in front of the direct path after the signal mixing and filtering operations.

To handle this random time delay, we carefully design a “virtual” transmission signal and mix it with the received signal to cancel out the system delay. Then we perform a fast Fourier transform (FFT) operation on the mixed signal to obtain the target location range bin. In this way, the signal which contains human heartbeat information is extracted. However, the heartbeat information and the respiration information, as well as the noise are mixed together. In heartbeat extraction module, we adopt the complete ensemble empirical mode decomposition method to extract heartbeat signal. After segmenting the heartbeat signal, the heartbeat rates and heartbeat intervals are estimated. We next describe each module of the system in detail.

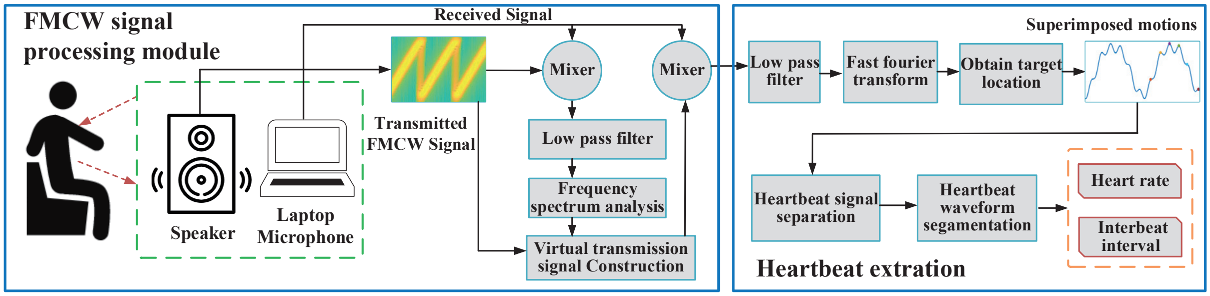


Fig. 3. System framework. The system consists of two modules: FMCW signal processing module and heartbeat extraction.

3.3 FMCW Signal Processing Module Design

In general, there exists a system delay when transmitting an audio signal using the smart speaker. This is because the audio signal is put into a buffer first and then transmitted out by the speaker. Therefore, there is a time delay between the signal transmission command and actual signal transmission. The amount of delay is random and thus can not be measured and removed easily. This issue is also discovered and reported in other works [25] [21].

To address this issue, existing approaches place the device at known reference positions [31] or move the device along a pre-defined trace to eliminate this time delay [40]. These solutions are effective but intrusive, requiring user interventions. Also these solutions only work when the delay is smaller than half the chirp period. If the delay is larger than half the chirp period, then after the signal mixing operation, the reflection path will appear in front of the direct path and fail existing approaches.

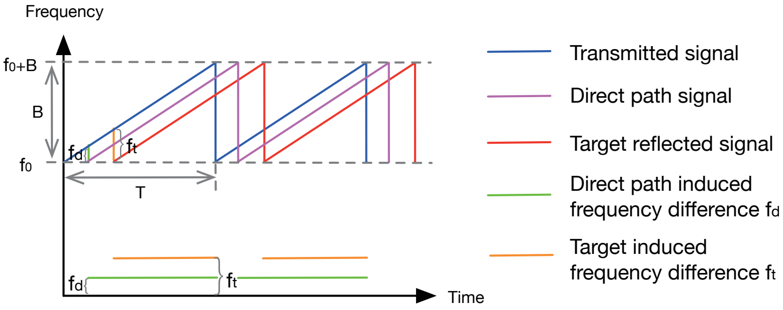
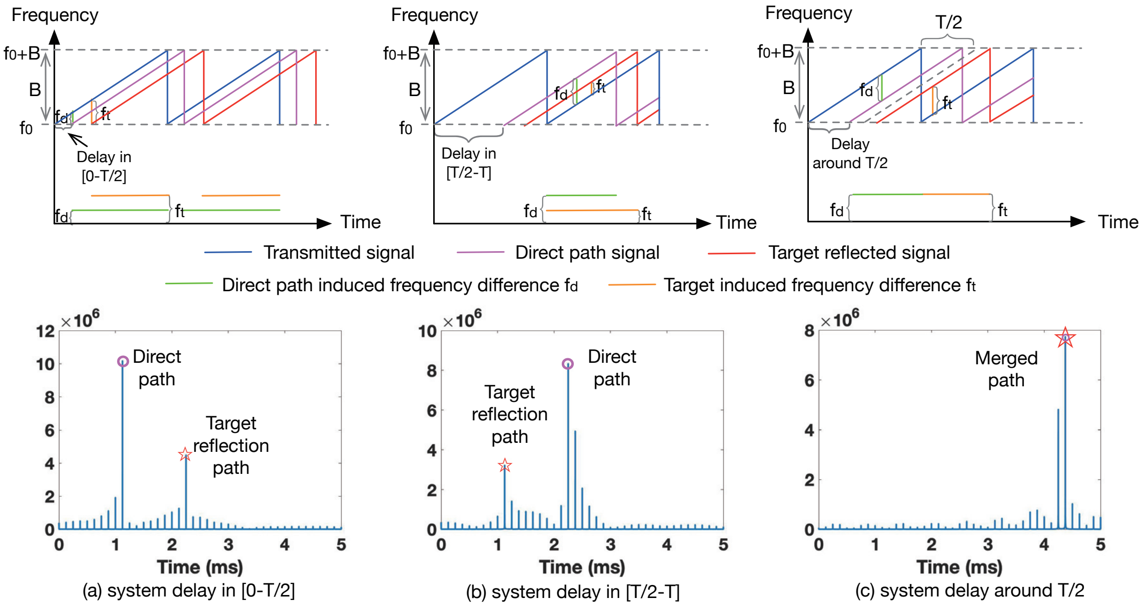


Fig. 4. Signals in FMCW.

Now we illustrate the direct path signal and target reflection signal in the time-frequency space. As shown in Figure 4, the blue line is the transmitted signal, the pink line is the received direct path signal and the red line is the received target reflection signal. The received signal contains both the direct path and reflection path



(a) system delay in [0-T/2]

(b) system delay in [T/2-T]

(c) system delay around T/2

Fig. 5. Three cases caused by different system delays.

signals. After the received signal is mixed with the transmitted signal and filtered to remove the high-frequency component, there are two frequency components left, corresponding to the direct path denoted as f_d and target reflection path denoted as f_t. In Figure 5(a), we can observe that the stronger direct path peak appears first before the target reflection peak as expected. However, when the time delay is larger than $T/2$, something interesting happens: the reflection path peak will appear in front of the direct path peak. In Figure 5, we illustrate the three different cases caused by different system delays.

- When the system delay is in the range [0 - T/2] as shown in Figure 5(a), we observe two peaks appear after the FFT operation. Note that for chirp signal, the frequency difference can be converted into equivalent time and distance differences. In Figure 5(a), the first peak corresponds to the direct path signal and the second one corresponds to the target reflection path signal. Although there is a system delay, the signal sequence is still correct: the direct path signal appears in front of the reflection path signal.
- When the system delay is in the range [T/2 - T], surprisingly, after the signal mix and filter operations, although there are still two peaks, the first peak now corresponds to the target reflection path signal while the second peak corresponds to the direct path signal as shown in Figure 5(b). The order of the direct path and reflection path in time domain is corrupted.
- When the system delay is $T/2$ as shown in Figure 5(c), f_d is obtained as the frequency difference between the direct path signal and first chirp of the transmitted signal. However, the f_t is obtained as the frequency difference between the target reflection path signal and the second chirp of the transmitted signal. In this scenario, f_d= f_t. The direct path signal peak and reflection path signal peak now coincide and only one merged peak can be observed.

Through the analysis above, we need to tackle two issues: 1) The random hardware-induced system delay introduces distance measurement errors; 2) Due to different amounts of time delays, the order of the direct path signal and reflection path signal in time can get corrupted. The traditional reference-based approaches do not work when the order relationship is corrupted. To address the above issues, we propose a two-phase signal mixing approach by introducing a “virtual” transmission signal, as shown in Algorithm 1. In *line 1*, we first mix the received signal with the transmitted signal and pass the mixed signal through a low-pass filter. Based on different amounts of system delays, we obtain one of the three cases described above. Since the direct path signal from the speaker to the microphone is much stronger[2] than the reflection path signal from the human chest, we can employ the signal strength to identify the direct path signal. As described in *line 2* to *line 4*, we perform the frequency analysis and select the strongest peak. The delay of this strongest peak is denoted as t_w. Now we create a “virtual” transmission signal which can be represented as:

$$x'_{tx}(t) = Acos(\phi(t + t_w)) = Acos(2\pi(f_0(t + t_w) + \frac{k(t + t_w)^2}{2})). \tag{5}$$

```
Algorithm 1: Construct "virtual" transmitted signal to obtain target location.
   Input: The transmitted signal $x_{tx}$,the received signal $x_{rx}$.
   Output: The target induced frequency $f_t$.
 1 Mix the received signal with transmitted signal and pass a low-pass filter, then perform FFT;
 2 Choose the highest amplitude peak of FFT and get $f_d$ which is frequency difference induced by direct path;
 3 $t_w = t' = \frac{f_d T}{2B}$;
 4 Shift the transmitted signal forward with time delay $t_w$ to construct "virtual" transmitted signal;
 5 Mix received signal using shifted “virtual” signal, and pass a low-pass filter then perform FFT;
 6 Select the highest amplitude peak of FFT and get $f_d$ which is frequency difference induced by direct path;
 7 if $f_d \neq 0$ then
 8     $t_w = T - t'$;
 9     return line 4
10 end
11 Select the second highest amplitude peak of FFT and get $f_t$ which is frequency difference induced by target;
12 return $f_t$
```

In Eq. 5, the virtual transmission signal is moved backward (to the left) by a time period of t_w, as dotted blue line shown in Figure 6. Then we utilize this “virtual” signal and the received signal to perform a second mixing operation again in *line 5*. The obtained mixed signal $x'_m(t)$ can be represented as:

$$x'_m(t) = x'_{tx}(t) \cdot x_{rx}(t). \tag{6}$$

Then the second mixed signal is also passed through a low-pass filter and performs the FFT operation. Now if the highest peak is located at timestamp “0" as shown in Figure 6(b), it indicates the random delay is removed properly. However, for the other case shown in Figure 7(a), even after the above operation, the direct path signal is still not located at timestamp “0". For this case, we move the original signal again backward (to the left) by a time period of $T - t'$. T is the chirp length which is known. In *lines 7-10*, we input $t' = \frac{f_d T}{2B}$ and return to *line 4*. Now the direct path peak is moved to timestamp “0" as shown in Figure 7(d). So with a maximum of two steps, the random time delay can be successfully removed and the direct-path-reflection-path order distortion issue is also addressed. Now, we are able to use the direct path signal as a reference to calculate the time difference between the target reflection signal and the direct path signal.

[2] The speaker and microphone are located very close to each other

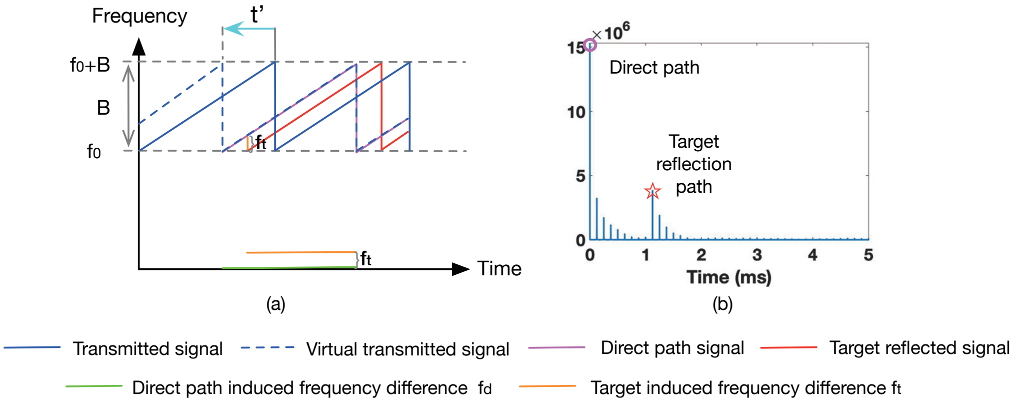


Fig. 6. “virtual" transmitted signal: the original transmitted signal is moved backward by a time period of t’.

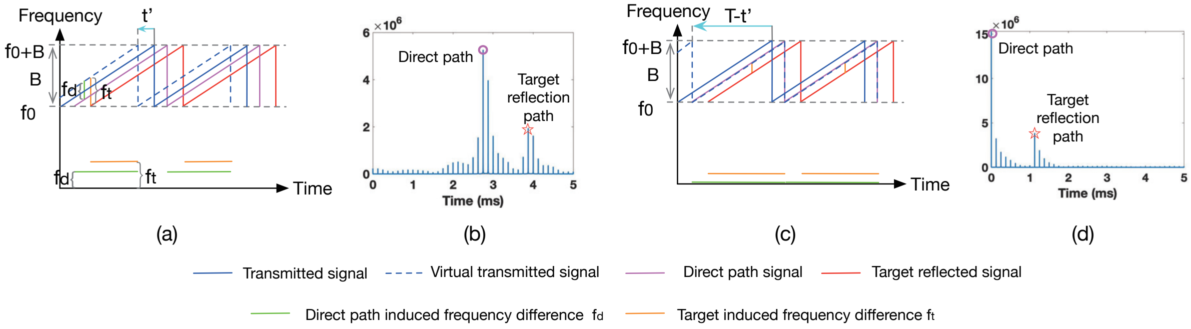


Fig. 7. The direct path peak is moved to the origin with two steps.

3.4 Understanding the Superimposed Respiration and Heartbeat Motions

In the previous section, we remove the random time delay and now we focus on the target reflection signal because it contains the heartbeat information. Suppose the subtle movement of human chest is Δd, the mixed signal $x'_m(t)$ is passed through a low-pass filter and then it can be represented as:

$$x'_m(t) = \frac{AA'}{2}cos\Big(2\pi(f_0\tau - \frac{k(\tau^2 - 2t\tau)}{2})\Big) \approx \frac{AA'}{2}cos(\frac{4\pi kRt}{c} + \frac{4\pi f_0\Delta d}{c}) = \frac{AA'}{2}cos(2\pi f_t t + \frac{4\pi f_0 \Delta d}{c}). \tag{7}$$

The above expression is decomposed into two terms. The first term $2\pi f_t t$ is an inter number of 2π phase change corresponding to the coarse-grained target distance. The second term $\frac{4\pi f_0 \Delta d}{c}$ is a phase change in the range of $0 - 2\pi$, corresponding to the fine-grained target distance. Here, a 2π phase change corresponds to a distance change of 10.7 mm. Note that the hearbeat-induced chest displacement is very small ($0.1 - 0.5\ mm$), thus, the subtle displacement will mostly only cause a phase change in the second term $\frac{4\pi f_0 \Delta d}{c}$. As shown in Figure 8, if the heartbeat-induced chest displacement is 0.5 mm, the phase change can be calculated as:

$$\frac{4\pi f_0 \Delta d}{c} = \frac{4\pi \times 16000 \times 0.0005}{343} = 0.093\pi = 16.8°. \tag{8}$$

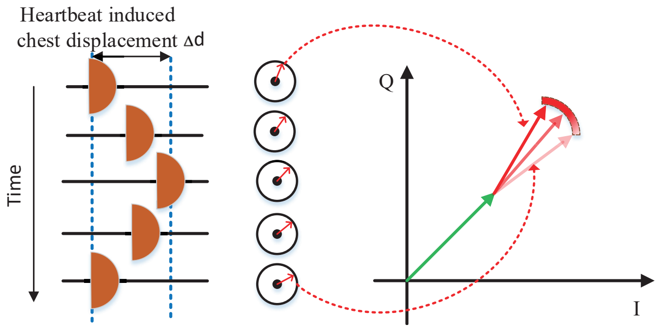
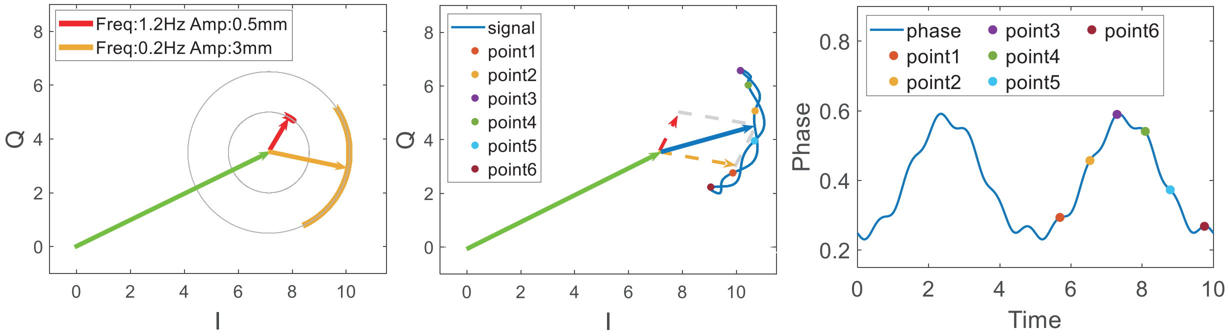


Fig. 8. Chest displacement due to heartbeat and corresponding induced phase change.

A chest displacement of $0.1 - 0.5\ mm$ will thus cause the signal phase change of 3.4° to 16.8°. Therefore, very accurate phase measurement is needed to monitor sub-millimeter heartbeat. However, one big challenge here is that the phase change of the signal caused by heartbeats is submerged by respiration-induced phase change. Respiration produces 10× larger chest displacement than that produced by the heartbeat. We model the two motions simultaneously in Figure 9a. We consider a respiration with a 3 mm displacement at 0.2 Hz frequency and a heartbeat with a 0.5 mm displacement at 1.2 Hz frequency. As shown in Figure 9b, we plot the signal variation caused by the two superimposed motions in the I-Q space. The phase change in Figure 9c shows the heartbeat information is buried in the respiration information. We can clearly observe the heartbeat caused signal variations (points 1 - 6) corresponding to the heartbeat frequency, which is 6 times larger than the respiration rate (frequency). We therefore need to separate the superimposed motions and obtain the breathing and heartbeat waveforms respectively.



(a) Signals of two separate motions (b) Signal of superimposed motions (c) Signal phase change

Fig. 9. The resultant signals of superimposed heartbeat and respiration motions.

3.5 Heartbeat Extraction

According to the analysis in the previous section, we observe that the heartbeat, respiration, and even small body movement can induce phase variation of a signal. We categorize different displacements of body movements and their frequencies in Table 1. The body movements are classified into four categories. The heartbeat and respiration frequency are [0.8 - 2 Hz] and [0.1 - 0.5 Hz], respectively. We can see that the slow human body movement has a low frequency of 0.1 - 2 Hz, which mainly refers to movement of the torso, such as leaning forward and backward. The fast body movement has high frequency in the range of 0.5 - 5 Hz, which mainly refers to the movement of certain parts of the body, such as hand or leg. We can observe that the signal frequencies caused by four components are overlapped, thus a simple band-pass filter cannot filter out the interference signal to retain the heartbeat information. To overcome this problem, we need to find a signal decomposition method to effectively isolate the signal of each frequency component.

Table 1. Summary of different displacements of body movements and their frequencies

Vital signs and body motion	Displacements	Frequency
Heart Rate	0.1-0.5mm	0.8-2Hz
Breathing Rate	1-5mm	0.1-0.5Hz
Slow body movement (torso)	Decimeter level	0.1-2Hz
Fast body movement (hand/leg shake)	Centimeter level	0.5-5Hz

Empirical mode decomposition (EMD) [14] is a signal time-frequency analysis technique that can decompose the signal into a superposition of independent frequency components. These frequency components are called intrinsic mode function (IMF) when they satisfy the following two requirements: 1) the number of extrema and the number of zero-crossings must either be equal or differ at most by one; 2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The procedure of extracting an IMF is called sifting. The sifting process is as follows.

Step 1: For a signal $x(t)$ containing superimposed motions, the local maximum and local minimum are first identified;

Step 2: Connect all the local maxima by a cubic spline as the upper envelope ($u(t)$), and then obtain the lower envelope $l(t)$ in the same way. Thus the average of the upper and lower envelopes can be calculated as $m(t) = \frac{u(t)-l(t)}{2}$.

Step 3: Let $h(t) = x(t) - m(t)$ and check whether $h(t)$ satisfies the conditions of IMF. If not, continue the above iteration process until an $h(t)$ satisfying the conditions of IMF is obtained.

In this way, the signal $x(t)$ is decomposed into a series of IMFs by applying the EMD method. Since the decomposition is based on the local characteristic-scale of the signal, it can be applied to nonlinear and nonstationary processes [14]. However, when the extremum distribution is not uniform, the mode mixing occurs and leads to similar frequency components in different IMF components or multiple different frequency components in a single IMF. Further, ensemble EMD (EEMD) [35] is proposed to improve the extremum characteristics by adding Gaussian white noise to the signal to achieve more uniform extremum distribution. However, the instability and incompleteness of the decomposed results appear due to the addition of white noise. Thus we employ complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [30], which adds an adaptive white Gaussian noise to further mitigate the mode mixing in the EMD method, and reduce the amount of similarity between decomposed IMF components to achieve a more complete separation.

Suppose $x(t)$ denotes the phase change of the received acoustic signal in Section 3.4, which contains the IMFs corresponding to different frequency components (heartbeat, breathing and other components). Let $\widetilde{C_i}$ be the i-th IMF mode component obtained by the CEEMDAN decomposition, $n_i(t)$ be the Gaussian white noise added at the i-th time and β_i be the signal-to-noise ratio coefficient. $E_k(\cdot)$ is the operator which produces the kth mode in the EMD process. Thus, with the aid of CEEMDAN, the heartbeat signal can be obtained with the following steps:

Step 1: We add white Gaussian noise to the signal $x(t)$ and repeat the process I times to construct the sequence $X_i(t) = x(t) + \beta_0 n_i(t)$, $i = 1, 2, \ldots, I$. For each $X_i(t)$, we employ the EMD algorithm to decompose the signal until obtaining the first IMF mode component. The first mode component of the CEEMDAN method is calculated as:

$$\widetilde{C_1} = \frac{1}{I} \sum_{i=1}^{I} E_1(X_i(t)). \tag{9}$$

Step 2: After we obtain the first IMF mode component ($k = 1$), we remove the first IMF mode component from the signal $r_1 = x(t) - \widetilde{C_1}$.

Step 3: We then continue to calculate the second IMF mode component by adding the white noise to the remaining signal: $r_1 + \beta_1 E_2(n_i(t))$, $i = 1, 2, ..., I$. The second IMF of $x(t)$ signal is calculated as:

$$\widetilde{C_2} = \frac{1}{I} \sum_{i=1}^{I} E_1(r_1 + \beta_1 E_1(n_i(t))), \tag{10}$$

Step 4: We continue the process. By removing the kth IMF mode component from the signal $r_k = r_{(k-1)} - \widetilde{C_k}$, the (k+1)-th IMF of $x(t)$ signal is calculated as:

$$\widetilde{C_{(k+1)}} = \frac{1}{I} \sum_{i=1}^{I} E_1(r_k + \beta_k E_k(n_i(t))), \tag{11}$$

Step 5: We iterate the process until the signal cannot be further decomposed. The coefficients $\beta_k = \varepsilon_k std(r_k)$ allow tuning the SNR at each iteration.

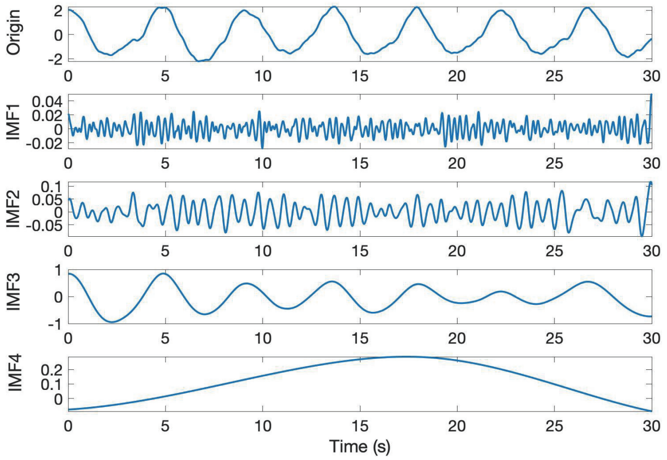


Fig. 10. The decomposed phase change signals.

Based on the above CEEMDAN decomposition method, we decompose a target reflection signal into four IMFs as shown in Figure 10. We plot the four IMFs in descending order of the frequency. IMF1 is the high-frequency noise and IMF2 is the heartbeat frequency component. IMF3 is the respiration frequency component and IMF4 is the low-frequency noise. By performing fast Fourier transform (FFT) on each component, it can be observed that the heartbeat IMF2 component is within 0.8 - 2 Hz and the IMF3 respiration component is within 0.16 - 0.6 Hz. In this way, we successfully extract the heartbeat signal of the human target.

To obtain heartbeat intervals, we then employ a dynamic heartbeat segmentation scheme by applying the EM algorithm [43]. Our method dynamically adjusts the duration of a heartbeat interval on a continuous heartbeat signal, and uses the dynamic programming method to match and iteratively optimize the heartbeat segmentation. As shown in Figure 11, we plot the segmentation results obtained. It can be observed that the segmented heartbeat intervals are consistent with the ground-truth ECG signals with only a small 0.05 s average deviation.

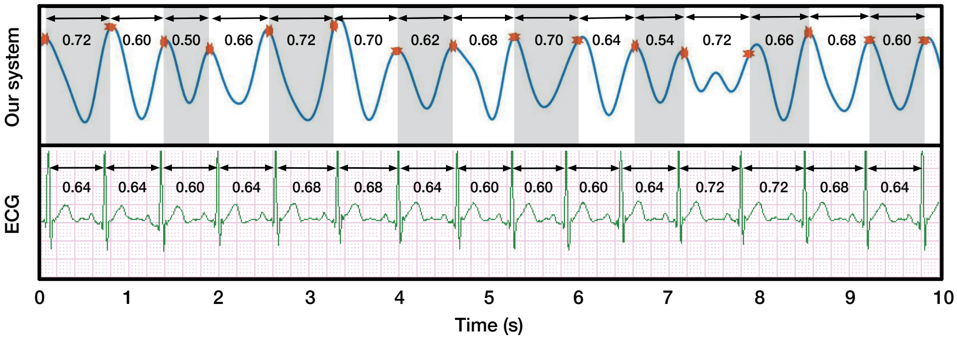


Fig. 11. Heartbeat segmentation result compared with ECG. The interbeat interval can be accurately derived with our methods.

4 EVALUATION

In this section, we implement the heartbeat monitoring system on top of a smart speaker. We conduct comprehensive experiments and report the results.

4.1 Experiment Setup

We implement our system using an off-the-shelf smart speaker (JBL Jembe, 6 Watt, 80 dB) and connect it to a laptop (MacBook Pro 2.6GHz with an Intel Core i7, 16 GB RAM) via the 3.5mm Audio Interface (AUX) as shown in Figure 12. The smart speaker is employed to transmit acoustic signals and the laptop with a built-in microphone is used to received the signal. The transceivers are place $60 - 120$ cm away from the user. We adopt f_c = 16 kHz, B = 5 kHz, T = 0.02 s to generate acoustic FMCW signals. The laptop employs a 48 kHz sampling rate and processes the signals in real time to monitor the target heartbeat. We employ a 3-lead ECG monitor, i.e., Heal Force PC-80B as shown in Figure 13 to measure the ground-truths. The electrocardiogram can be obtained to calculate heartbeat rate (HR) and interbeat interval (IBI). We define two metrics to evaluate the performance of the system:

- Heartbeat rate estimation error: this error is defined as the absolute value of the difference between the estimated heartbeat rate and the ground-truth rate. The unit is beats per minute (bpm).
- IBI estimation error: this error is defined as the absolute time difference between the estimated heartbeat interval and the ground truth, which is an important indicator to measure the accuracy of the boundaries of each heartbeat. The unit is millisecond.

We develop a web-based user interface to show the heartbeat signal variations in real time, as shown in Figure 14. The demo video can be found at https://youtu.be/b5So4tN6UEc.

4.2 Overall Performance

We first evaluate the overall performance of our system in terms of heartbeat rate error and heartbeat interval error. We compare the proposed system with two contact-based commercial solutions, i.e., Kiwi App [1] and Kangyuan App [2], and the state-of-the-art contact-free smartphone-based solution, i.e., ACG [26]. Kiwi requires

Fig. 12. Our device.

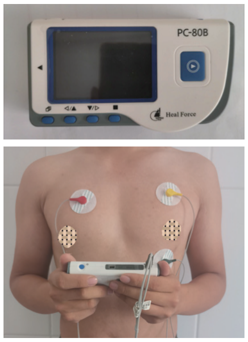


Fig. 13. ECG device.

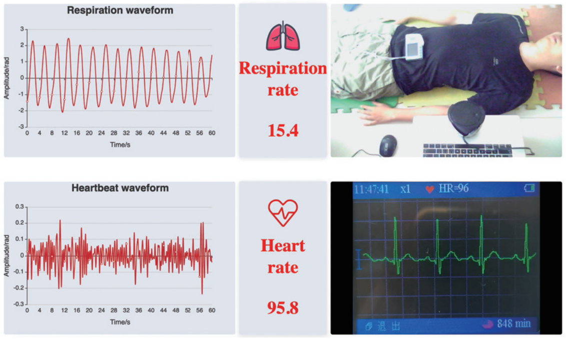
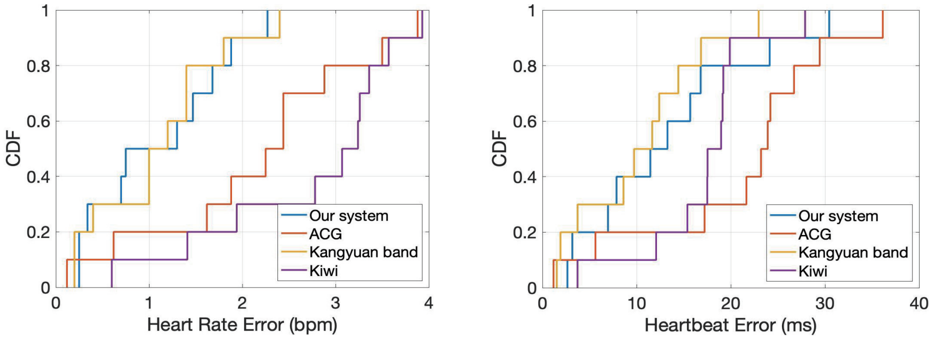


Fig. 14. Heart rate monitoring system.

the user to place a finger on the phone camera, which shines a light into the fingertip and measures the light absorbed by oxygenated haemoglobin. Kangyuan App requires the user to wear the Kangyuan band on the wrist and the band emits green light on the skin to monitor the heartbeat rate. We implement ACG as an APP on Huawei Nova 5 Pro with Android 10 OS. We find that the smartphone based acoustic sensing system (ACG) operates in a range below 30 cm. Thus, in our experiments, we place the smartphone at a distance of 20 cm from the user. In this experiment, five participants are involved. We monitor each participant for a period of 5 mins continuously and repeat the monitoring process twice.

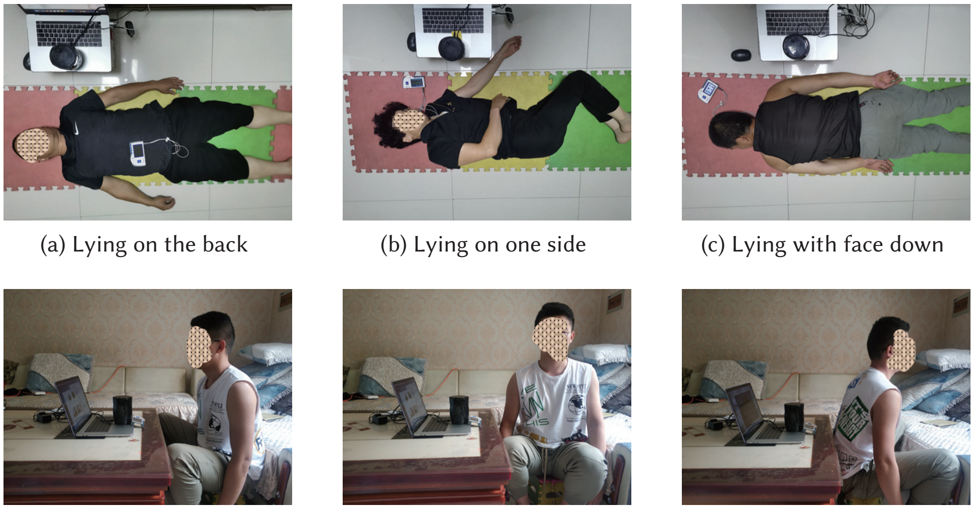
Figure 15a plots the cumulative distribution function (CDF) of the heartbeat rate measurement errors of the four approaches. The achieved median errors of our system, Kangyuan band, ACG and Kiwi are 0.75 bpm, 1.01 bpm, 3.07 bpm and 3.24 bpm, respectively. Our system achieves slightly better performance than Kangyuan band and much better performance than ACG and Kiwi. Comparing the performance of the two commercial devices, Kangyuan achieves a much better performance than Kiwi. The reason might be that the smartband (Kangyuan) is tightly fastened on the wrist. In contrast, Kiwi requires the user to touch a finger at the camera and the contact between the finger and camera is not that stable during the process. Another possible reason is that the white flash light from the camera does not perform as well as as the single-color green light. The error of ACG comes from two aspects. First, in certain cases, the direct path is not completely eliminated due to the strict requirement



(a) Estimation of Heart rate

(b) Estimation of Heartbeat interval

Fig. 15. Overall performance of our system.



(a) Lying on the back

(b) Lying on one side

(c) Lying with face down

(d) Sitting facing the device

(e) Sitting sideways

(f) Sitting with back facing the device

Fig. 16. Experiment environment and setup.

of signal alignment which is hard to achieve in reality. Second, ACG employs filters to separate signals and even a small leakage from the respiration can still severely interfere with the monitoring of subtle heartbeat.

The CDF of heartbeat interval error in Figure 15b demonstrates similar evaluation results. Specifically, our system achieves a median error of 13.28 ms (1.73%), which is better than ACG (a median error of 3.1%) , Kiwi (a median error of 3%) and slightly lower than smart band (a median error of 1.57%). Note that IBI can be used for heartbeat rate variability analysis and the achieved accuracy is high enough to detect most heart diseases.

4.3 The Impact of Subject Diversity

We recruit eight participants including three females and five males in the age range of 10 - 65 to evaluate the effect of subject diversity. The participants are asked to behave naturally in sitting and lying postures, as shown in Figure 16a and Figure 16d. For each posture, we record five groups of 5-min monitoring data for each participant.

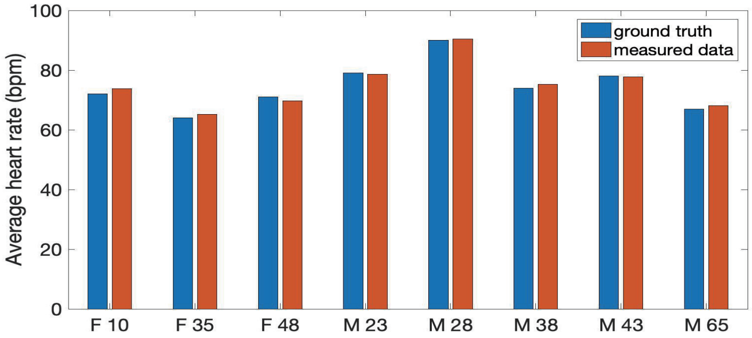


Fig. 17. Heart rates of different participants. (F-Female, M-Male, the following number is the age of this subject)

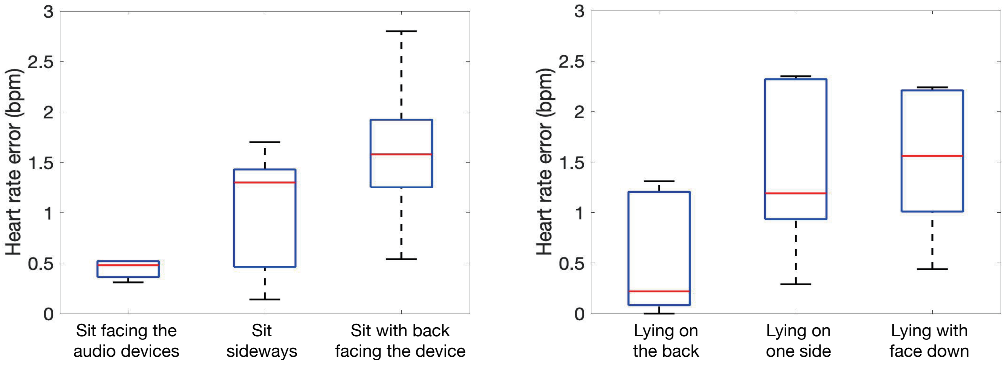
During the measurement process, the 3-lead ECG monitor is attached to the chest of the participants to record the ground truths.

Figure 17 shows the obtained heartbeat rates of eight participants and the ground truths. For the child (10 years old) and elderly (65 years old), we achieve a median error of 1.8 bpm and 1.1 bpm, respectively. We observe a slightly lower accuracy for child and we believe this is because children have weaker chest motions and smaller body size. For young adults, we always achieve a median error below 1 bpm. Compared with the EEG solution which needs to attach sensors to the body, all participants agree the proposed contact-free system is more convenient for home use. Note that the device placed on the human target in Fig. 16 is for ground-truth measurements and is not part of our system.

4.4 The Impact of Different Sitting and Lying Postures

In order to evaluate the performance under different postures, we ask the participants to sit or lie with different postures. As shown in Figure 16, for the sitting scenario, the participants sit facing the audio devices, sit sideways and sit with back facing the device. For the lying scenario, the participants also have three typical postures, i.e., lying on the back, lying on one side, lying with face down. Figure 18a and Figure 18b show the heartbeat estimation error when the subject is sitting and lying with different postures, respectively. The results show that when the participant sits facing the device, the median error of heartbeat monitoring is the smallest (0.48 bpm). When the participant sits sideways, the heart rate monitoring error increases to 1.3 bpm. This is because when the user changes the orientation, the body reflection surface goes from the chest to the side of his/her body. The effective signal reflection surface and motion displacement are both reduced. Therefore, the error slightly increases. Besides, even the participant sits with his/her back facing the device, the heart rate monitoring still works well and the median error remains low at 1.58 bpm.

In addition, for three lying postures, the media errors of our system are 0.22 bpm, 1.19 bpm and 1.56 bpm, respectively. Compared with sitting scenario, the accuracy of heartbeat rate monitoring in lying postures is higher. This is because when the participant is in a sitting posture, involuntary movement of human body interferers the heartbeat monitoring. On the other hand, the body of the participant in a lying posture is more stable and thus it is easier for us to extract the heartbeat waveform. The experiment results indicate that our system is able to accurately monitor human heartbeat for all the common sitting/lying postures in real world environment.

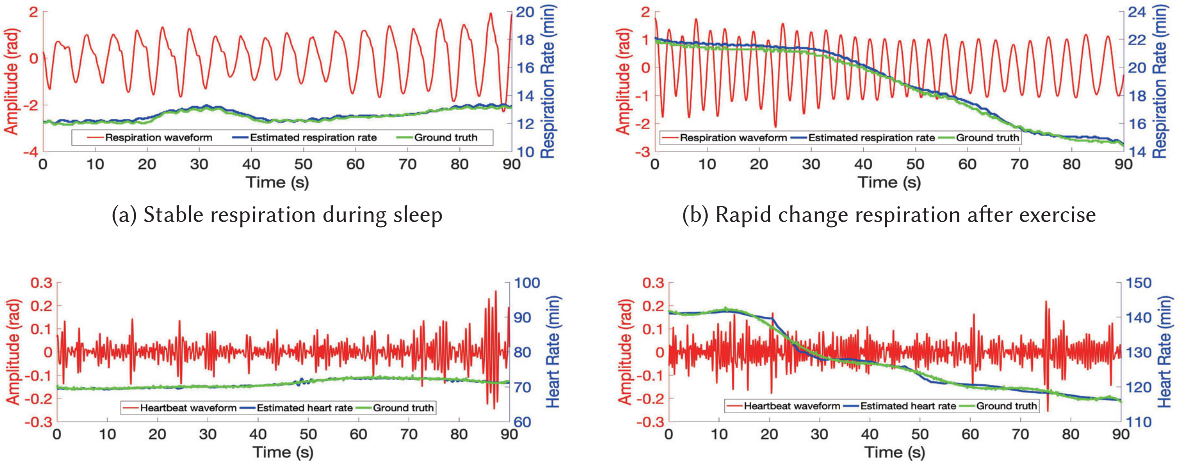


(a) Sitting scenario (b) Lying scenario

Fig. 18. Impact of different sitting and lying postures.

4.5 The Impact of Heart Rate Change

Note that when we sleep at night, the heartbeat rate is relative slow and stable. During the process of exercise, the heartbeat rate increases rapidly. We expect our system to be able to capture not just the stable information but also the respiration and heartbeat changes. Figure 19a and Figure 19c show that a participant in sleeping state has stable respiration and heartbeat. The corresponding respiration rate and heartbeat rate are 12.6 bpm and 72 bpm. Then the participant is asked to do high-intensity exercises (i.e., running), which increases the heartbeat rate significantly. Then, we let the participant sits in the chair and monitor his respiration rate and heart rate. Figure 19b and Figure 19d show the monitoring results of the participant after high-intensity exercise. The red line shows the signal variation and the blue line indicates the respiration and heartbeat rates. It can be seen that the heartbeat rate of the participant changes from 141 bpm to about 116 bpm. This is because the heartbeat rate gradually falls down to a normal level after exercise. During the whole process, the proposed system is able to accurately track the fine-grained heartbeat rate change in real time.



(a) Stable respiration during sleep

(b) Rapid change respiration after exercise

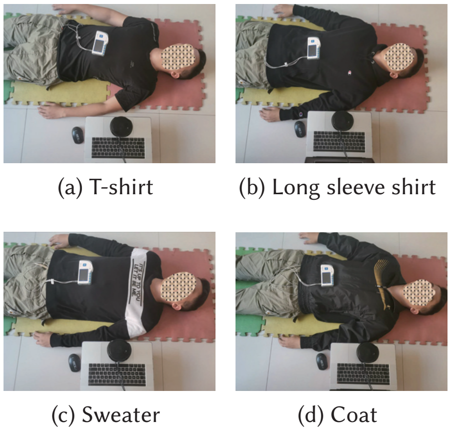
(c) Stable heartbeat during sleep

(d) Rapid change heartbeat after exercise

Fig. 19. Impact of heart rate change.

4.6 The Impact of Different Clothing

To evaluate the impact of clothing, the participants are asked to wear different clothes and observe their heartbeat monitoring accuracy. As shown in Figure 20, the performance of our system is examined under four different clothing, including (1) a lightweight T-shirt; (2) a long sleeve shirt; (3) a sweater; (4) a thick coat. The results of heartbeat rate estimation error are shown in Figure 21. We find that for all different clothing, our system achieves a median estimation error below 2.1 bpm for heartbeat rate monitoring. Moreover, our system performs better when the participant wears less. When the participant wears a lightweight T-shirt, the median error is just 0.22 bpm. In contrast, the median errors of wearing a long sleeve shirt, a sweater and a coat achieve 1.17 bpm, 0.82 bpm and 2.08 bpm, respectively. We believe there are two possible reasons for this performance degradation. The first reason is that a thicker clothing attenuates the signal more and therefore a weaker signal is received. The second reason is that, when the participant wears a thin clothing, the clothing moves together closely with the chest. However, for a thicker clothing, it does not move as much and therefore a smaller movement. In our experiment, even if the participant wears a thick coat on top of a sweater, the achieved accuracy is still relatively high.


(a) T-shirt (b) Long sleeve shirt

(c) Sweater (d) Coat

Fig. 20. Different clothing scenarios.

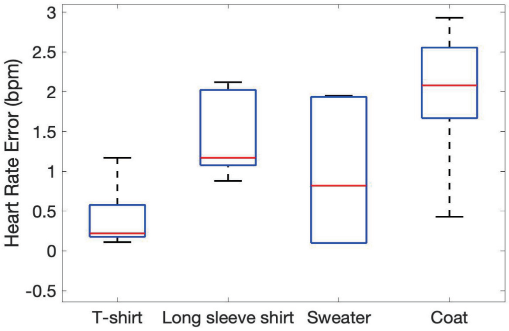
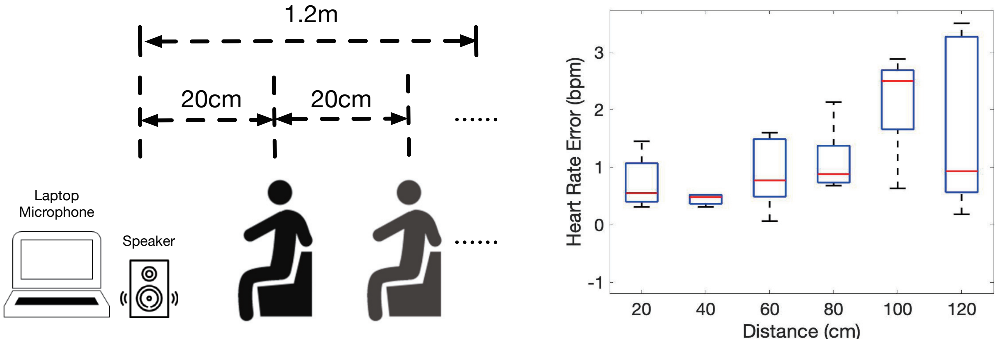


Fig. 21. Impact of different clothing.

4.7 The Impact of User-device Distance

Now we evaluate the performance of our system when users are located at various distances away from the audio devices. We vary the distance between the device and the participant from 0.2 m to 1.2 m at a step size of 0.2 m, as shown in Figure 22a. At each position, we monitor the heartbeat for five minutes. Figure 22b shows the average heartbeat rate estimation errors under different distances. It can be seen that the proposed system achieves an error of 0.55 bpm, 0.48 bpm, 0.77 bpm, 0.88 bpm at 20 cm, 40 cm, 60 cm and 80 cm, respectively. When we further increase the distance to 1 m and 1.2 m, we observe a slightly larger error increase due to weak reflection signals. However, we want to emphasize that even at 1.2 m, the achieved accuracy is still high enough to meet the requirement of most applications. Also this achieved distance (1.2 m) significantly outperforms the state of the art (30 cm) achieved in acoustic-based heartbeat sensing.

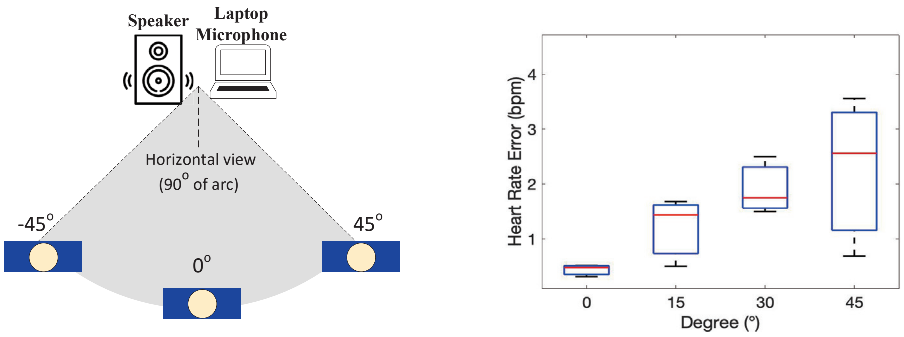


(a) Different distance scenarios (b) Heart rate

Fig. 22. Impact of user-device distance.

4.8 The Impact of User-device Direction

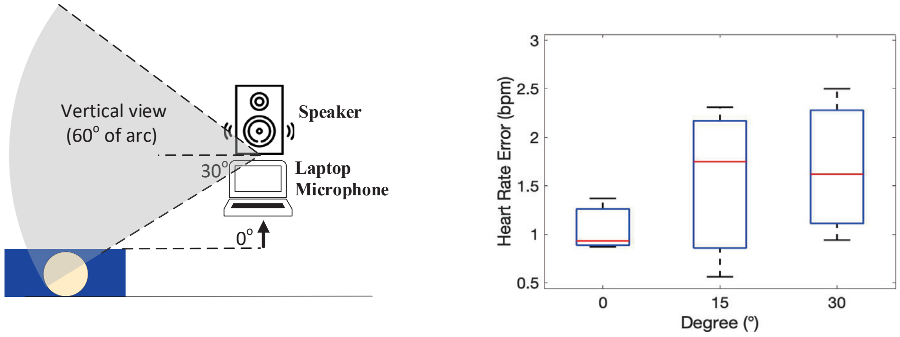
To evaluate the impact of user-device direction, we fix the location of the device and ask the participant to sit at different locations at an angle of 0°, 15°, 30° and 45° with respect to the device as shown in Figure 23a. The distance between the participant and the device is set as 60cm. As shown in Figure 23b, our system achieves the highest accuracy when the participant is located at 0° with respect to the device. The estimation error increases from 0.48 bpm to 2.48 bpm when the angle is 45°. When the angle is larger than 45°, it is difficult to receive the reflection signal and sense the heartbeats. Thus, we conclude that the horizontal angle coverage of our acoustic sensing system is around 90°. On the other hand, for the lying scenario in Figure 24a, we place the audio device on one side of the human body, and vary the height of the device. The initial height of the device is the same as the human chest (30 cm above the ground) and increased to 45cm. The angle of the participant with respect to the device changes from 0° to 30° in the vertical plane. During the process, the heartbeat rate estimation error increases from 0.48 bpm to1.62 bpm. When we further increase the height, the signal cannot reach the participant's chest any more. We therefore conclude that the vertical angle coverage of our sensing system is around 60°.



(a) Sitting at different horizontal angles

(b) Horizontal angle vs. estimation error

Fig. 23. Impact of user-device sitting direction.



(a) Lying at different vertical angles

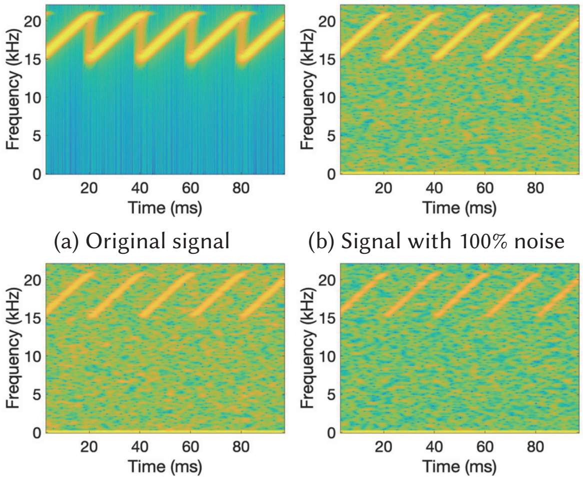
(b) Vertical angle vs. estimation error

Fig. 24. Impact of user-device lying direction.

4.9 The Impact of Noise Interference

To evaluate the impact of ambient noise, we evaluate the system performance under different types of ambient noises including human laughter, music and the Gaussian white noise. We observe that the frequency bands of human laughter and music are typically in the range of 300 Hz – 4 kHz, which is far away from the signal frequency (16 kHz – 21 kHz) used in the proposed system. Therefore, by applying a simple high-pass filter, these noises can be mostly removed and have little effect on the sensing performance. In our experiments, when we set the signal strength of the human laughter[3] and music as 200% of that used for heartbeat sensing, the achieved median errors are 1.93 bpm and 1.07 bpm, respectively. Compared with the result obtained when there is no noise (0.75 bpm), the accuracy degradation is very small. From these results, we can conclude that the human laughter and music have little effect on the performance of the proposed system

We further add different levels of Gaussian white noise to the transmitted signal to see the effect. Note that white noise spans across the whole frequency band. We increase the strength of the white noise from 0% to 300% of that of the transmitted signal at a step size of 100%. We show the frequency spectrum of the transmitted signal with different levels of white noise in Figure 25. It can be seen that when there is no noise, the signal chirp is clear. With 300% white noise added, the transmitted signal becomes obscured, but can still be identified owning to the chirp design adopted. Figure 26 shows the heartbeat estimation error under different levels of white noise. A slight increase can be observed with a larger white noise. However, even with a very large 300% noise, the median error is still as low as 2.5 bpm. The main reason of this strong anti-noise capability is the FMCW chirp design spanning across a relatively large bandwidth (5 kHz in our design). Single-frequency and OFDM signals will be more easily affected by ambient noise.



(a) Original signal

(b) Signal with 100% noise

(c) Signal with 200% noise

(d) Signal with 300% noise

Fig. 25. Transmitted signal with different noise levels.

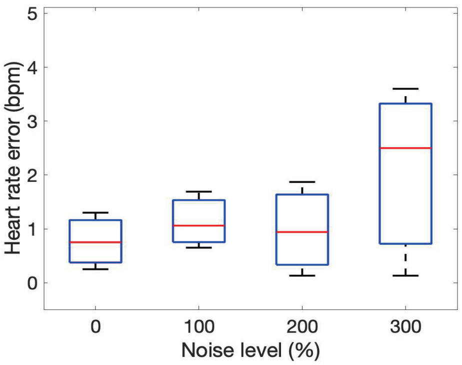


Fig. 26. Noise level vs. estimation error

5 LIMITATIONS AND DISCUSSION

In this work, we focus on utilizing the smart speaker to achieve contact-free human heartbeat monitoring. We briefly discuss the limitations and also some promising future research directions below.

[3] Note that to precisely tune the signal strength of the laughter, we record down the laughter and play it through a speaker

- **Multi-target Heartbeat Sensing.** In this paper, we focus on heartbeat sensing of a single target. When multiple targets exist, reflection signals from multiple targets get mixed at the receiver and it is challenging to separate these signals and sense each individual target. For our current design, if multiple targets are far away from each other, our system can still work. However, the performance degrades when the distance between two targets are less than 2 m. In our future work, we plan to exploit the unique opportunity of multiple microphones available at the smart speaker to separate signals in spatial domain for multi-target sensing.
- **Self interference** The proposed system works well when the target is static (e.g., during sleeping). When the target is walking or performing continuous activities such as typing, the proposed system still has difficulties to accurately sense the subtle heartbeat. This is because compared to tiny heartbeat motion, even typing induces much larger signal phase changes, interfering with heartbeat sensing. The heartbeat-induced signal variations can be easily submerged without being detected. Note that the proposed signal decomposition method works well in separating the heartbeat information from the respiration information because the respiration process is periodic and the chest movement is still small (i.e., 0.5 cm).
- **Practical usage.** With the popularity of microphones and speakers embedded in home appliances, we believe the proposed system is a promising solution for contact-free vital sign monitoring at home. We want to emphasize that acoustic sensing can be used for not just heartbeat monitoring but also other applications such as fall detection, gesture recognition and respiration sensing. We envision that the smart speaker and other microphone/speaker equipped devices have a great potential to form an ecosystem of acoustic sensing, providing ubiquitous sensing service at home.

6 CONCLUSION

In this paper, we enable smart speaker to monitor subtle human heartbeats in a contact-free manner for the first time. We analyze the effect of system delay on commodity smart speaker, and propose a novel approach to address this delay. We also model the relationship between the signal changes and the superimposed respiration and heartbeat motions. Through a deep understanding of the underlying mechanisms, we employ a series of novel signal separation methods to extract the subtle heartbeat motion in the presence of strong interference from respiration. We build a prototype system to demonstrate the sensing performance in real-life environments. Comprehensive experiment results show the effectiveness of our system in achieving comparable performance as the commodity wearable devices. We believe the smart speaker is a powerful platform capable of realizing a large range of wireless sensing based applications in our everyday lives.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 61802373, No. 62072450), the Youth Innovation Promotion Association, Chinese Academy of Sciences (No. 2020109), the EU CHIST-ERA RadioSense Project, the EU Horizon 2020 research and innovation programme IDEA-FAST (No. 853981), the Project 2019BD005 supported by PKU-Baidu Fund.

REFERENCES

[1] 2014. Kiwi application. http://patrickhealthappreview.blogspot.com/2014/06/app-review-kiwi-instant-heart-rate.html.
[2] 2018. Kangyuan bracelet. https://www.kangyuanai.com/index/index/product.html.
[3] 2019. Heart Disease and Stroke Statistics-2019 At-a-Glance. https://healthmetrics.heart.org/wp-content/uploads/2019/02/At-A-Glance-Heart-Disease-and-Stroke-Statistics-%E2%80%93-2019.pdf.
[4] 2020. Apple watch. https://www.apple.com/watch/.
[5] 2020. Cardiovascular Diseases. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1.
[6] 2020. Fitbit wrist band. https://www.fitbit.com/.
[7] 2020. Heart Rate Monitor. https://play.google.com/store/apps/details?id=com.repsi.heartrate&hl=en.

[8] 2020. smart speaker market. https://marketingland.com/more-than-200-million-smart-speakers-have-been-sold-why-arent-they-a-marketing-channel-276012.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2015. Smart Homes that Monitor Breathing and Heart Rate. *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 837–846.

[10] Lili Chen, Jie Xiong, Xiaojiang Chen, Sunghoon Ivan Lee, Daqing Zhang, Tao Yan, and Dingyi Fang. 2019. LungTrack: Towards Contactless and Zero Dead-Zone Respiration Monitoring with Commodity RFIDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 79 (2019), 22 pages.

[11] Mingshi Chen, Panlong Yang, Jie Xiong, Maotian Zhang, Youngki Lee, Chaocan Xiang, and Chang Tian. 2019. Your Table Can Be an Input Panel: Acoustic-Based Device-Free Interaction Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1 (2019), 21.

[12] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. 2017. Zero-Effort In-Home Sleep and Insomnia Monitoring Using Radio Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3 (2017), 18.

[13] J. Hu, Y. He, J. Liu, M. He, and W. Wang. 2019. Illumination Robust Heart-rate Extraction from Single-wavelength Infrared Camera Using Spatial-channel Expansion. *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3896–3899.

[14] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454, 1971 (1998), 903–995.

[15] Young K. Jang and Joe F. Chicharo. 1993. Adaptive IIR comb filter for harmonic signal cancellation. *International Journal of Electronics* 75, 2 (1993), 241–250.

[16] R. Jia, M. Jin, Z. Chen, and C. J. Spanos. 2015. SoundLoc: Accurate room-level indoor localization using acoustic signatures. *IEEE International Conference on Automation Science and Engineering (CASE)*, 186–193.

[17] Hampton John and Joanna Hampton. 2019. *The ECG Made Easy E-Book.* Elsevier Health Sciences.

[18] Seong-Hoon Kim, Zong Woo Geem, and Gi-Tae Han. 2019. A Novel Human Respiration Pattern Recognition Using Signals of Ultra-Wideband Radar Sensor. *Sensors (Basel, Switzerland)* 19, 15 (2019).

[19] Leonard S Lilly. 2012. *Pathophysiology of heart disease: a collaborative project of medical students and faculty.* Lippincott Williams & Wilkins.

[20] L. Lu, J. Yu, Y. Chen, H. Liu, Y. Zhu, Y. Liu, and M. Li. 2018. LipPass: Lip Reading-based User Authentication on Smartphones Leveraging Acoustic Signals. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. 1466–1474.

[21] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-Precision Acoustic Motion Tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*. 69–81.

[22] Marcel Miynczak and Gerard Cybulski. 2016. Improvement of Body Posture Changes Detection During Ambulatory Respiratory Measurements Using Impedance Pneumography Signals. *Mediterranean Conference on Medical and Biological Engineering and Computing* (2016).

[23] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless Sleep Apnea Detection on Smartphones. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys 2015)*. 45–57.

[24] Giardino ND, Lehrer PM, and Edelberg R. 2002. Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. Psychophysiology. *Frontiers in public health* 39, 2 (2002), 246–253.

[25] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. 2007. BeepBeep: A High Accuracy Acoustic Ranging System Using COTS Mobile Devices. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems (SenSys '07)*. 1–14.

[26] K. Qian, C. Wu, F. Xiao, Y. Zheng, Y. Zhang, Z. Yang, and Y. Liu. 2018. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. 1574–1582.

[27] M. S. Raheel and et al. 2019. Breathing and Heartrate Monitoring System using IR-UWB Radar. *13th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 1–5.

[28] A. Santra, R. V. Ulaganathan, T. Finke, A. Baheti, D. Noppeney, J. R. Wolfgang, and S. Trotta. 2018. Short-range multi-mode continuous-wave radar for vital sign measurement and imaging. In *2018 IEEE Radar Conference (RadarConf18)*. 0946–0950.

[29] Nirjon Shahriar, Robert F. Dickerson, Qiang Li, Philip Asare, John A. Stankovic, Dezhi Hong, Ben Zhang, Xiaofan Jiang, Guobin Shen, and Feng Zhao. 2012. Musicalheart: A hearty way of listening to music. *In Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, 43–56.

[30] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin. 2011. A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 4144–4147.

[31] Anran Wang and Shyamnath Gollakota. 2019. MilliSonic: Pushing the Limits of Acoustic Motion Tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. 1–11.

[32] Lei Wang, Kang Huang, Ke Sun, Wei Wang, Chen Tian, Lei Xie, and Qing Gu. 2018. Unlock with your heart: Heartbeat-based authentication on commercial mobile phones. *In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 140.

[33] T. Wang, D. Zhang, L. Wang, Y. Zheng, T. Gu, B. Dorizzi, and X. Zhou. 2019. Contactless Respiration Monitoring Using Ultrasound Signal With Off-the-Shelf Audio Devices. *IEEE Internet of Things Journal* 6, 2 (2019), 2959–2973.

[34] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4 (2018), 20.

[35] Tong Wang, Mingcai Zhang, Qihao Yu, and Huyuan Zhang. 2012. Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal. *Journal of Applied Geophysics* 83 (2012), 29–34.

[36] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. 2017. Robust heart rate from fitness videos. *Physiol Meas* 38, 6 (2017), 1023–1044.

[37] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. *In Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*, 82–94.

[38] Y. Xie, F. Li, Y. Wu, S. Yang, and Y. Wang. 2019. D3-Guard: Acoustic-based Drowsy Driving Detection Using Smartphones. *IEEE Conference on Computer Communications, INFOCOM*, 1225–1233.

[39] Wei Xu, ZhiWen Yu, Zhu Wang, Bin Guo, and Qi Han. 2019. AcousticID: Gait-based Human Identification Using Acoustic Signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 25.

[40] Sangki Yun, Yi-Chao Chen, and Lili Qiu. 2015. Turning a Mobile Device into a Mouse in the Air. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys 2015)*. 15–29.

[41] Fusang Zhang, Zhaoxin Chang, Kai Niu, Jie Xiong, Beihong Jin, Qin Lv, and Daqing Zhang. 2020. Exploring LoRa for Long-Range Through-Wall Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 68 (2020), 27 pages.

[42] Fusang Zhang, Daqing Zhang, Jie Xiong, Hao Wang, Kai Niu, Beihong Jin, and Yuxiang Wang. 2018. From Fresnel Diffraction Model to Fine-grained Human Respiration Sensing with Commodity Wi-Fi Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (2018).

[43] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion Recognition Using Wireless Signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*. Association for Computing Machinery, 95–108.