LoEar: Push the Range Limit of Acoustic Sensing for Vital Sign Monitoring

LEI WANG, Peking University, China
WEI LI, Peking University, China
KE SUN, University of California San Diego, USA
FUSANG ZHANG, Institute of Software, Chinese Academy of Sciences and University of Chinese Academy of Science, China
TAO GU, Macquarie University , Australia
CHENREN XU, Peking University, China
DAQING ZHANG*, Peking University, China and Telecom SudParis, France

Acoustic sensing has been explored in numerous applications leveraging the wide deployment of acoustic-enabled devices. However, most of the existing acoustic sensing systems work in a very short range only due to fast attenuation of ultrasonic signals, hindering their real-world deployment. In this paper, we present a novel acoustic sensing system using only a single microphone and speaker, named LoEar, to detect vital signs (respiration and heartbeat) with a significantly increased sensing range. We first develop a model, namely *Carrierforming*, to enhance the signal-to-noise ratio (SNR) via coherent superposition across multiple subcarriers on the target path. We then propose a novel technique called *Continuous-MUSIC* (Continuous-MUltiple SIgnal Classification) to detect a dynamic reflections, containing subtle motion, and further identify the target user based on the frequency distribution to enable *Carrierforming*. Finally, we adopt an adaptive Infinite Impulse Response (IIR) comb notch filter to recover the heartbeat pattern from the Channel Frequency Response (CFR) measurements which are dominated by respiration and further develop a peak-based scheme to estimate respiration rate and heart rate. We conduct extensive experiments to evaluate our system, and results show that our system outperforms the state-of-the-art using commercial devices, *i.e.,* the range of respiration sensing is increased from 2 m to 7 m, and the range of heartbeat sensing is increased from 1.2 m to 6.5 m.

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

Additional Key Words and Phrases: Vital sign, Acoustic sensing

*Corresponding author.

Authors' addresses: Lei Wang, Key Laboratory of High Confidence Software Technologies (Ministry of Education), School of Computer Science, Peking University, China, Email:wang_l@pku.edu.cn; Wei Li, School of Computer Science, Peking University, China; Ke Sun, University of California San Diego, USA; Fusang Zhang, Institute of Software, Chinese Academy of Sciences and University of Chinese Academy of Science, China; Tao Gu, Macquarie University , Australia; Chenren Xu, School of Computer Science, School of Electronics Engineering and Computer Science, Peking University, Beijing, China; Daqing Zhang, Key Laboratory of High Confidence Software Technologies (Ministry of Education), School of Computer Science, School of Electronics Engineering and Computer Science , Peking University, China and Telecom SudParis, France, dqzhang@sei.pku.edu.cn.

2474-9567/2022/9-ART145 $15.00
https://doi.org/10.1145/3550293

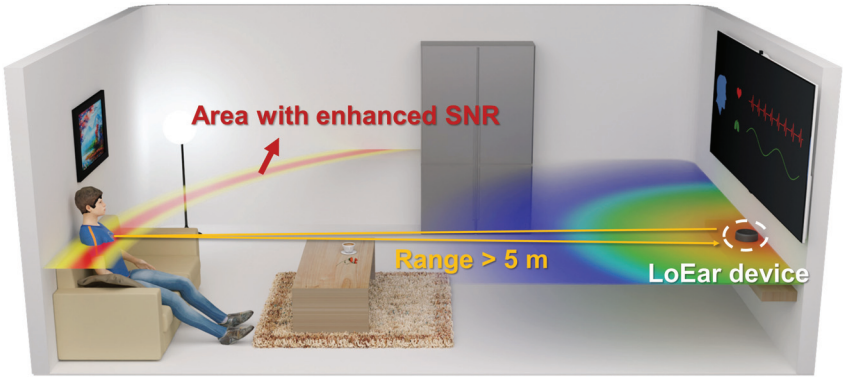


Fig. 1. Typical application scenario for LoEar

ACM Reference Format:
Lei Wang, Wei Li, Ke Sun, Fusang Zhang, Tao Gu, Chenren Xu, and Daqing Zhang. 2022. LoEar: Push the Range Limit of Acoustic Sensing for Vital Sign Monitoring . *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 145 (September 2022), 24 pages. https://doi.org/10.1145/3550293

1 INTRODUCTION

Acoustic sensing for vital sign detection has become a prevailing research topic in the mobile sensing community. Benefiting from the high sensing granularity [1] of acoustic signal and the wide availability of acoustic hardware (*i.e.,* speakers and microphones), many acoustic sensing techniques have been developed for respiration monitoring [2–5], and heartbeat monitoring [6, 7]. While most of the systems can achieve good accuracy, they have a common drawback–very short sensing range. This is constrained by the property of its fast acoustic signal strength decay over distance, and the effective sensing range is 1.2 m only for vital sign detection [6, 7], significantly limiting its large-scale deployment. It is worth noting that although the state-of-the-art WiFi-based sensing system [8] is capable of monitoring the respiration at the room scale, it is difficult to enable monitoring heartbeat due to the limitation of sensing granularity. Another wireless signal solution is capable of detecting heartbeat [9], but it uses a dedicated FMCW radar, which is very expensive.

To extend the sensing range of gesture detection, beamforming [10–12] can be applied to improve signal quality and deep learning [10, 13] has been widely used to mitigate interference. However, they may not be easily applied to Commercial off-the-shelf (COTS) devices due to hardware cost, computation cost, and battery life. Furthermore, their performance for vital sign detection is unclear since gesture detection typically has larger displacements. In this paper, we propose *Carrierforming*, a novel signal processing technique to extend acoustic sensing range with a pair of speaker and microphone. The design of *Carrierforming* leverages OFDM signal as a basis. Motivated by our observation of the pre-determined phase difference between subcarriers, the basic idea is that *Carrierforming* strategically delays the phase information across multiple subcarriers and aligns the peak to enable coherent superposition for SNR enhancement at a specific location (unlike a direction in beamforming).

However, the subject-of-interest has to be detected (or known) in prior to vital sign monitoring. In reality, it is challenging to differentiate weak subject-reflected signals from ambient-reflected signals. To address this challenge, inspired by the MUSIC algorithm, we first propose a novel technique, named *Continuous-MUSIC*, to detect the objects with **subtle motion** (*i.e.,* defined as the motion with the displacement of mm-level), even when the objects are further away. Note that subtle motion may include vital signs (*i.e.,* heartbeat and respiration), vibrations from household appliances, *etc.* Comparing to the standard MUSIC algorithm which is based on signal

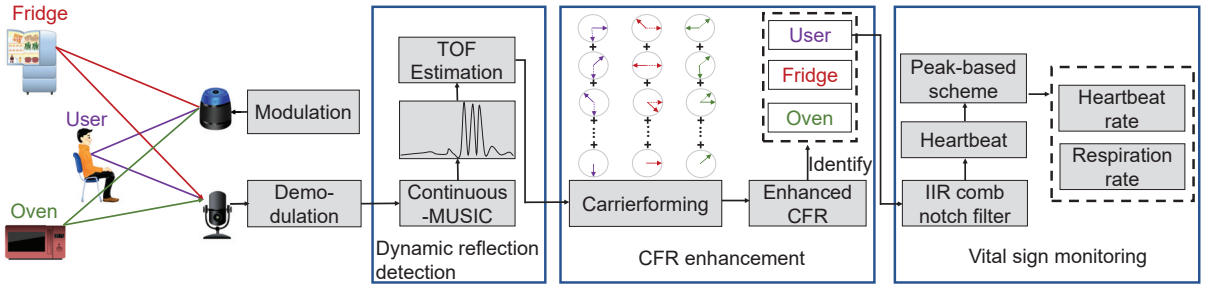


Fig. 2. System overview

samples of the same frame, our technique is built on received signals from multiple continuous frames. Both the subtle and larger scale motions will be singled out in this method. We can then identify the target user with vital signs based on the property of frequency distribution in CFR enhancement. The basic idea is that user respiration is a quasi-periodical movement with dominant frequency components in the range of 0.15 $\sim$ 0.3 Hz, while the frequency components of other subtle motions do not fall in this range. In this way, the proposed two-step technique has a superior advantage, *i.e.,* being able to distinguish the target user with vital signs from some dynamic objects and other static objects such as desks, walls, and the floor.

By significantly enhancing SNR, LoEar can sense high-resolution subtle movement. To demonstrate its sensing capacity, we apply LoEar in respiration and heartbeat monitoring as a case study. Compared to the state-of-the-art, LoEar increases the sensing range from 2 m to 7 m in respiration monitoring, and from 1.2 m to 6.5 m for heartbeat monitoring, significantly pushes the sensing range to room-scale. In addition, LoEar has the capacity of identifying multiple users with different path lengths and monitoring their respiration and heartbeat simultaneously. As shown in Fig. 1, LoEar can be applied to a wide range of acoustic devices such as voice assistants to build a health surveillance system for smart home.

In summary, the main contributions of this work are summarized as follows:

• To our best knowledge, LoEar appears to be the first to explore the feasibility of acoustic sensing for vital signs, including both respiration and heartbeat, at room-scale using a single microphone and speaker.

• We develop a SNR enhancement sensing system based on the constructive superposition across multiple subcarriers.

• We propose a novel technique, built on received signals from multiple continuous instants, to detect the dynamic objects containing objects with subtle motions, even though object is further away. We further employ the property of frequency distribution to identify the target user based on the quasi-period of respiration.

• We build a real-time vital sign monitoring system using COTS acoustic devices and extensive experiments to demonstrate its performance in the presence of environmental interference.

2 SYSTEM OVERVIEW

Fig. 2 gives an overview of LoEar. In LoEar, a loudspeaker transmits OFDM-modulated acoustic signals, and the reflected signals will be recorded by microphones and then demodulated to obtain CFR measurements (Section 3.1) for motion detection. The design goal of LoEar is to push the range limit of acoustic sensing, which aims to fulfill the following three requirements.

(i) Detecting target user (Section 4). LoEar is able to detect the subtle movement of target, *i.e.,* vital signs, within a range of 7 m from the transceiver by using a single pair of speaker and microphone. We first propose a new technique, called *Continuous-MUSIC*, to detect the dynamic reflections containing the subtle motions. We then identify the target user using spectral concentration in the specific range related to respiration.

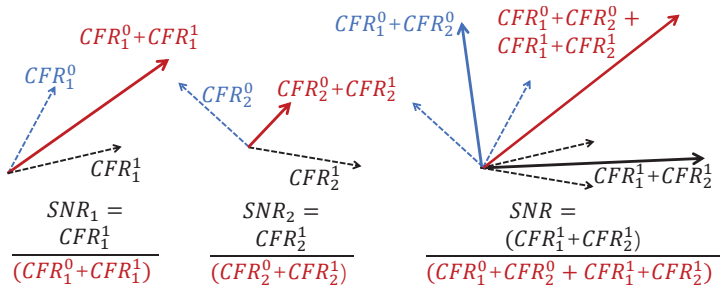


Fig. 3. Direct addition of CFRs across two subcarriers

(ii) CFR enhancement (Section 3). Once we estimate the TOF of the target user, we design *Carrierforming* to enhance the SNR of the signal at the reflection path. The key idea is to compensate the phase offsets across multiple subcarriers with different frequencies to align the peaks of the selected subcarriers at the target path.

(iii) Vital sign monitoring (Section 5). By enhancing SNR, we derive the CFR changes mainly caused by vital signs, including respiration and heartbeat. As the faint heartbeat signals are drowned out by respiration, we focus on recovering heartbeat signals from the mixed CFR measurements. The key idea is that the heartbeat signal is a quasi-periodic movement with a specific fundamental frequency (FF) different from the respiration signal. Then, we develop a peak-based scheme to estimate the instant rhythms of heartbeat and respiration, respectively.

3 UNDERSTANDING CARRIERFORMING

In this section, we describe the detailed design of *Carrierforming*. We first introduce the property of Orthogonal Frequency-Division Multiplexing (OFDM) and how to derive the CFR measurements. Second, we propose the basic model to enhance the SNR on a particular path. Finally, we analyze the effect of factors, *i.e.,* bandwidth and frame size, on the performance in depth.

3.1 Primer on CFR Measurements

Our modulation and demodulation schemes are based on OFDM [14–16]. The advantage of OFDM is that it splits up the bandwidth into orthogonal subcarriers, which means that "crosstalk" between the sub-channels is eliminated and inter-carrier guard bands are not required. This significantly simplifies the process of modulation and demodulation to extract the CFR measurements of multiple subcarriers.

Similar to the prior works [15, 16], we choose the Zadoff-Chu (ZC) sequence as baseband signals to achieve optimal auto-correlation properties. We define that the length of the ZC sequence is N_s, the bandwidth of the baseband signal is B, and the sampling rate is $f_s = 48$ kHz, resulting in the frame length of $N = \frac{N_s \times f_s}{B}$ for each modulated signal. The central frequency f_c is set to ensure the bandwidth after modulation $(f_c - \frac{B}{2}) \sim (f_c + \frac{B}{2})$ is inaudible to most people.

We follow the standard steps of OFDM modulation and demodulation at the transmitter and receiver end, as illustrated in prior works [15, 16]. Finally, we derive the CFR as follows:

$$CFR_n = \sum_{i=0}^{M-1} A_i e^{-j2\pi f_n \tau_i}, \tag{1}$$

where $n \in [1, N_s]$ is the subcarrier index, M is the number of sound propagation paths, A_i is the attenuation coefficient of path i, φ_i is the corresponding phase shift caused by propagation, τ_i is the Time of Flight (TOF) of

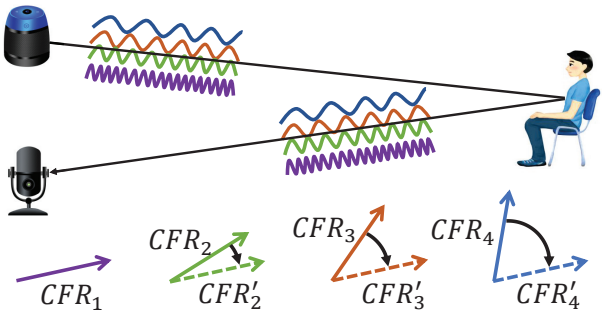


Fig. 4. Basic idea of Carrierforming. CFR_1, CFR_2, CFR_3, and CFR_4 denote the CFR of four consecutive subcarriers for a specific propagation path, respectively. They have different phases due to different frequencies. We can compensate for these phase differences so that the CFR of all subcarriers have the same phase (*i.e.,* $\angle CFR'_1 = \angle CFR'_2 = \angle CFR'_3 = \angle CFR'_4$) to further enable constructive CFR summation across multiple subcarriers.

path i, f_n is denoted as the frequency of n-th subcarrier. Note that OFDM guarantees that the subcarriers are orthogonal to each other, with the frequency of the n-th subcarrier as $f_n = f_1 + (n-1)f_s/N$.

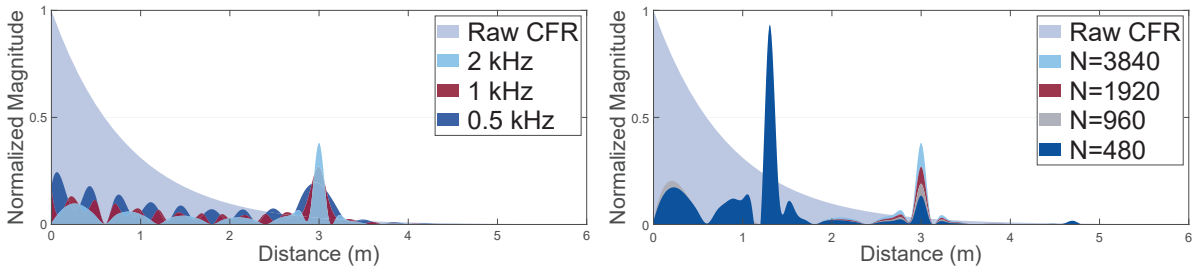
3.2 Basic Model

With only one pair of a transceiver, a promising way for SNR enhancement is to mitigate the multi-path effect by combing the measurements from multiple subcarriers due to the frequency diversity. Based on the measurements obtained from different subcarriers, LLAP [17] utilizes linear regression over multiple subcarriers to mitigate the multi-path effect. Detailedly, they select the subcarriers which fit the distance change curve during a short period and then use the measurements from these subcarriers to enable a tracking system. Since this scheme is based on the high SNR of the raw measurement, it does work when the movement is close to the transceiver. However, when the movement is far away from the transceiver, this scheme fails due to the low SNR.

How can we efficiently combine measurements obtained from multiple subcarriers to enhance SNR? An intuitive method is to add up measurements obtained from multiple subcarriers directly. To study the feasibility, we give the Figure 3 to illustrate the adding process. We assume there are only two paths in the environment, *i.e.,* reference path 0 and target path 1, respectively. For the subcarrier n, we use vectors CFR_n^0 and CFR_n^1 in complex domain to represent the CFR measurements corresponding to these two paths. After taking addition operation between CFR measurements of two subcarriers, the resulting SNR corresponding to the target path has a slight increase compared to subcarrier 1 while it has an evident decrease comparing to subcarrier 2. This means that the direct CFR superposition across multiple subcarriers is infeasible to increase SNR due to the phase shift caused by the frequency diversity.

To address this challenge, we propose a new scheme, namely *Carrierforming*, to increase the SNR within a specified range so as to enable far sensing. The key insight is similar to another famous signal processing technique, beamforming. However, compared to the beamforming scheme, which increases the SNR of the received signals based on received microphone/antenna array, we propose to combine shifted CFR measurements of multiple subcarriers using only one pair of transceiver.

Now, we will introduce the process idea of *Carrierforming*. *Given the TOF τ of the target path derived previously, we now explain how to estimate it in Section 4 in detail.* Since the frequencies of the subcarriers are distributed with linearly spaced frequencies from f_1 to $f_{N_s} = f_1 + (N_s - 1)f_s/N$ in OFDM signals, the received phase caused by the same propagation path is also linearly distributed from $2\pi f_1 \tau$ to $2\pi f_{N_s} \tau$. Fig. 4 shows an example of CFR



(a) Depth of field with various bandwidths

(b) Grating lobes with various frame size

Fig. 5. The effect of different factors on the enhancement.

phase results from four consecutive subcarriers. Our key insight is to compensate for these phase differences at the transmitter side or receiver side so that the CFR of all the subcarriers will have the same phase for a specific propagation path, and then we can align the phase of subcarriers to achieve constructive CFR summation.

Next, we formally prove *Carrierforming* by assuming that there is a single propagation path with TOF τ_0 and the attenuation coefficient A_0 for the target path changes exponentially with propagation delay [18]. We define sound attenuation amplitude as $A_0 = Ae^{-\alpha\tau_0}$, where A and α are the constant values. The CFR of n-th subcarrier can be expressed as $CFR_n = Ae^{-\alpha\tau_0}e^{-j2\pi f_n\tau_0}$. Then, for the n-th subcarrier, we compensate the CFR measurement corresponding to the target path with the constructive offset $e^{j2\pi n\Delta f\tau}$, where τ is denoted as the delay offset measured by our algorithm and Δf is the frequency spacing between two consecutive subcarriers. The synthetic CFR measurements for all subcarriers relative to the target path τ_0 can be expressed as follows:

$$CFR_{e_0} = CFR_1 \sum_{n=1}^{N_s} e^{j2\pi(n-1)\Delta f(\tau-\tau_0)}. \tag{2}$$

Ideally, if τ_0 is equal to τ, $|CFR_{e_0}|$ is equal to $N_s|CFR_1| = N_sA_0$, which indicates that the amplitude of synthesized CFR at the target path will be improved by N_s times than that of CFR in a single subcarrier. In contrast, the synthesized amplitude at other paths experiences varying degrees of attenuation due to delay difference, leading to the synthesized CFR of smaller than N_sA_0. Therefore, we can significantly improve the SNR at the target path to enable stronger sensing capability.

3.3 Depth of Field Quantification

Based on the *Carrierforming* model, we improve SNR for the target path with a fixed delay. However, it is unclear about the changes in SNR when the path is around the target. In this section, we focus on quantifying **Depth of Field (DOF)**, defined as the target sensing range, in which the SNR should be efficiently enhanced and decreased when out of range.

We assume that there are m propagation paths from speaker to microphone in total, and the ToF of path i is denoted as τ_i, where $i = 0, \ldots, M-1$. For path i, the amplitude of the enhanced CFR can be rewritten as:

$$\begin{aligned} CFR_{e_i} &= CFR_1 \frac{1 - e^{j2\pi N_s\Delta f(\tau-\tau_i)}}{1 - e^{j2\pi\Delta f(\tau-\tau_i)}} \\ &= CFR_1 \frac{\sin(N_s\pi\Delta f(\tau-\tau_i))}{\sin(\pi\Delta f(\tau-\tau_i))}. \end{aligned} \tag{3}$$

(a) Setup for sensing heartbeat (b) Setup for sensing fan rotation

Fig. 6. Experimental settings for verification.

Since the multipath is mainly distributed around the target leading to $\sin(\pi\Delta f(\tau-\tau_i)) \sim \pi\Delta f(\tau-\tau_i)$, the amplitude of the superimposed CFR is approximated to:

$$|CFR_{e_i}| \quad = \quad N_s|CFR_1||\text{sinc}(N_s\pi\Delta f(\tau-\tau_i))|. \tag{4}$$

Fig. 5(a) simulates the result of multipath interference. The superimposed CFR has an evident degradation trend with the value increase of $|\tau-\tau_i|$. This indicates that the destructive effect of noisy multipath can also be suppressed when the target path and the multipath are further apart. However, for the nearby multipath, the detailed effect is unclear. Therefore, we need further to formulate the relationship between the target path and the surrounding multipaths.

We use the measured ToF τ_β with Half Power Delay Width (HPDW) as the metrics to character the side lobe interference caused by surrounding multipaths, in which the radiation pattern decreased by 50% (or -3 dB) from the peak of the main lobe. Therefore, we derive the quantitative relationship between the bandwidth and ToF τ_β, as follows:

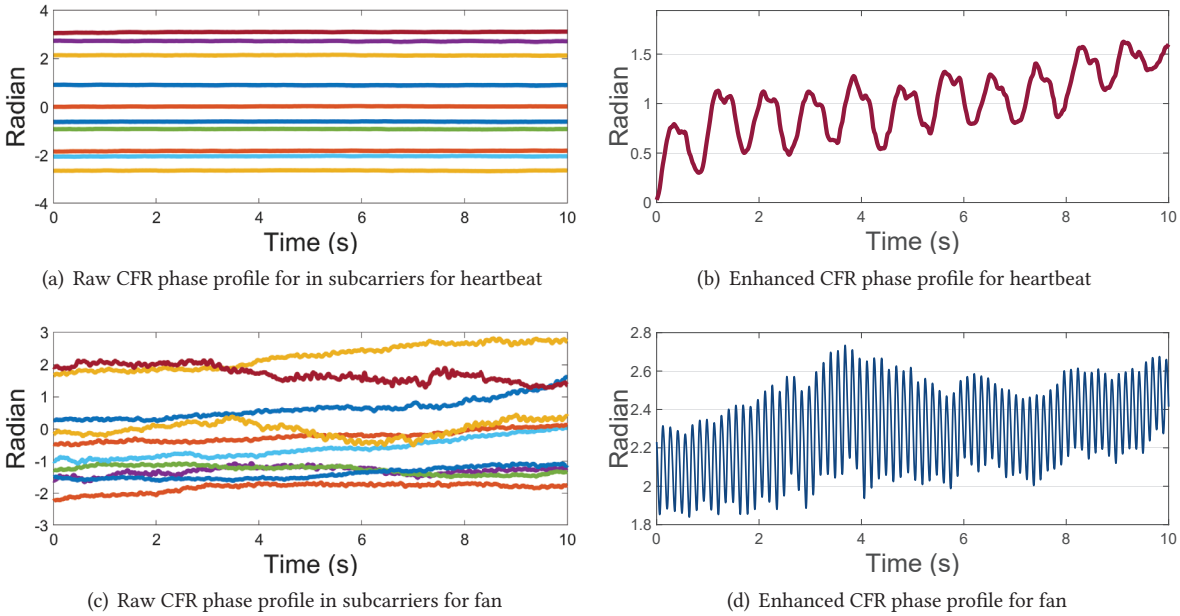
$$\mathcal{B} \quad = \quad N_s\Delta f \simeq \left|1.4/[\pi(\tau_\beta-\tau_0)]\right|, \tag{5}$$

where $\mathcal{B}$ represents the bandwidth of our transmitted signals. Eq. (5) shows that the selected bandwidth is anti-correlated with HPDW, which denotes the valid enhanced range. For the path with $|\tau_i-\tau_0| > |\tau_\beta-\tau_0|$, the power degrades by more than 50%, and the decay increases with the difference between τ_i and τ_β. Therefore, we can effectively reduce multipath's destructive effect outside the sensing range by selecting a suitable bandwidth. The larger bandwidth implies the narrower enhanced range, which may be inside the range of motion change. In contrast, the smaller bandwidth may bring the multipath interference in the vicinity. To balance the trade-off, we empirically choose 3.25 cm (path length range of 6.5 cm due to round-trip reflection) as the enhanced sensing range. The TOF difference between HPDW and the target path can be calculated as $|\tau_\beta-\tau_0| = (2\times 0.0325)/343 = 0.022$ ms, which results in a bandwidth of about 2 kHz according to Eq. (5).

3.4 Enhanced SNR vs. Frame Length

As illustrated in Fig. 5(b), the SNR of both the target path and other paths may be enhanced. We denote the lobes corresponding to some non-target paths, of which the SNR are also constructively enhanced, as **grating lobes**. The main reason for grating lobes is that the frequency spacing $\Delta f = B/N_s = f_s/N$ is too large which leads to the phase ambiguity of phase compensation $e^{-j2\pi(n-1)\Delta f(\tau-\tau_0)}$ for different paths, which means:

$$|2\pi\Delta f(\tau-\tau_0)| \quad = \quad k2\pi, k = 0, 1, 2, \ldots \tag{6}$$



(a) Raw CFR phase profile for in subcarriers for heartbeat

(b) Enhanced CFR phase profile for heartbeat

(c) Raw CFR phase profile in subcarriers for fan

(d) Enhanced CFR phase profile for fan

Fig. 7. CFR phase profile changes with enhancement.

where k is an integer number. Obviously, the main lobe corresponding to the target path is identified by $k = 0$ and other integer values of k will identify the grating lobes. Eq. (6) shows that the grating lobes must exist when the sensing range is sufficiently large. In contrast, for the limited range, we can cancel the grating lobes by:

$$|2\pi\Delta f(\tau - \tau_0)| \quad < \quad 2\pi. \tag{7}$$

Moreover, the requirement to satisfy Eq. (7) is given by:

$$\Delta f \quad < \quad 1/|\tau - \tau_0|_{max}, \tag{8}$$

where $|\tau - \tau_0|_{max}$ is the maximal value for $|\tau - \tau_0|$. For example, to avoid grating lobes within 6m (about 12m in path length difference due to the round-trip reflection) from the target, the frame length N should be greater than 1678 to ensure that the frequency spacing is less than 28.6 Hz.

Furthermore, as discussed in Section 3.2, the enhanced SNR is correlated with the number of subcarriers N_s. Since we have determined the bandwidth in Section 3.3, the number of subcarriers equals to $N_s = B \times N/f_s$, which indicates that larger N will give rise to higher SNR. Ideally, we want to improve SNR as much as possible. However, the larger length also brings a serious issue, namely, increasing the range with multipath effect. As discussed in Section 3.2, the SNR of the target path in DOF will be enhanced significantly, while the SNR of paths further from DOF will be decreased. However, the SNR of paths close to DOF can also be slightly enhanced, indicating the existence of multipath effect. Suppose the overall transmission amplitude is A, if the amplitude of CFR_{ei} is larger than A, the propagation in path i is constructively enhanced. Our target is to enable constructive enhancement in the small range around the target, with the destructive effect outside the range. Given the bandwidth, *i.e.,* $N_s\Delta f$, the enhanced range $|\tau - \tau_0|$ is inversely correlated with the number of subcarriers, which is linearly proportional to N. As a result, the larger N will lead to the wider enhanced range, implying that the SNR of multipath in this range will also be improved. For example, when N is 3840, multipath reflections within

the path length range of 62.7 cm from the target will be enhanced to varying degrees. Overall, there is a trade-off between enhanced SNR and frame length. We finally set N to 1920 so that the system can avoid grating lobe within 6.86 m (path length range of 13.72 m) from the transceiver, covering most space in the room. Furthermore, the multipath effect with a range difference of more than 22.64 cm (path length difference of 45.28 cm) from the target will be destructively mitigated.

3.5 Model Verification with Experiments

To verify the effect of signal enhancement with *Carrierforming*, we ask a volunteer to sit 5 m away from the transceiver holding his breath in a relatively static environment, as shown in Fig. 6(a). At the moment, there are no dynamics other than subject's heartbeat in the environment. In this way, the weak heartbeat-caused signals are not buried into respiration or other environmental dynamics. Similar to prior work [14, 17], we derive the profile changes of CFR phase in selected 10 subcarriers, as shown in Fig. 7(a). We cannot observe any regular changes in heartbeat except for environmental noise. However, when the compensated CFR measurements in multiple subcarriers are added up together, we can clearly observe the pattern caused by heartbeat, as shown in Fig. 7(b). The above results demonstrate the superior capacity of the *Carrierforming* model in enhancing SNR with long range.

To further verify the enhanced performance of *Carrierforming*, we try to monitor the rotation of electric fans at a long distance. As shown in Fig. 6(b), we replace the user with an electric fan at the same position. The result further demonstrates the superiority of *Carrierforming* in enhancing SNR. Fig. 7(d) shows that the profile after *Carrierforming* is periodic corresponding to the fan rotation, while the raw profiles are irregular due to low SNR, as shown in Fig. 7(c).

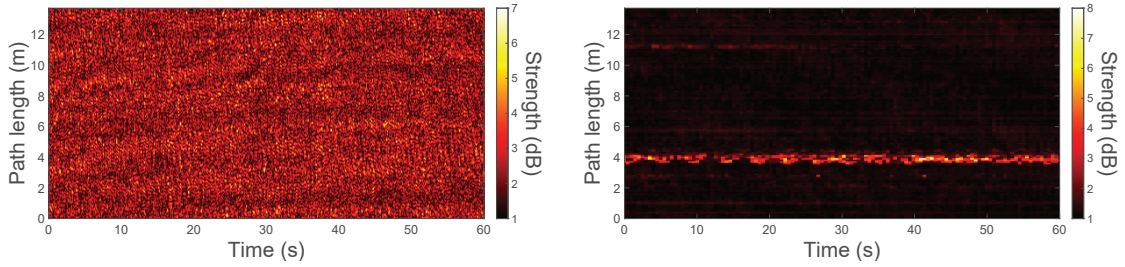
4 TARGET USER EXTRACTION

The purpose of modeling *Carrierforming* is to enhance SNR with only one pair of speaker and microphone. However, without the known initial path length of the target, it is challenging to derive an accurately constructive offset. In this section, we focus on searching for the target candidates, and differentiating the moving target path from the interfering multi-paths.

4.1 Target Candidate Detection

This section focuses on detecting target candidates (*i.e.,* all the dynamic reflections, including the large-scale and subtle motions) in the environment. A natural solution is to use the time-delay profile based method to detect the dynamic paths. Similar to ResTracker [15], we can first get the complex-valued CIR after converting CFR_n back to the time domain by Inverse Discrete Fourier Transform (IDFT). We then take difference of the CIR estimation along the time axis to distinguish those dynamic reflection paths. In this way, the static reflection paths can be removed, and we can detect the target's path delay through the corresponding lightened peak. This solution makes sense when the target is nearby the transceiver with evident movements such as pushing hand and nodding. However, when the user keeps still with only subtle motions (respiration and heartbeat), it turns out to be invalid. As shown in Fig. 8(a), it is challenging to identify the user path when the user sits 2 m away from the transceiver. These results will be worse when the target is further away due to the attenuation of SNR.

To address this issue, we propose *Continuous-MUSIC* to detect and estimate the TOF of the target path. The MUSIC algorithm has been developed to measure Angle of Arrival (AOA) and TOF using multiple antennas and subcarriers jointly [19–21]. Inspired by this method, we adopt incident signals on multiple subcarriers to estimate TOF. The key insight is that incident signals from different distances introduce different amounts of phase changes on each subcarrier.



(a) Differential CIR profile (b) Continuous-Music profile

Fig. 8. Profile comparison when the user sits at 2 m from the transceiver.

Now, we describe how to estimate TOF based on MUSIC using multiple subcarriers. Assume there are N subcarriers at the receiver, and the target is located with propagation delay τ. The frequency spacing between adjacent subcarriers is set to Δf which should avoid the grating lobes. Obviously, the phase difference of adjacent subcarrier can be expressed as $e^{-j2\pi\Delta f\tau}$. We then use steering vector $\mathbf{a}(\tau)$ to describe the phase shift related to the first subcarrier as a function of the TOF of the target path:

$$\mathbf{a}(\tau) \quad = \quad [1, e^{-j2\pi\Delta f\tau}, \ldots, e^{-j2\pi(N-1)\Delta f\tau}]^T. \tag{9}$$

The received signals at the subcarrier array with time t can be expressed as:

$$\mathbf{X}(t) \quad = \quad \mathbf{a}(\tau)s(t) + \mathbf{N}(t), \tag{10}$$

where $\mathbf{N}(t)$ denotes the noise vector, $s(t)$ denotes the signal received at the first subcarrier, and $\mathbf{X}(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$. Additionally, when there are K incident signals with time delay of $\tau_1, \tau_2, \ldots, \tau_K$, the received signals vector can be further expressed as:

$$\mathbf{X}(t) \quad = \quad \mathbf{AS}(t) + \mathbf{N}(t), \tag{11}$$

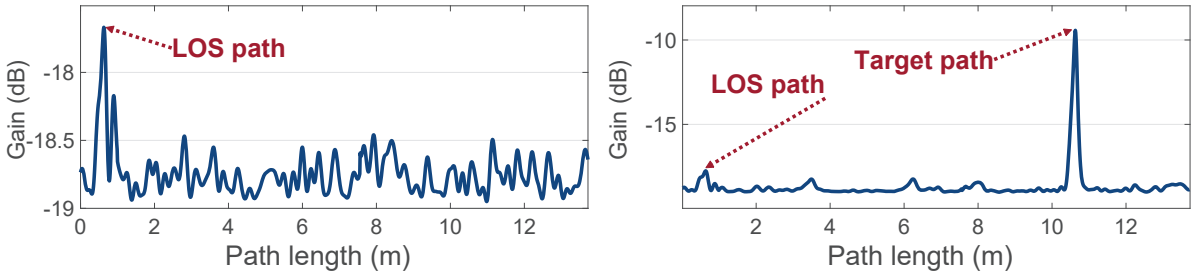
where $\mathbf{A} = [\mathbf{a}(\tau_1), \ldots, \mathbf{a}(\tau_2), \mathbf{a}(\tau_K)]$ and $\mathbf{S}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^T$.

The key idea behind MUSIC is that the eigenvectors with the largest eigenvalues of $\mathbf{X}(t)\mathbf{X}(t)^H$ (H represents the conjugate transpose operation) corresponds to the signal subspace and the remaining eigenvectors construct the noise subspace. Besides, these two subspaces are orthogonal to each other.

Based on this theory, the next standard process is to compute the eigenvectors of the correlation matrix $\mathbf{X}(t)\mathbf{X}(t)^H$ with the $(N-K)$ smallest eigenvalues and then construct the corresponding noise vector space $\mathbf{E}_{N-K}(t) = [\mathbf{e}_{K+1}(t), \mathbf{e}_{K+2}(t), \ldots, \mathbf{e}_N(t)]$. Due to orthogonality, the spectrum function of traveling delay can be expressed as:

$$P_t(\tau) \quad = \quad \frac{1}{\mathbf{a}^H(\tau)\mathbf{E}_{N-K}(t)\mathbf{E}_{N-K}(t)^H\mathbf{a}(\tau)}. \tag{12}$$

The peaks in the traveling delay profile indicate the target's path. This standard method is indeed effective for measuring the delay of the strong LOS or for measuring the delay of the singly obvious moving target's reflection when LOS is relatively weak. However, when a subject is further away from the transceiver with low SNR or keeps always still, the prior method will be disabled. For example, we set the distance between microphone to 1 m and ask a subject to sit 5 m away from the midway point between them. As shown in Fig. 9(a), we can see the only strong peak indicating the LOS path. This is mainly because the strength of LOS is much greater than the reflection from the target and surrounding objects (*e.g.,* desk, wall, *etc.*).



(a) Gains with the standard MUISC algorithm

(b) Gains with Continuous-MUSIC algorithm

Fig. 9. Comparison between the standard MUSIC and Continuous-MUSIC algorithm.

To address this problem, we propose the *Continuous-MUSIC* technique, which is founded on the received signals from multiple continuous timing points, to accurately detect the weak target reflection with subtle motions. In detail, we first build the received signal matrix as $\mathbf{X} = \{\mathbf{X}(t_i) \mid 1 \leq i \leq l\}$, which consists of l continuous frames of received signals. Second, we take eigen decomposition on the correlation matrix $\mathbf{X}\mathbf{X}^H$ and derive the $(N-K)$ eigenvectors (*i.e.,* $\mathbf{E}_{N-K} = [\mathbf{e}_{K+1}, \mathbf{e}_{K+2}, \ldots, \mathbf{e}_N]$) with the smallest eigenvalues corresponding to noise. Finally, similar to Eq. (12), we derive the spectrum function of traveling delay as follows:

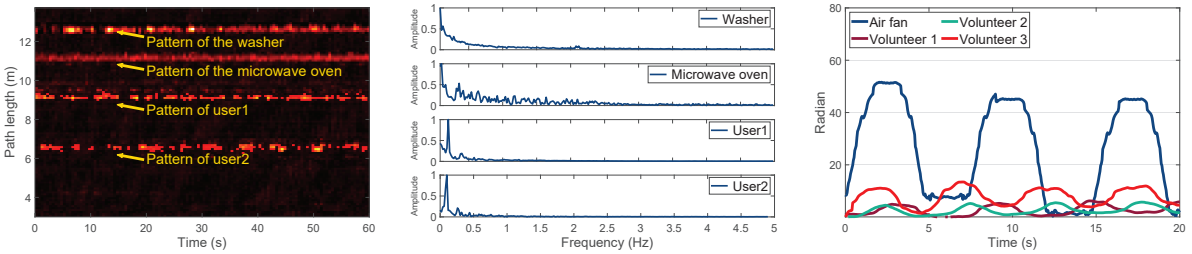
$$P(\tau) \quad = \quad \frac{1}{\mathbf{a}^H(\tau)\mathbf{E}_{N-K}\mathbf{E}_{N-K}^H\mathbf{a}(\tau)}. \tag{13}$$

This new model has one superior advantage: It can distinguish dynamic reflections, including subtle motions, from other static reflections such as desk, wall, and the floor. Fig. 9(b) shows that the peak corresponding to target is significantly highlighted compared to Fig. 9(a) based on the *Continuous-MUSIC* scheme. In comparison, the LOS peak has been reduced significantly in the profile since it retains static in continuous time. To further demonstrate the superiority over the prior solution used in ResTracker, we process the same CFR measurements with *Continuous-MUSIC*. Fig. 8(b) shows the resultant Continuous-MUSIC profile change within a period of 60 s. We observe that respiration causes a bright line in the correlated path.

We now focus on detecting the target candidates mathematically and measuring their path lengths. As shown in Fig. 9(a), the standard MUSIC can detect the strongest LOS path by localizing the maximum peak in the profile. Although the gain regarding the LOS path is significantly weakened in *Continuous-MUSIC*, it still has a more remarkable peak relative to other static paths, as shown in Fig. 9(b). Benefiting from MUSIC, we can derive the gain regarding the LOS path in the *Continuous-MUSIC* profile. On the other hand, the gains caused by the dynamic reflections are significantly improved in *Continuous-MUSIC*, resulting in much larger gains related to all static paths. Thus, we propose to use a threshold-based scheme to detect the dynamic reflections (target candidate), *i.e.,* once the local peak in *Continuous-MUSIC* result exceeds the empirical threshold Thr (set as three times the gain regarding the LOS path in *Continuous-MUSIC* result), we then determine that there is a target candidate in the environment. Meanwhile, the path length of a target candidate can be calculated by localizing the peaks.

4.2 Target User Identification

In this section, we need to identify and signal out the paths corresponding to the target users. As discussed in Section 4.1, *Continuous-MUSIC* can detect multiple subjects with motions and add them into the candidate set. However, it is difficult to identify the target user from the candidate set. We take the household appliances, *i.e.,* washer and microwave oven, as an example. As shown in Fig. 10(a), it is difficult to determine the pattern



(a) Continuous-MUSIC result (b) Spectrum distribution (c) Phase profile for the fan and respiration

Fig. 10. Pattern comparison between the household appliances and users.

corresponding to the target user in the *Continuous-MUSIC* profile when we put a working washer and microwave oven in the room meanwhile. Fortunately, we observe that the respiration of the user is a quasi-periodical movement with dominant frequency components in the range of 0.15 ∼ 0.3 Hz, while the frequency components of the appliances are not concentrated in this range, as illustrated in Fig. 10(b). Thus, we regard the spectral concentration in the specific range as the unique feature to identify the target user. In detail, we qualify the spectral concentration by calculating the ratio of the spectral amplitude superposition in the specific range of 0.15 ∼ 0.3 Hz and overall range of 0.1 ∼ 5 Hz as follows:

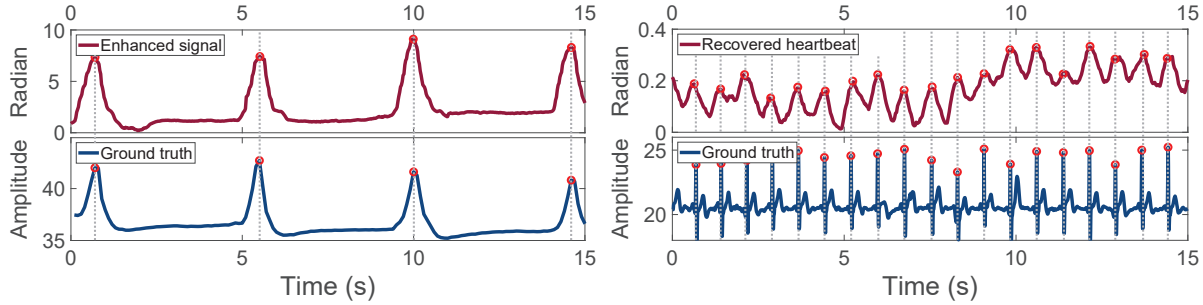
$$r_{res} = \frac{\sum_{f=0.15}^{0.3} |FFT(f)|}{\sum_{f=0.1}^{5} |FFT(f)|}. \tag{14}$$

If the ratio r_{res} is larger than the threshold Thr_{rt}, the pattern is corresponding to the user. We set the experienced Thr_{rt} to 0.7 to avoid false alarms. Note that the frequency range is mainly concentrated in 0.1 ∼ 5 Hz since the lower frequency mainly corresponds to the Direct Component (DC) component, and the higher frequency mainly corresponds to the small environmental noise. The current scheme indeed works for most indoor environments. However, when there are objects in the environment with similar frequency patterns, such as the clock pendulum and air fan with a periodic waving head, it may fail to identify the target user accurately. Since natural respiration induces a mm-level (*i.e.,* 5 ∼ 12 mm) displacement of the chest [22], while most objects have much larger scale displacements, it is promising to utilize the phase range in motion to identify the target user. We have three subjects in this experiment. Two of them breathe naturally, and the third subject takes a deep breath. We compare the result with that of the air fan, as shown in Fig. 10(c). We observe that the phase range for natural respiration is about 5.3 radians, nearly ten times lower than that for the air fan (*i.e.,* 51.2 radians). Even though the deep respiration has a phase range of 13.4 radians, it is still far below the air fan. Based on the above observations, we set the threshold of phase range for natural respiration as 15 radians to filter out most interfering objects with similar frequency patterns in the indoor environment. Additionally, our system also has the capacity of separating multiple users based on different path lengths relative to the transceiver. As shown in Fig. 10(a), two users with two path lengths of 6.7 m and 9.1 m can be clearly identified.

5 VITAL SIGN MONITORING

5.1 Heartbeat Recovery

In this subsection, we focus on recovering the heartbeat from the enhanced CFR measurements. Fig. 11(a) shows the waveform of the enhanced CFR when the volunteer sits at a distance of 5 m from the transceiver. As the volunteer is still, the CFR measurements mainly consist of two components, *i.e.,* respiration and heartbeat.



(a) Waveform of the enhanced signal

(b) Reconstructed heartbeat signal

Fig. 11. Vital signs recovery.

However, the chest movement caused by respiration is larger than that caused by the heartbeat, resulting in the heartbeat signals buried in the respiration signals.

To recover the heartbeat, we first perform DFT on the enhanced CFR measurement over a period of 10 s to obtain the spectrum of the signal. As the heartbeat is a quasi-periodic movement of which the FF is in the specific range of 0.8 ∼ 2 Hz [7], we second search for the prominent peak pointing to the FF in the spectrum. Third, we employ an adaptive IIR comb notch filter [23], which is normally used to filter out the FF and the corresponding harmonic components, on the CFR measurements. In detail, the order of the notching filter is set as the ratio of the sampling rate and FF (*i.e.,* f_{CFR}/FF). The heartbeat is, thus, recovered by taking the difference between the enhanced CFR measurements and the filtered result, as shown in Fig. 11(b). Note that the above scheme has a potential issue: whether the heart rhythm overlaps with the respiration rhythm. Even partial overlap will make it fail to recover the heartbeat. To explore the distribution of respiration and heart rates, we collected 7,000 samples of respiration and heart rhythms, respectively, from 15 subjects (7 females and 8 males) aged from 21 to 56. Fig. 12 summarizes these subjects. During the collection, we ask all subjects to be in different postures, *i.e.,* sit, lie down and stand. From Fig. 13, we observe that there is no overlap in the distribution of respiration and heart rhythms, which further demonstrates the feasibility of the heartbeat recovery process.

5.2 Instant Rhythm Estimation

Although the average heart rhythm (AHR) can be inferred via DFT operation in Section 5.1, it is insufficient to qualify the heart rhythm variability (HRV), which is an important index for heart disease detection [24–26]. We develop a peak-based scheme to estimate the instant heart rhythm (IHR) over a certain period. The scheme undergoes the following three steps. First, we locate all peaks from the phase sequence of enhanced CFR measurements and rearrange them in order of their values. Second, we build the candidate set with the initial element of the highest peak. Starting from the second highest peak, we add the peaks into the candidate set in the descending order of values. When the sampling interval of the current peak and one of the candidate peaks is smaller than the threshold τ_p, the current peak will be removed. Here, we set the threshold τ_p to 0.5 s by assuming that the heart rhythm of the user is between 50 Beats Per Minute (BPM) and 120 BPM [27]. Accordingly, the beat interval is in the range of 0.5 ∼ 1.2 s. Third, the candidate peaks need to go through a further pruning step to avoid false alarms. For example, if the user's heart rhythm is 50 BPM with the interval of 1.2 s for two continuous peaks, the noisy peaks between these two neighboring peaks may be falsely put into the candidate set. As shown in Fig. 11(b), the prominence of the noisy peak (*e.g.,* the third peak from the left of the figure) is far lower than the actual peaks. Based on this observation, the candidate set can be further refined by wiping out the peaks with

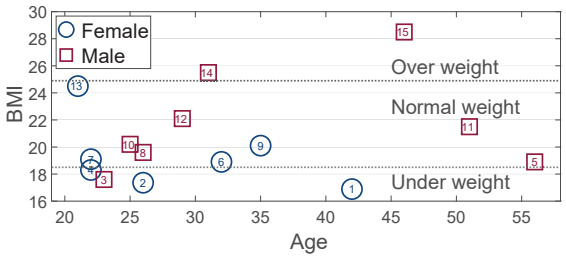


Fig. 12. Characteristics of participated subjects. All subjects are indexed by the value of BMI in increasing order.

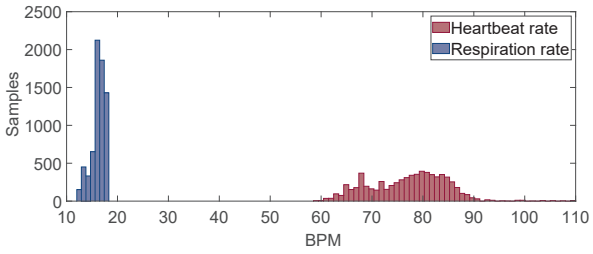


Fig. 13. Distribution of respiration and heartbeat rates.

low prominence. In detail, we find the two minimum points in the interval between the candidate peak and the two neighboring peaks on the left and right, respectively. The higher of the two minimum points is specified as the reference. If the height of the candidate peak above the reference is smaller than the empirical threshold Thr, it will be shifted out. Specially, for the first and the last candidate peak, the left and right endpoints will be regarded as the corresponding neighboring peaks. We set Thr to 0.2 times the difference between the maximum and minimum over the period. The resulting peaks are marked as the red dots, as shown in Fig. 11(b), and the IHR can be derived by calculating the interval of the consecutive peaks. Note that instantaneous respiratory rhythm (IRR) can be estimated using the same peak-based scheme. We have omitted the description for the sake of space.

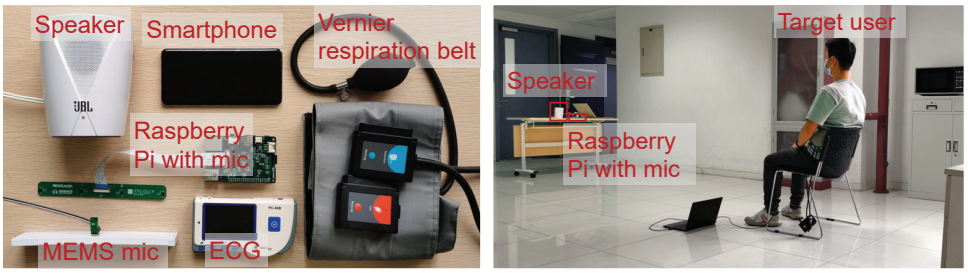
6 IMPLEMENTATION

We implement LoEar on three acoustic devices, including *i)* Samsung S10 smartphone embedded with a pair of loudspeaker and microphone, *ii)* Raspberry Pi 3B+ connected with a ReSpeaker 4-mic linear microphone array and JBL Jemebe loudspeaker, and *iii)* Laptop (DELL XPS 15 9500) embedded with a MEMS microphone (InvenSense ICS-40730) and a loudspeaker (*i.e.,* JBL Jemebe), as shown in Fig. 14(a). Note that ReSpeaker is equipped with a 4-mic linear microphone array, and we employ one of the microphones for receiving signals. Acoustic signals are transmitted and received at a sampling rate of 48 kHz with a frequency band ranging from 19 to 21 kHz, which is inaudible to most people [28]. The acoustic signals received by the smartphone and MEMS microphone will be sent to the laptop through WiFi for further processing using MATLAB. For the ReSpeaker, the processing is done on Raspberry Pi using Python. We collect the ground truth of respiration signals using the Vernier respiration belt [29]. The ground truth of heartbeat signals is captured by a 3-lead Electrocardiogram (ECG) monitor (Heal Force PC-80B) [30], which is intended for measuring and recording the ECG waveform and heartbeat patterns.

7 EVALUATION

7.1 Experimental Setup

We evaluate the system's performance in detecting heartbeat and respiration rates in various scenarios. Unless otherwise specified, we use the Raspberry Pi 3B+ connected with a ReSpeaker 4-mic linear microphone array and JBL Jemebe loudspeaker. The transceivers are placed close to each other so that the path length is about twice the distance from the target to the transceiver. A subject keeps relatively still throughout the experiment where small body movements (leg shaking and finger tapping) are allowed. We conduct experiments in a room with 7 m $\times$ 7 m. The environmental noise level stays at 40 dBA for frequencies below 5000 Hz and 30 dBA for frequencies ranging from 15 kHz to 22 kHz. We keep the speaker power at 1.035 W, which is about the lower bound of the standard power consumption of the loudspeaker [31]. This ensures that sensing range improvement is not brought trivially by increased speaker power. The transmitted sound is 55 dBA at the distance of 5 m,



(a) Devices (b) Scenario

Fig. 14. Experimental setup.

which is below the safety threshold of 85 dBA specified in Occupational Noise Exposure by the National Institute for Occupational Safety and Health (NIOSH) [32].

To evaluate LoEar, we recruit 15 volunteers (7 females and 8 males) with ages from 21 to 56 years old and weights in the range of 49 to 88 Kg. Fig. 12 summarizes their ages, genders, and somatotypes characterized by Body Mass Index (BMI). Note that all subjects are indexed by the BMI value in an increasing order, and the indexes are put into the marker. In prior to the experiments, we have a ten-minute briefing session about the testing equipment and data collection. All experiments are conducted upon approval from the institutional review board (IRB) at our institute. To reduce the impact of subject diversity, we collect data for each experiment as follows. Each data collection runs for 2 minutes and repeats for 10 times. Unless other specified, subjects are asked to sit 3 m away from the transceiver, face the transceiver, and breathe normally, as shown in Fig. 14(b).

7.2 Experimental Results

7.2.1 Overall Performance. In this section, we evaluate the overall performance of our system in monitoring vital signs.

Distance: We explore the effect of distance between subject and transceiver on vital sign monitoring. As shown in Fig. 15, the system achieves a relatively small error in both respiration and heartbeat rates at a distance ranging from 3 m to 6.5 m. At longer distances (*i.e.,* $\geq$ 7 m), the accuracy of the heartbeat rate decreases significantly, but the respiration rate achieves a reasonable accuracy. In general, our system is able to measure the heartbeat rate with negligible error at a distance of 6.5 m. Note that we do not test the maximum distance for respiration due to the space constraint.

Path length estimation error: In this experiment, we evaluate the accuracy of path length estimation between the target user and transceiver at different distances from 3 m to 7 m. We first measure the path length with a ruler for the ground truth. We then compare it with the estimation measured by *Continuous-MUSIC* to derive the estimation error. We repeat this experiment for each subject 500 times at different distances relative to the transceiver from 3 m to 7 m. The final cumulative distribution function is shown in Fig. 16. The results show that *Continuous-MUSIC* achieves a median error of 0.029 m and a maximum error of 0.11 m. This shows that *Continuous-MUSIC* can estimate the path of the target accurately.

Device type: In this experiment, we verify the generality of LoEar on different acoustic devices, including smartphone, Raspberry Pi and laptop, as mentioned in Section 7.1. As shown in Fig. 17, "ics", "ras", and "pho" represent the laptop setup, Raspberry Pi setup, and smartphone setup, respectively. In both Raspberry Pi and laptop setups, we achieve a high accuracy in monitoring vital signs when the subject is at different distances. In

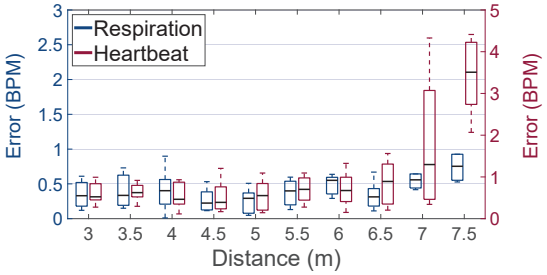


Fig. 15. User distance.

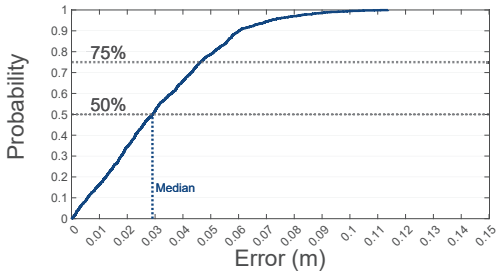


Fig. 16. Path length estimation error.

comparison, the smartphone can monitor heartbeat and respiration when the subject is in the only range of 1 m and 3 m, respectively, mainly because of the lower sound volume of the loudspeaker. Besides, another underlying reason may be the smaller distance between the speaker and microphone in a smartphone. The much smaller distance (*e.g.,* about 1 cm for Samsung S10) will bring about a stronger LOS path, resulting in much lower SNR and, thus, weaker capability in monitoring vital signs. Although the performance of using a smartphone is worse than the other two devices, it is sufficient to meet our daily requirements since the smartphone is usually near users.

Benchmarking: We now compare the performance of LoEar with the state-of-the-art acoustic approaches, *i.e.,* FMCW-based [6, 7] and OFDM-based systems [14, 15]. For a fair comparison, we deploy all comparative systems on the same transceiver consisting of the ReSpeaker and smart speaker. Fig. 18 shows that LoEar outperforms other approaches in terms of sensing range and the resolution. This illustrates the great usability of our system in daily life, as most acoustic-based devices (*i.e.,* televisions and smart speakers) are at a fixed position in the room.

7.2.2 Evaluation on User Issues. In this section, we evaluate the impact of user-related factors on the performance of our system in monitoring vital signs.

Spacing: In this experiment, we evaluate the effect of the spacing between two subjects on vital sign monitoring. The spacing of two users is denoted as the difference in the distance of the two users from the transceiver. In the beginning, we ask a subject to sit at 3 m and 0 ° relative to the transceiver while another subject sits at 3.2 m and 10 ° relative to the transceiver, as shown in Fig. 19. The spacing of these two subjects is regarded as 20 cm, leading to a path length difference of about 40 cm (*i.e.,* nearly twice the spacing). The second subject was then asked to change the spacing by moving away along the line at 10 ° relative to the transceiver. As shown in Fig. 20, LoEar monitors vital signs by increasing the spacing from 20 to 100 cm with a fixed step of 20 cm. The result shows that the BPM error of monitored vital signs increases in the presence of another subject. Specially, the monitoring performance has a severe degradation when the spacing is 20 cm. This is reasonable due to the destructive superposition of reflected acoustic signals from two subjects, resulting in interference from each other. Fortunately, the attenuation of performance was effectively alleviated with the spacing of greater than 40 cm. For example, LoEar achieves median errors of 0.81 BMP and 0.85 BMP for respiration and heartbeat when a spacing is 40 cm, respectively. When the spacing is greater than 60 cm, the performance tends to be stable, which is comparable to the previous standard scenario.

Orientation: We also tested the system performance when subjects had different orientations (facing directions). These measurements were conducted when the subject was at a distance of 5 m from the transceiver. To avoid the impact of other factors, we asked all users to wear the same shirt and keep the same sitting posture. The subject changed his orientation towards the transceiver (*i.e.,* from 0 ° to 90 °) during this experiment. Results in Fig. 21 show that the respiration recovery is more immune to user orientation compared to the heartbeat. This is

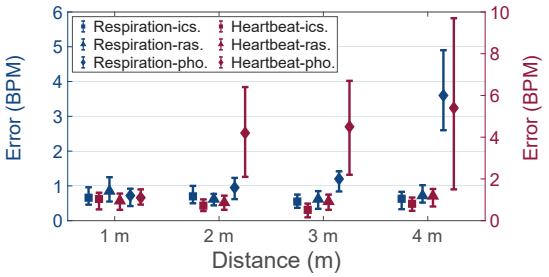


Fig. 17. Device type.

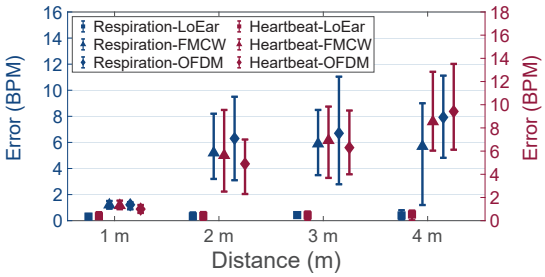


Fig. 18. Benchmarking.

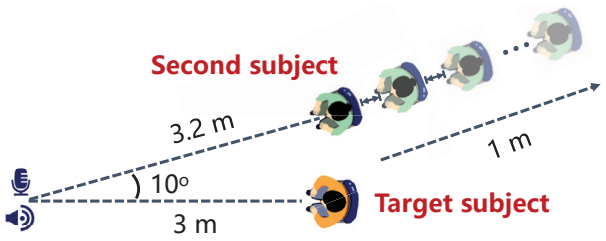


Fig. 19. Different spacings between two subjects.

reasonable since the chest displacement caused by the respiration is larger than the heartbeat from the side view. Furthermore, we observe that LoEar achieves similarly excellent performance in monitoring both respiration and heartbeat when the orientation is in the range from 0 to 15 °. This indicates that the orientation has little impact on the vital signs monitoring when it is below 15 °. However, this effect increases as the orientation is greater than 15 ° and leads to the unacceptable BPM error when it is beyond 60 °. Overall, our system has good orientation tolerance of up to 120 ° (*i.e.,* −60 ∼ 60 °) according to symmetry.

Number of users: To explore the capacity of LoEar in monitoring vital signs in the presence of multiple users, we ask multiple subjects to sit at different locations in the room at the same time and breathe normally, as shown in Fig. 22. To reduce the interference, the spacing of two adjacent subjects is set to 60 cm based on the observation in Fig. 19. Additionally, the angle difference of two adjacent subjects relative to the transceiver is set to 10 ° to ensure there is no occlusion between each other. To further study the effect of user number, we asked a varying number of subjects to sit at different locations randomly and then measured all subjects' respiration and heart rhythms. Fig. 23 shows that LoEar supports simultaneous measurements of up to 4 subjects. Although the heartbeat measurements exhibit a median error of 1.39 BPM for 4 subjects, it is sufficient to satisfy daily monitoring requirements. This result demonstrates that our system is capable of supporting room-level respiration and heartbeat measurements with multiple users.

Clothing: In this experiment, we evaluate the performance when subjects wear different types of clothes. We ask a subject to repeat the experiments at a distance of 3 m from the transceiver, wearing different clothes. As shown in Fig. 24, LoEar achieves a reasonable accuracy for heartbeat monitoring for most types of clothes (*i.e.,* coat, T-shirt, shirt) except sweater. This is probably because the sweater is thick, resulting in signal blocking.

Diversity: Different subjects are diverse in age, gender, and BMI, which may affect the characteristics of the vital signs. Thus, it is necessary to explore the impact of human diversity on system performance. In this experiment, we report the monitoring accuracy of each subject, which corresponds to the index in Fig. 12, respectively, as

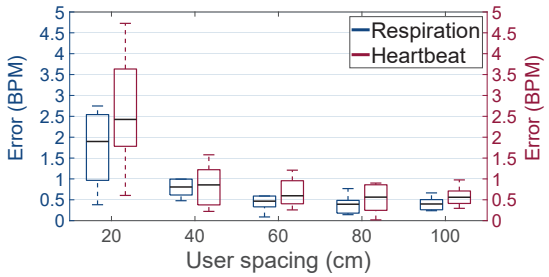


Fig. 20. User spacing.

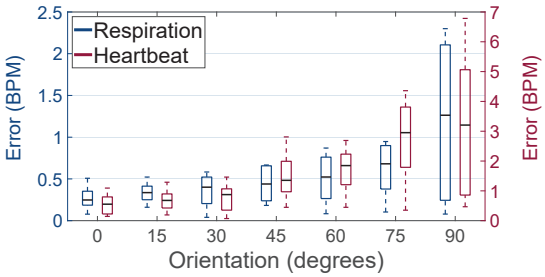


Fig. 21. User orientation.

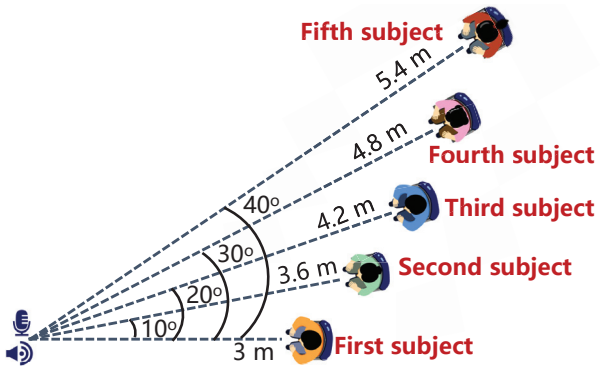


Fig. 22. Specific positions for multiple users.

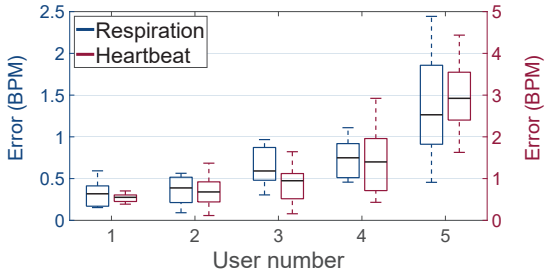


Fig. 23. User number.

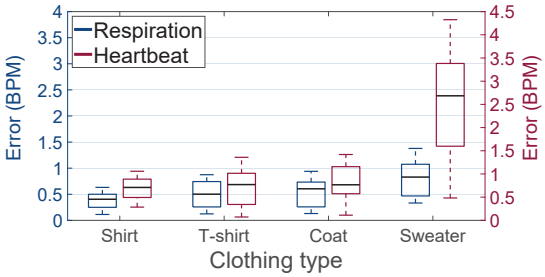


Fig. 24. User clothing.

shown in Fig. 25. We observe that gender and age have no significant influence on the performance in monitoring the vital signs. In comparison, the value of BMI is highly correlated to the performance in monitoring heartbeat. Specifically, we observe that the error in monitoring heartbeat has an evident increasing trend with the increased value of BMI while there is no significant pattern in respiration monitoring. This is reasonable since the larger BMI will make the thoracic muscle thicker to block the weak heartbeat signals. In contrast, the respiration, which causes displacement of the chest, is not affected. LoEar still achieves a median heart rhythm error of 0.83 BPM for the overweight subject (indexed as 15) [33], sufficient for daily life. Overall, the performance remains robust on different subjects, demonstrating the system's generality in measuring vital signs.

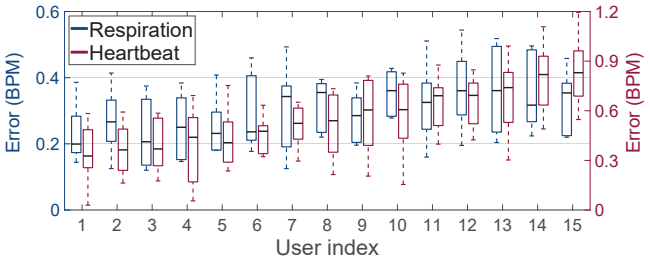


Fig. 25. User diversity.

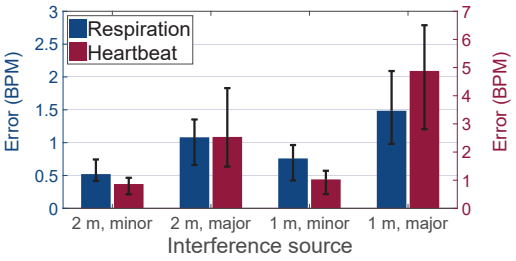


Fig. 26. Nearby human movements.

7.2.3 Evaluation on Noise Issues. In this section, we evaluate the impact of noise-related factors on the performance of our system in monitoring vital signs.

Movements: In this experiment, we evaluate the impact of nearby human movements on the performance of LoEar. We ask a subject to take the measurement while another interfering subject performs subtle (*i.e.,* shaking legs) and large (*i.e.,* walking back and forth) movements as an interference source. Similar to Fig. 19, the interfering subject is located at the spacing of 1 m and 2 m from the target subject (*i.e.,* path lengths of about 0.98 m and 1.98 m from the target subject), respectively. As shown in Fig. 26, the respiration rate stays robust when a major interference source is 1 m away from the target subject. However, major movement brings a significant effect on the heartbeat rate measurement at both locations. This is because the heartbeat motion is only mm-level, which is easily drowned out by large human movement nearby. LoEar identifies the period with only subtle movements and uses this period to measure vital signs with high accuracy.

Ambient noise: We conducted several experiments to reveal how ambient noise (*i.e.,* human voice, background music, and television sound) affect the reconstructed signal. The noise sources are placed 3 m from the transceiver and make sure the sound at the ears is as strong as it is commonly heard in daily scenarios. Fig. 27 shows that the performance has little degradation under common ambient noise. This demonstrates that LoEar is capable of working under daily indoor scenarios with common ambient noise sources.

Appliance interference: Besides ambient noise sources, we also test the performance of the system when there is interference from other appliances. Specifically, we select four appliances (*i.e.,* washer, microwave oven, fan, and air conditioner (AC) that produce mechanical vibration and thus mechanical waves that can interfere with our transmitted signal. The user is at 3 m from the transceiver, and the appliances are placed 5 m from the transceiver and turned on throughout the data recording period. The result in Fig. 28 shows that the washer and the oven have little influence on the heartbeat and respiration monitoring accuracy while the performance has a slight degradation when using fan and AC. This is because these two appliances accelerate airflow around the user and consequently blur peaks in the recovered signal.

7.2.4 Performance Comparison. In this section, we compare the performance LoEar with LASense [34] in vital signs monitoring. We use several metrics including the accuracy at various user-device distances, the impact of dynamic environment, and generalizability.

Accuracy: To ensure a fair comparison, we evaluate LASense and LoEar when monitoring both respiration and heartbeat in settings where the user is asked to sit from 1 m to 7 m in front of the transceiver at a step of 2 m. Fig. 29 shows that LASense achieves a low median error of 0.41 BPM for respiration monitoring when the user is 5 m away from the transceiver while dropping to 0.82 BPM as the distance increases to 7 m. On the other hand, LASense achieves a low median error of 0.42 BPM for heartbeat monitoring with the user-distance of 1 m, and the median error substantially increases to 1.12 BPM with the user-distance of 3 m. In comparison, LoEar

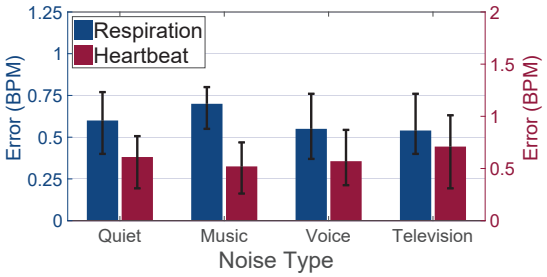


Fig. 27. Ambient noise.

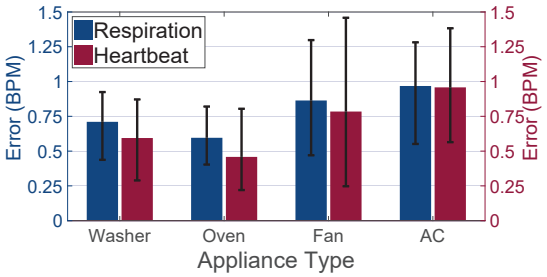


Fig. 28. Appliance interference.

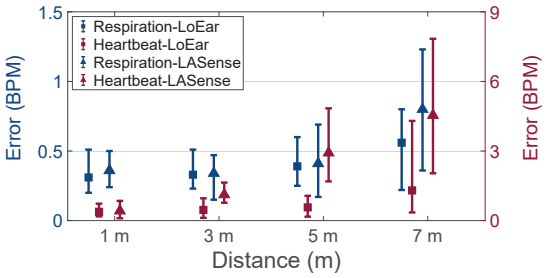


Fig. 29. Comparison with various distances.

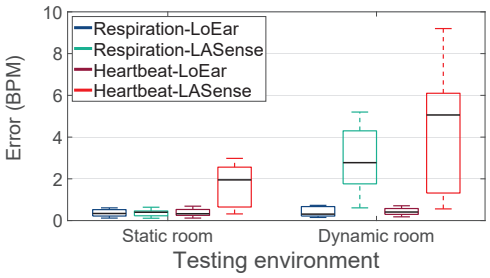


Fig. 30. Comparison under dynamic environment.

still achieves much lower median errors of 0.56 BPM and 1.3 BPM for respiration and heartbeat monitoring, respectively, even with the user-device distance of 7 m.

Impact of dynamic environment: To intuitively evaluate the impact of surrounding dynamics on both systems, we conduct comparisons in both static and dynamic environments. In this experiment, we use four types of household appliances (*i.e.,* washer, microwave oven, fan, and AC). All appliances are in the power-off condition for the static environment while in the running condition for the dynamic environment. As illustrated in Fig. 30, LASense achieves a low median error of 0.41 BPM for respiration monitoring in the static environment while substantially increasing to 2.78 BPM in the dynamic environment, even worse for heartbeat monitoring. In comparison, LoEar can robustly monitor respiration and heartbeat with high accuracy in both environments. Compared with LASense, the robustness of LoEar in the dynamic environment mainly benefits from the target user identification process, which focuses on identifying the target user from multiple subjects with motions based on the quasi-period of respiration, as discussed in Section 4.2.

Generalizability: We further compare LoEar with LASense in terms of monitoring range, user detection rate, bandwidth, and filtering, as listed in Tab. 1, to demonstrate generalizability. Not only is LoEar far superior to LASense in terms of monitoring range, but in a dynamic environment, LoEar can identify users with higher accuracy. Beyond that, LoEar is implemented with a bandwidth of only 2 kHz, while LASense requires 4 kHz, which indicates that LoEar is easier to be deployed on bandwidth-limited speakers. The enhanced signal in LASense has to undergo a filtering step to smooth the raw signals for further respiration monitoring. The additional filtering may increase the computation cost and filter out some valuable information related to vital movement. In contrast, the enhanced signal in LoEar is sufficiently strong to monitor vital signs so that no additional filtering process is required. Overall, LoEar is more generalizable than LASense for real-world deployment.

Table 1. Overall comparison between LASense and LoEar.

System	Respiration range	Heartbeat range	User detection rate	Band	Need Filtering
LASense	6 m	3 m	23.2 %	4 kHz	Yes
LoEar	7 m	6.5 m	95.8%	2 kHz	No

Table 2. Time consumption.

Platform	CFR measurment	Continuous-MUSIC	Carrierforming	Rhythm Estimation	Total
DELL XPS 15 9500	4.13 ms	10.69 ms	4.15 ms	1.35 ms	20.32 ms
Raspberry Pi 3B+	6.06 ms	16.23 ms	6.12 ms	2.26 ms	30.67 ms

7.2.5 Processing Latency. This section provides an evaluation on the processing time of key steps used in LoEar, including CFR measurement, *Continuous-MUSIC*, *Carrierforming*, and Rhythm Estimation.

We measure the processing time for one frame acoustic signal on different devices, including the laptop (DELL XPS 15 9500) and Raspberry Pi 3B+. The procedure of LoEar includes four serial components: CFR measurement, *Continuous-MUSIC*, *Carrierforming*, and Rhythm Estimation. Since the duration of each audio frame signal is 40 ms under the sampling rate of 48 kHz, the processing time for each frame should be shorter than 40 ms to enable the real-time processing. To avoid random error, we measure the average time consumption for 1500 frames in 1 minute. As shown in Tab. 2, LoEar achieves total latency of 20.32 ms and 30.67 ms for the laptop and Raspberry Pi, respectively. The responsiveness on the two platforms is sufficient to meet real-time processing requirement. Overall, our system design is light-weighted so that it can be easily deployed on various smart-home platforms.

8 RELATED WORK

Recent work related to LoEar is in the following categories.

Low-resolution Acoustic Sensing: A lot of research efforts have been made to utilize acoustic signals to sense hand gestures. Lazik *et al.* [35] present an indoor ultrasonic location tracking system with a localization error of around 1 m. AAMouse [36] uses the Doppler shift to estimate the velocity and the hand moving distance with the error of 1.4 cm. Multiwave [37] leverages the Doppler effect to translate hand movements into user interface commands. AudioGest [38] and Soundwave [39] present a device-free system that can sense the hand in-air movement around user's devices. Recent works [40, 41] enable highly accurate hand gesture recognition based on the deep learning technique. Beyond sensing hand gestures, significant works have focused on sensing large-scale movements. CovertBand [42] tracks individuals' locations and daily activities both within a room using acoustic signals. AcousticID [43] demonstrates the feasibility of gait recognition by analyzing the Doppler effect of various body parts on acoustic signals while walking. All these schemes can only sense low-resolution human activities, while our system can sense high-resolution activities, indicating broader use in real life.

High-resolution Acoustic Sensing: Due to the relatively high bandwidth, acoustic signals are superior in sensing subtle movements. Many prior works have been devoted to subtle motion sensing using acoustic signals. FingerIO [14] tracks subtle finger motion using the reflected acoustic signals. LLAP [17] senses the finger movement direction and distance by using the phase changes of the sound signals with a tracking accuracy of 3.5

mm. CAT [44] develops a FMCW system achieving an accuracy of 5 mm. Sun *et al.* [45] propose an acoustic-based tapping scheme to detect finger tapping movement with high accuracy. Strata [46] deploys a fine-grained finger tracking system by selecting the corresponding channel tap. In addition to finger gesture sensing, there are many significant works on sensing subtle vital signals, comprising respiration [3–5] and heartbeat [6, 7]. However, all these sensing systems work at a short range. Instead, our system intends to push the limited range of vital sign monitoring to the room scale using commodity acoustic devices.

Long-range Acoustic Sensing: There have been a few attempts to extend the acoustic sensing range. FM-Track [11] enables tracking of hand-sized targets with an error of 4 cm. Mao *et al.* [10] realize hand motion tracking within 4.5 m and achieves $1.2 - 3.7$ cm error. DeepRange [13] develops a ranging system with single speaker and microphone, and the distance estimation is within 1 cm at 4 m. These approaches either require multiple microphones or rely on deep learning techniques for long-range sensing. Not only that, the ranging resolution is centimeter-level, which can not be applied to sense subtle motions. In recent pioneer work parallel with us, Li *et al.* [34] proposed a new system called LASense to increase the range for sensing respiration to 6 m using a single speaker and microphone. In comparison, our system is capable of monitoring both respiration and heartbeat with high accuracy in the range of 7 m and 6.5 m, respectively, using a single speaker and microphone.

9 LIMITATIONS AND DISCUSSIONS

We now briefly discuss the limitations of LoEar and our future work as follows.

First, LoEar can simultaneously enhance multi-subject reflection on different paths. However, when the paths of two subjects are close in length, severe interference between subjects occurs, leading to failure in differentiation and further enhancement. This usually happens when multiple people sit in front of the device side by side. In this scenario, they have paths of similar length even though their AOAs relative to the device are different. Obviously, LoEar fails to differentiate subjects in this scenario. In our future work, we will in-depth explore this issue from two perspectives. For one thing, we consider modeling the multi-user separation problem as a blind source separation (BSS) problem. Since the reflected signals of multiple users are linearly mixed at different subcarriers and concurrently independent of each other, we can first take the CFR of multiple subcarriers as the multiple references respectively (*i.e.,* CFR_1 in Eq. (2)) and derive the corresponding enhancements at multiple subcarriers. Then, we can apply the independent component analysis (ICA) method to separate the signals of each user from the enhancements at multiple subcarriers. For another, we plan to use microphone arrays available at the receiver to separate signals from multiple users at different locations in 2D space, even though they have the same path lengths.

Second, *Carrierforming* also works in a relatively static environment, where the significantly environmental dynamics should be far from the target. For example, when another person is walking around the target (*e.g.,* ≤ 1 m), it is practically impossible to sense the subtle motions of the target. This is mainly because that although the drastic interference is destructively reduced in the *Carrierforming* model, it is still so strong that it dominates the CFR measurements. One possible solution is to enhance the target's reflections in the time domain so that the enhanced reflection can be separated from surrounding reflections, and we leave it for our future work.

Third, LoEar works well when the target user keeps relatively still with small movements, such as limb-shaking and finger tapping. However, LoEar may fail when the user is walking. There are two reasons for this. First, the length of the user path in motion varies largely, making it difficult to track the real-time TOF with *Continuous-MUSIC*, leading to failure in enhancement. Second, movements in other body parts during walking will seriously interfere with the tiny vital signals. For our future work, we consider using machine learning schemes, such as deep contrastive learning [47] and variational encoder network [48], to refine vital signs under large body movements.

10 CONCLUSION

In this paper, we developed LoEar, an acoustic-based sensing system that pushes the acoustic sensing range for vital signs monitoring. We focus on increasing the high-resolution sensing range using a single microphone and speaker without applying machine learning. We propose a systematic approach to address the challenges associated with high-resolution, long-range acoustic sensing. We demonstrate that LoEar has a strong capacity for sensing subtle motions in different scenarios. We believe the system will significantly benefit many long-range acoustic sensing applications.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (Grant No. 62102006). This work is also supported by the National Natural Science Foundation of China A3 Foresight Program (No.62061146001), in part by National Natural Science Foundation of China (Grant No. 61802007, 62022005, 12071460, 62172394, 62061146001) and the Youth Innovation Promotion Association, Chinese Academy of Sciences (No. 2020109).

REFERENCES

[1] Chao Cai, Zhe Chen, Henglin Pu, Liyuan Ye, Menglan Hu, and Jun Luo. Acute: Acoustic thermometer empowered by a single smartphone. In *Proceedings of ACM SenSys*, 2020.
[2] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of ACM MobiSys*, 2015.
[3] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. Contactless infant monitoring using white noise. In *Proceedings of MobiCom*, 2019.
[4] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. C-fmcw based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–20, 2018.
[5] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. Spirosonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of ACM MobiCom*, 2020.
[6] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *Proceedings of IEEE INFOCOM*, 2018.
[7] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. Your smart speaker can" hear" your heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–24, 2020.
[8] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–26, 2019.
[9] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. In *Proceedings of ACM MobiCom*, 2016.
[10] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. Rnn-based room scale hand motion tracking. In *Proceedings of ACM MobiCom*, 2019.
[11] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. Fm-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of ACM SenSys*, 2020.
[12] Anup Agarwal, Mohit Jain, Pratyush Kumar, and Shwetak Patel. Opportunistic sensing with mic arrays on smart speakers for distal interaction and exercise tracking. In *Proceedings of IEEE ICASSP*, 2018.
[13] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. Deeprange: Acoustic ranging via deep learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–23, 2020.
[14] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of ACM CHI*, 2016.
[15] Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. Resptracker: Multi-user room-scale respiration tracking with commercial acoustic devices. In *Proceedings of IEEE INFOCOM*, 2021.
[16] Lei Wang, Xiang Zhang, Yuanshuang Jiang, Yong Zhang, Chenren Xu, Ruiyang Gao, and Daqing Zhang. Watching your phone's back: Gesture recognition by sensing acoustical structure-borne propagation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–26, 2021.
[17] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of ACM MobiCom*, 2016.
[18] Kanji Ono. A comprehensive report on ultrasonic attenuation of engineering materials, including metals, ceramics, polymers, fiber-reinforced composites, wood, and rocks. *Applied Sciences*, 10(7):2230, 2020.

[19] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In *Proceedings of ACM SIGCOMM*, 2015.
[20] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *Proceedings of USENIX NSDI*, 2013.
[21] Xiang Li, Shengjie Li, Daqing Zhang, Jie Xiong, Yasha Wang, and Hong Mei. Dynamic-music: Accurate device-free indoor localization. In *Proceedings of ACM UbiComp*, 2016.
[22] C Lowanichkiattikul, M Dhanachai, C Sitathanee, S Khachonkham, and P Khaothong. Impact of chest wall motion caused by respiration in adjuvant radiotherapy for postoperative breast cancer patients. *SpringerPlus*, 5(1):1–8, 2016.
[23] Young K Jang and Joe F Chicharo. Adaptive iir comb filter for harmonic signal cancellation. *International Journal of Electronics Theoretical and Experimental*, 75(2):241–250, 1993.
[24] Gary G Berntson, J Thomas Bigger Jr, Dwain L Eckberg, Paul Grossman, Peter G Kaufmann, Marek Malik, Haikady N Nagaraja, Stephen W Porges, J Philip Saul, Peter H Stone, et al. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34(6):623–648, 1997.
[25] Reham Mohamed and Moustafa Youssef. Heartsense: Ubiquitous accurate multi-modal fusion-based heart rate estimation using smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–18, 2017.
[26] Federica Landreani, Mattia Morri, Alba Martin-Yebra, Claudia Casellato, Esteban Pavan, Carlo Frigo, and Enrico G Caiani. Ultra-short-term heart rate variability analysis on accelerometric signals from mobile phone. In *Proceedings of IEEE EHB*, 2011.
[27] Carolyn Jarvis. *Physical Examination and Health Assessment E-Book*. Elsevier Health Sciences, 2019.
[28] A Rodríguez Valiente, A Trinidad, JR García Berrocal, C Górriz, and R Ramírez Camacho. Extended high-frequency (9–20 khz) audiometry reference thresholds in 645 healthy subjects. *International journal of audiology*, 53(8):531–545, 2014.
[29] https://www.vernier.com/product/go-direct-respiration-belt/.
[30] http://www.healforce.com/en/html/products/portableecgmonitors/healthcare-equipment-portable-ECG-monitors-PC-80B.html.
[31] https://www.pdf-manuals.com/pdf/jbl-jembe-wireless-speakers-jbljembebtam-b-h-photo-video-240865-user-manual.pdf.
[32] https://www.cdc.gov/niosh/docs/98-126/pdfs/98-126.pdf?id=10.26616/NIOSHPUB98126.
[33] Cora E Lewis, Kathleen M McTigue, Lora E Burke, Paul Poirier, Robert H Eckel, Barbara V Howard, David B Allison, Shiriki Kumanyika, and F Xavier Pi-Sunyer. Mortality, health outcomes, and body mass index in the overweight range: a science advisory from the american heart association. *Circulation*, 119(25):3263–3271, 2009.
[34] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. Lasense: Pushing the limits of fine-grained activity sensing using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–27, 2022.
[35] Patrick Lazik and Anthony Rowe. Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In *Proceedings of ACM SenSys*, 2012.
[36] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proceedings of ACM MobiSys*, 2015.
[37] Corey R Pittman and Joseph J LaViola Jr. Multiwave: Complex hand gesture recognition using the doppler effect. In *Proceedings of ACM GI*, 2017.
[38] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. Audiogest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of ACM UbiComp*, 2016.
[39] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: Using the doppler effect to sense gestures. In *Proceedings of ACM CHI*, 2012.
[40] Kang Ling, Haipeng Dai, Yuntang Liu, Alex X Liu, Wei Wang, and Qing Gu. Ultragesture: Fine-grained gesture sensing and recognition. *IEEE Transactions on Mobile Computing*, 2020.
[41] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing*, 2020.
[42] Rajalakshmi Nandakumar, Alex Takakuwa, Tadayoshi Kohno, and Shyamnath Gollakota. Covertband: Activity information leakage using music. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–24, 2017.
[43] Wei Xu, ZhiWen Yu, Zhu Wang, Bin Guo, and Qi Han. Acousticid: gait-based human identification using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–25, 2019.
[44] Wenguang Mao, Jian He, and Lili Qiu. Cat: High-precision acoustic motion tracking. In *Proceedings of MobiCom*, 2016.
[45] Ke Sun, Wei Wang, Alex X Liu, and Haipeng Dai. Depth aware finger tapping on virtual displays. In *Proceedings of ACM MobiSys*, 2018.
[46] Sangki Yun, Yichao Chen, Huihuang Zhang, Lili Qiu, and Wenguang Mao. Strata: Fined-grained device-free tracking using acoustic signals. In *Proceedings of ACM MobiSys*, 2017.
[47] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. Movi-fi: motion-robust vital signs waveform recovery via deep interpreted rf sensing. In *Proceedings of SCM MobiCom*, 2021.
[48] Tianyue Zheng, Zhe Chen, Shujie Zhang, Chao Cai, and Jun Luo. More-fi: Motion-robust and fine-grained respiration monitoring via deep-learning uwb radar. In *Proceedings of ACM MobiCom*, 2021.