LoEar: Push the Range Limit of Acoustic Sensing

Anonymous Author(s)

ABSTRACT

Acoustic-based sensing has been currently used in numerous applications with the wide-deployment of acoustic-enabled devices. However, most acoustic-based sensing systems can only work in the limited range, leading to serious hindrance to the application in real-world environments. The main reason for the limited sensing range is the fast attenuation of acoustic signals. In this paper, we present a novel high-resolution sensing system called LoEar, which is the first to focus on increasing the sensing range greatly using only single microphone and speaker without considering machine learning technologies. To achieve this, we first develop a novel model, namely *Carrierforming*, to enhance the SNR on the target path. Second, we propose a new technique called *Continuous-MUSIC* to identify the target with subtle motions to enable *Carrierforming*. Finally, we extend the LoEar to 2D scenario to sense multiple objects' subtle motions when they are at different locations. To demonstrate the capacity of LoEar, we choose the respiration and heartbeat as the case study. For the respiration, our system increases the sensing range from the state-of-the-art 2 *m* to 7 *m* using the commercial devices. For the heartbeat, our system increases the sensing range from the state-of-the-art 1.2 *m* to 5 *m*. Additionally, our system has the capacity of detecting multiple static users with different locations in 2D space and monitoring their respiration and heartbeat simultaneously.

1 INTRODUCTION

Recently, acoustic sensing has become a prevailing research topic in the mobile sensing community. Benefiting from its high sensing granularity [1] and the fact that the sensing hardware (*i.e.,* speakers and microphones) is already widely deployed in today's mobile devices, many application-driven acoustic-based sensing techniques have been proposed for finger tracking [2–8], gesture recognition [9–12], respiration monitoring [13–16], and heartbeat monitoring [17, 18]. While most of these proposals have demonstrated the efficacy in their domain-specific sensing accuracy, they are commonly in short of in one important metric: limited sensing range. Constrained by the nature of its fast signal strength decay over distance, the effective sensing range is limited by ** m in **, ** and ** respectively. This limited range, however, can significantly constrain the practicality of acoustic sensing application from large-scale deployment, especially for room-level sensing with limited number of sensing device

Figure 1: Ultrasound propagation paths from the transceiver to the target in the indoor environment. All paths can be classified into direct paths and indirect paths. The direct path corresponds to the direct reflection from only the target, and the indirect path corresponds to the indirect reflection from the target via some static objects such as desk and wall.TODO: Multiple people. Show signal map with peak location enhancement. Two scenarios.

with no coordinated sensing capability yet. While beamforming [7, 11, 19] and deep learning [6, 7] have shown their success in extending the sensing range from signal quality improvement and interference mitigation perspectives, they are not compatible with COTS device with its limited power budget and number of acoustic transceivers.

In this paper, we propose *Carrierforming*, a novel signal processing technique dedicated to extending acoustic sensing range with a single pair of speaker and microphone. The *Carrierforming* design leverages multi-subcarrier signal (*i.e.,* OFDM or its variants) as a basis. At a high level, idea of *Carrierforming* is rather simple: by observing the pre-determined phase difference between the subcarrier, it strategically delays the phase across multiple subcarriers and align the peak of them for SNR enhancement at a specific location (unlike a direction in beamforming). Analytic results show that one it improves the SNR by 29x, by using 80 subcarriers with 25 Hz spacing under a 2 KHz bandwidth, for example. Using vital sign monitoring as a representative human-centric acoustic sensing example, *Carrierforming* can substantially extend the sensing range from sub-1.2m to beyond 5 m with a single transceiver, making such applications much more easy to deploy in typical room-scale spaces including home, office and hospital.

There is an important precondition when applying *Carrierforming* in sensing applications – one needs to detect and locate the subject-of-interest(s) first, in terms of ranging and

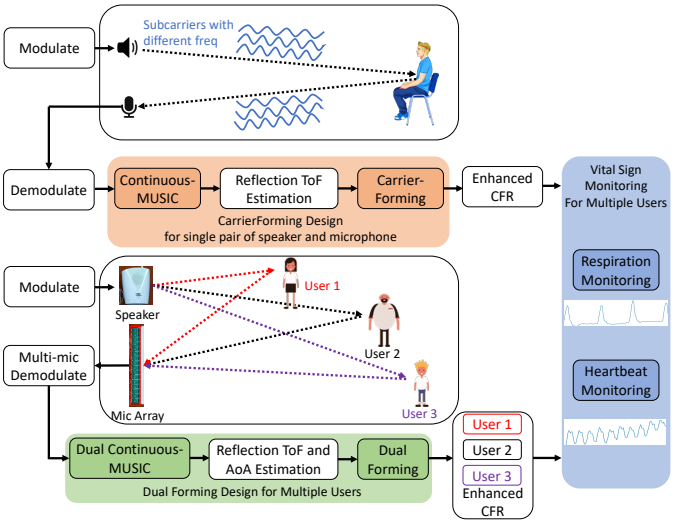


Figure 2: System overview of LoEar.

orientation determination. In fact, there are two practical challenges towards this goal. First, one needs to detect and differentiate the (weak) subject's reflection signal from other ambient reflection before the target reflection signal is enhanced. Second, there could be multiple subjects in the same space, and ambient reflection.

To address this challenge, we propose a new technique, called *Continuous-MUSIC*, to detect the target path even when the target with subtle motion is far away from the device. This technique is inspired by the MUSIC algorithm. Comparing to the standard MUSIC algorithm which is based on the same moment , our scheme is founded on received signals from multiple continuous timing points. With the dynamic signals from continuous frames, the rank of the correlation matrix in the MUSIC algorithm will be increased so as to detect multiple subjects. Meanwhile, the environmental dynamics only consists of the subjects' subtle motions. Thus, the subtle motions will be highlighted in this method. In this way, the new scheme has two superior advantages. Firstly, it can distinguish multiple objects simultaneously. Secondly, it can distinguish the target with subtle motions (e.g., respiration, heartbeat) from other static reflections such as desk, wall, and the floor.

So far, our system can simultaneously enhance the multiple subject reflection on different paths. However, when the paths of two subjects are close in length, there will be severe interference between the subjects leading to the failure in the differentiation and further enhancement. This usually happens when multiple people sit in front of the device side by side. In this scenario, they have the paths of similar length even though their Angle of Arrival (AOA) relative to the device are different. Obviously, *Continuous-MUSIC* fails to differentiate subjects in this scenario. Thus, the third challenge is to differentiate multiple subjects and further enhance the corresponding SNR.

To address this challenge, we extend the *Continuous-MUSIC* technique to *Dual Continuous-MUSIC* technique and propose *Dual-Forming* model by combining multiple subcarriers and microphones to differentiate multiple subjects with subtle motions when they are at different locations in 2D space and enhance the corresponding SNR, respectively. Furthermore, we make a depth analysis on the influence factors, including the microphone array width and microphone spacing, in the enhanced model. Note that they are closely related to the lowest and highest frequency in OFDM signals, respectively.

By significantly enhancing the SNR, LoEar can sense the high-resolution subtle movement. To demonstrate the sensing capacity, we choose the vital signals, including respiration and heartbeat, as a case study. For the respiration, our system increases the sensing range from the state-of-the-art 2 m to 7 m. For the heartbeat, our system increases the sensing range from the state-of-the-art 1.2 m to 5 m. As the heartbeat signal is a sub-millimeter level movement ($0.1-0.5$ mm), our system significantly pushes the sub-millimeter level sensing range from 1.2 m to the room-scale. Additionally, our system has the capacity of identifying multiple static users with different locations in 2D space and monitoring their respiration and heartbeat simultaneously.

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

- To our knowledge, our high-resolution sensing system is the first to focus on increasing the sensing range greatly using only single microphone and speaker without considering machine learning technologies.
- We propose a new model to identify the target path even when the target with subtle motion is far away from the device.
- We enable the system in 2D scenario, in which multiple users with subtle motions can be clearly separated and the reflections can be further enhanced.

2 SYSTEM OVERVIEW

Figure 2 illustrates the system architecture of LoEar. LoEar transmits the OFDM-modulated acoustic signals using the loudspeaker and then records and demodulates the reflected sound by microphones to obtain the channel frequency response (CFR) measurements (Section 3). The design goal of LoEar is to push the range limit of acoustic ranging, which needs to achieve three requirements.

(i) Long-range detection (Section 5). LoEar is able to detect the subtle movement of target, *i.e.,* , human vital signs, within 5 meter distance by using a single pair of speaker and

microphone. We propose a new technique, called *Continuous-MUSIC*, to locate the human by detecting the reflection signals with subtle TOF variety caused by vital signs.

(ii) High-SNR sensing (Section 4). Once we estimate the location of targeted user, we design *Carrierforming* to enhance the signal SNR of the detected reflection path. Our basic idea is to compensate the phase across multiple subcarriers with different frequency and wavelength to align the peak of all the subcarriers at the location of detected reflection path.

(iii) Multi-user sensing. Although both *Continuous-MUSIC* and *Carrierforming* can be used by any single pair of speaker and microphone, a single pair of them will introduce ambiguity when there are multiple users at different location of the 2D space. To address this dilemma, we extent our method to the platform with a single speaker and a microphone array, and design *Dual-Forming* techniques by combining both traditional beamforming with *Carrierforming* to achieve high resolution sensing capability in the 2D space. With *Dual-Forming*, LoEar is able to monitor the respiration and heartbeat of multiple users in the 2D space simultaneously.

3 CFR MEASUREMENTS EXTRACTION

In this section, we mainly introduce our OFDM-based modulation and demodulation scheme to obtain the CFR measurements of multiple subcarriers with high sampling rate.

To extract the CFR measurements of multiple subcarriers conveniently, our modulation and demodulation schemes are based on the well-known Orthogonal Frequency-Division Multiplexing (OFDM) scheme [2].

We define that the length of transmitting sequence is N_c, the bandwidth of the baseband signal is B, and the sampling rate is $f_s = 48$ kHz. 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. Note that these parameters need to be carefully designed to maximize the enhancement performance of *Carrierforming* as shown in Section **TODO:** 4.

To demodulate the passband signal, we should undergo the down-conversion process to derive the CFR as:

$$CFR_n = \sum_{i=0}^{M-1} A_i e^{-j2\pi f_n \tau_i} \quad (1)$$

where M is the number of paths, A_i is attenuation coefficient of path i, φ_i is the corresponding phase shift caused by the propagation, τ_i is the time of flight (ToF) of path i, including the random system delay for restart, f_n is denoted as the frequency of n-th subcarrier. Note that we have the relationship between the frequency of 1-th subcarrier and n-th subcarrier as: $f_n = f_1 + (n-1)F_s/N$.

After getting the CFR measurements of each subcarrier, we encounter one pressing challenge: CFR measurements

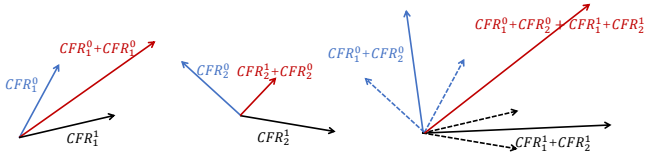


Figure 3: Direct CFR superposition between two subcarriers

are derived at low sampling rates. The low sampling rate in CFR estimation may bring about phase ambiguity when the movements change quickly. The key insight in addressing this issue is to use the property of time-frequency characteristic [20] to upsample the CFR estimations in the frequency domain. We denote the CFR at time t as $CFR_n[t]$. We can upsample the CFR measurements with time step Δt instead of getting only one estimation for each frame.

4 CARRIERFORMING DESIGN

In this section, we introduce the detailed design of *Carrierforming*, which focuses on high-resolution sensing within a large range. We first propose the basic model to enhance the SNR on the particular path. We then analyze the effect of some essential factors on the enhancement performance in depth.

4.1 The 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 [4] 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.,* target path 0 and path 1, respectively. For the subcarrier n, we use vector 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 offset 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.

In order to describe the *Carrierforming* scheme conveniently, let's first make some quantitative assumptions. As the attenuation coefficient A_0 for target path changes exponentially with the traveling delay [21], we set $A_0 = Ae^{-\alpha\tau_0}$, where A and α are the constant values. We can express the CFR at subcarrier n as $CFR_n = Ae^{-\alpha\tau_0}e^{-j2\pi f_n\tau_0}$. Then, for subcarrier n, we multiply 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. We next add up the improvements on the selected subcarriers for the given delay τ_0 as follows:

$$\begin{aligned} CFR_e^{1d} &= \sum_{n=1}^{N} CFR_n e^{j2\pi(n-1)\Delta f\tau} \\ &= CFR_1 \sum_{n=1}^{N} e^{j2\pi(n-1)\Delta f(\tau-\tau_0)} \end{aligned} \tag{2}$$

where CFR_e^{1d} is denoted as the new superimposed CFR measurements for all selected subcarriers relative to the target path.

Ideally, if τ_0 is equal to τ, $|CFR_e^{1d}|$ is equal to $N|CFR_1|$, which indicates that the amplitude of synthesized CFR will be N times as much as the primitive. In contrast, for $\tau_0 \neq \tau$, the amplitude of CFR_e^{1d} is smaller than N times as much as the original CFR. This means that the CFR with delay τ will be enhanced through the constructive offsets across certain subcarriers.

4.2 Depth of Field Vs. Bandwidth

The amplitude of enhanced CFR on the path with delay τ will be theoretically N times larger than the original CFR. However, there are many other noisy paths surrounding the targeted path in the ambient. For example, when the target path is denoted as the path from the user to the device with direct round-trip reflection, there also exists other corresponding reflection paths caused by the desk or the wall. Furthermore, other people may also perform actions around the target leading to series of dynamic paths. Based on the above scenario, there are two issues that need to be carefully considered. Firstly, how to formulate the effect of the noisy multipath on the target path quantitatively? Secondly, how to reduce the deconstructive effect of the multipath effectively?

We assume that there are total M paths from the device to the target, and the traveling delay of any path i is denoted as τ_i, $i = 0, \ldots, M-1$. For any path i, the amplitude of the enhanced CFR can be rewritten as:

$$\begin{aligned} CFR_e^{1d} &= CFR_1 \frac{1 - e^{j2\pi N\Delta f(\tau-\tau_i)}}{1 - e^{j2\pi\Delta f(\tau-\tau_i)}} \\ &= CFR_1 \frac{[1 - e^{j2\pi N\Delta f(\tau-\tau_i)}]e^{j\pi\Delta f(\tau-\tau_i)}}{[1 - e^{j2\pi\Delta f(\tau-\tau_i)}]e^{j\pi\Delta f(\tau-\tau_i)}} \\ &= CFR_1 \frac{\sin(N\pi\Delta f(\tau-\tau_i))}{\sin(\pi\Delta f(\tau-\tau_i))} \end{aligned} \tag{3}$$

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:

$$|CFR_e^{1d}| = |N \cdot CFR_1||\mathrm{sinc}(N\pi\Delta f(\tau-\tau_i))| \tag{4}$$

As the simulated Figure 4a shows, 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 be almost suppressed when the target path and the multipath are far apart. However, for the nearby multipath, the detailed effect is unclear. Therefore, we need to further formulate the relationship between the target path and multipath.

We denote the Half Power Delay Width (HPDW) as the relative difference $|\tau_B - \tau_0)|$, in which the radiation pattern decrease by 50% (or -3 dB) from the peak of the main beam. Therefore, the HPDW should meet the following requirement:

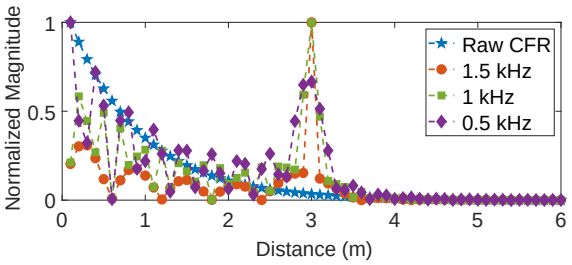
$$|\mathrm{sinc}(N\pi\Delta f(\tau_B - \tau_0))|^2 = 0.5 \tag{5}$$

From Eq. (5), we have:

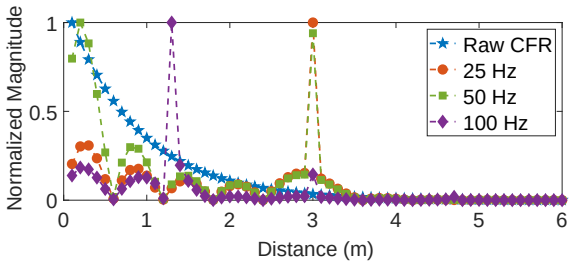
$$B = N\Delta f \cong |1.4/[\pi(\tau_B - \tau_0)]| \tag{6}$$

where B represents the selected bandwidth for the enhancement. Eq. (6) shows that the selected bandwidth is anti-correlated with HPDW, which denotes the valid enhanced range. For the path with $|\tau_i - \tau_0| > |\tau_B - \tau_0|$, the power degrades by more than 50%, and the decay increases with the relatively more significant distance. Therefore, we can effectively reduce the multipath's destructive effect outside the enhanced range by selecting the appropriate bandwidth.

In our system, we aim at sensing the activity within 3.25 *cm* from the target, the valid enhanced range will be given



(a) Enhancement with various bandwidths



(b) Enhancement with various frequency spacing

Figure 4: The effect of different factors on the enhancement

by $|\tau_B - \tau_D| = (2 \times 0.0325)/343 = 0.022$ ms due to round-trip reflection. Thus, we should set the bandwidth as about 2 KHz to enable SNR enhancement in the sensing range and simultaneously reduce the destructive effect of multipath outside the range.

4.3 Grating Lobes Vs. Frequency Spacing

As illustrated in Figure 4b, not only the SNR of the target path can be enhanced, but also those of other paths may also be enhanced. We denote the lobes corresponding to some non-target paths, of which the SNR are also constructively enhanced, as the **grating lobes**, *e.g.,* the purple curves in Figure 4b. The main reason for the existence of the grating lobes is due to the periodicity of the $e^{-j2\pi(n-1)\Delta f(\tau-\tau_i)}$ with $n = 1, \ldots, N$ in Eq. (2) (*i.e.,* $1 = e^{-j2\pi\Delta f(\tau-\tau_i)} = \ldots = e^{-j2\pi(N-1)\Delta f(\tau-\tau_i)}$). Given the existence of grating lobes, we have:

$$|2\pi\Delta f(\tau - \tau_i)| \quad = \quad k2\pi, k = 0, 1, 2, \ldots \tag{7}$$

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. (7) shows that the grating lobes must exist when the sensing range is sufficiently large. In contrast, for the limited range, *i.e.,* $(\tau_\alpha, \tau_\beta)$, we can cancel the grating lobes by:

$$|2\pi\Delta f(\tau - \tau_i)| \quad < \quad 2\pi \tag{8}$$

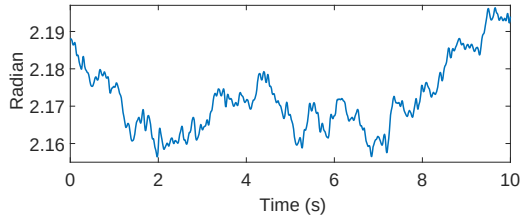
where τ and τ_i are both belong to the range of $(\tau_\alpha, \tau_\beta)$. Moreover, the requirement to satisfy Eq. (8) is given by:

$$\Delta f \quad < \quad 1/|\tau - \tau_i|_{max} \tag{9}$$

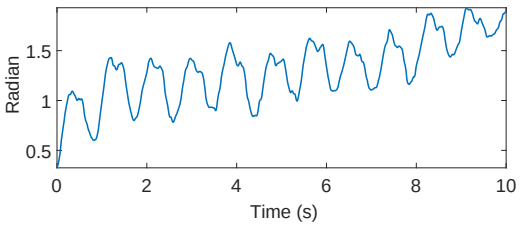
where $|\tau - \tau_i|_{max}$ is the maximal value for the $|\tau - \tau_i|$. Due to $\tau_\alpha < \tau_i < \tau_\beta$ and $\tau_\alpha < \tau < \tau_\beta$, $|\tau - \tau_i|_{max}$ must be smaller than $2|\tau_\beta - \tau_\alpha|$. Therefore, the requirement for avoiding grating lobes can be represented by:

$$\Delta f \quad < \quad 1/(2|\tau_\beta - \tau_\alpha|) \tag{10}$$

In our system, we set the bandwidth as 2 KHz with the frequency spacing of 25 Hz. In this way, we can enhance the SNR within the round-trip range around 7.5 cm from the target and avoiding the grating lobes within the round-trip range of 13.7 m.



(a) Raw CFR profile



(b) Enhanced CFR profile

Figure 5: CFR phase profile changes before and after enhancement

4.4 Model Verification

To verify the effect of signal enhancement with *Carrierforming* model, we ask a volunteer to sit 5 m away from the transceiver holding his breath in a realistic scenario. At the moment, there is no activity other than the user's heartbeat in the environment. Similar to prior work [2, 4], we derive the profile changes of CFR phase under certain subcarrier, as shown in Figure 5a. Apparently, we cannot observe any regular changes in the heartbeat except for drastic noise. However, when the compensated CFR measurements in multiple subcarriers are added up together, we can clearly observe the corresponding regular changes caused by the heartbeat, as shown in Figure 5b. The above results demonstrate the capacity of the *Carrierforming* model in enhancing the SNR with the long range.

5 CONTINUOUS-MUSIC TARGET DETECTION

In 1D Scenario, the purpose of modeling *Carrierforming* is to enhance the SNR with only one pair of speaker and microphone. However, without the known initial location of the target, it is challenging to derive an accurately constructive offset, which is essential for the enhancement. In this section, we focus on searching the target candidates and, then, differentiating the moving target path from the interfering multi-paths.

A natural solution is to use the time delay profile based method to derive the path length estimation. Similar to ResTracker [22], we can firstly get the complex-valued Channel Impulse Response (CIR) after converting CFR_n back to the time domain by IFFT. Then, we can 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 get the target's path delay through the corresponding lightened peak, as shown in Figure 6a. This solution makes sense when the target is nearby the transceiver and turns out to be invalid when the target is far away due to the attenuation of SNR. Figure 6b shows that it is hardly to identify the user path when the user is at 2 m.

To address this issue, we propose a new technique, namely Continuous-MUSIC, inspired by the MUSIC algorithm [23] to detect and estimate the TOF of the target path. As we know, the state-of-the-art MUSIC-based algorithm has been developed to measure AOA and TOF using multiple antennas and subcarriers jointly [24–26]. Inspired by this method, we adopted incident signals on multiple subcarriers to estimate the TOF. The key insight is that incident signals from different distances introduce different amounts of phase changes on each subcarrier.

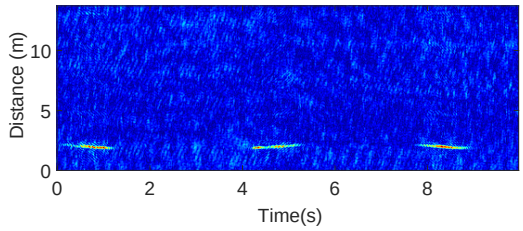
Now, we describe how to estimate the TOF based on MUSIC algorithm using multiple subcarriers. Assume there are N subcarriers at the receiver, and the target is located with the traveling delay τ. The frequency spacing between adjacent subcarriers is set to Δf which should avoid the grating lobes. Obviously, 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 relative to the first subcarriers, as a function of the TOF of the target path:

$$\mathbf{a}(\tau) \quad = \quad [1, e^{-j2\pi\Delta f\tau}, \ldots, e^{-j2\pi(N-1)\Delta f\tau}]^T \qquad (11)$$

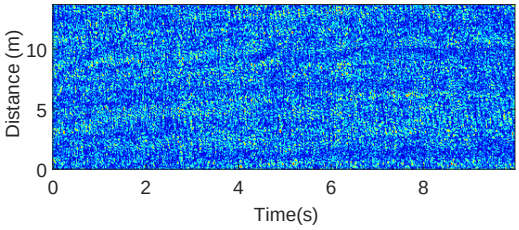
The received signals at subcarrier array with time t can be expressed as:

$$\mathbf{X}(t) \quad = \quad \mathbf{a}(\tau)s(t) + \mathbf{N}(t) \qquad (12)$$

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



(a) Differential CIR map when the user pushes hand back and forth at 1 m



(b) Differential CIR map when the user pushes hand back and forth at 2 m

Figure 6: Differential CIR map when the user is at different distances

can be further expressed as:

$$\mathbf{X}(t) \quad = \quad \mathbf{AS}(t) + \mathbf{N}(t) \qquad (13)$$

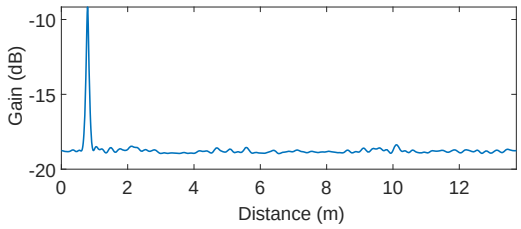
where $\mathbf{A} = [\mathbf{a}(\tau_1), \ldots, \mathbf{a}(\tau_2), \mathbf{a}(\tau_K)]$ and $\mathbf{S}(t) = [s_1(t), s_2(t), \ldots, s_K(t)]^T$.

The key idea behind MUSIC algorithm 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 remanning eigenvectors construct the noise subspace. Besides, these two subspaces are orthogonal to each other.

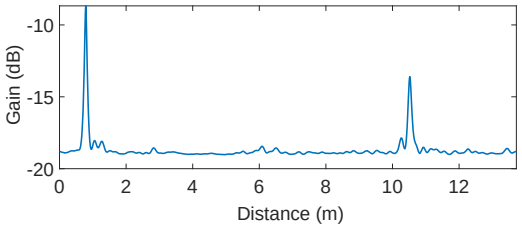
Based on this theory, the next standard process is to compute the eigenvectors of the correlation matrix $\mathbf{X}(t)\mathbf{X}(t)^H$ with the $(N-K)$ smallest eigenvalues and then construct the corresponding noise vector space $\mathbf{E}_{N-K}(t) = [\mathbf{e}_{K+1}(t), \mathbf{e}_{K+2}(t), \ldots,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

$$P_t(\tau) \quad = \quad \frac{1}{\mathbf{a}^H(\tau)\mathbf{E}_{N-K}(t)\mathbf{E}_{N-K}(t)^H\mathbf{a}(\tau)} \qquad (14)$$

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 the volunteer is far 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 0.5 m and ask a volunteer to sit at 4.8 m from the midway point between them. As shown in Figure 7a, we can see the only strong peak indicating the LOS path. This is mainly because the strength of LOS is much

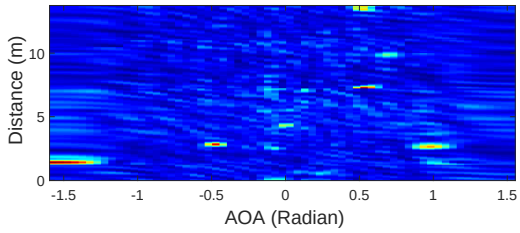


(a) Gains for different distances with standard MUISC algorithm

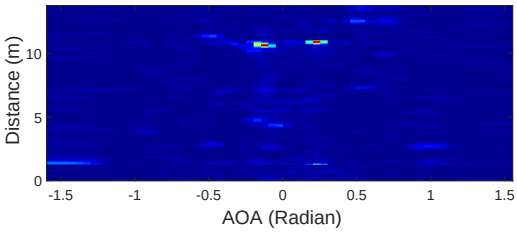


(b) Gains for different distances with Continuous-MUSIC algorithm

Figure 7: comparison between the standard MUSIC and Continuous-MUSIC algorithm



(a) The profile with standard MUSIC



(b) The profile with Continuous-MUSIC

Figure 8: The profile comparison between standard MUSIC and Continuous-MUSIC in 2D scenario

greater than the reflection from the target and surrounding objects (*e.g.,* desk, wall, *etc.*).

A significant issue naturally arises: why is it difficult to distinguish the target reflection when there are some other strong paths (*e.g.,* LOS path, reflection path from the wall, *etc.*) around in the state-of-the-art MUSIC model. There are mainly two reasons for this result. The first reason is that the received signals vector $\mathbf{X}(t)$ in most prior MUSIC model are coming from the same timing points with the size of $(N, 1)$. According to the linear algebra theory [27], the rank of $\mathbf{X}(t)\mathbf{X}(t)^H$ will follow the relationship as:

$$rank(\mathbf{X}(t)\mathbf{X}(t)^H) \leq rank(\mathbf{X}(t)) = 1 \tag{15}$$

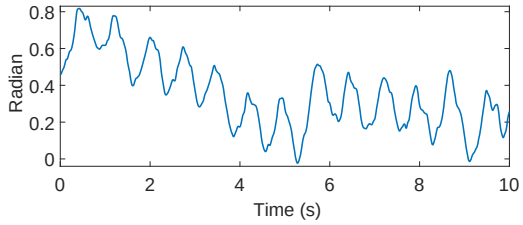
It indicates that there is only one non-zero eigenvalue corresponding to the strongest incident signal and the remaining zero eigenvalues are corresponding to the noise. This means that the state-of-the-art MUSIC model just has the capacity of detecting the strongest path. Note that even if the SpotFi [24] proposes to construct the received signal matrix with $(N, 2)$, the model can only detect the two strongest paths.

The second reason is that the target reflection is so weak that it is easily buried in the stronger reflections in the environment. Although Dynamic-MUSIC [26] proposes to distinguish the target path based on the observation that the large-scale mobile reflection (with strong reflection) may disappear for a period. However, this scheme fails when the user keeps still with only subtle motions, such as respiration and heartbeat.

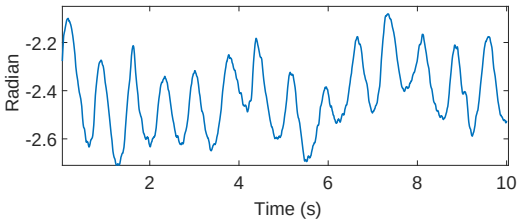
Instead, we propose the *Continuous-MUSIC* technique, which is founded on the received signals from multiple continuous timing points, to accurately detect the weak target reflection with subtle motions. Detailedly, we firstly construct the received signal matrix as $\mathbf{X} = [\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_l)]$, where $\mathbf{X}(t_1)$, $\mathbf{X}(t_2)$,..., $\mathbf{X}(t_l)$ are the continuous received signals at the subcarrier array. In general, we set $l \geq N$ to maximize the rank of $\mathbf{X}$ so as to detect multiple incident signals. Secondly, 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 smallest eigenvalues corresponding to the noise. Finally, similar to Eq. (14), we derive the spectrum function of traveling delay as:

$$P(\tau) = \frac{1}{\mathbf{a}^H(\tau)\mathbf{E}_{N-K}\mathbf{E}_{N-K}^H\mathbf{a}(\tau)} \tag{16}$$

This new model has two superior advantages: Firstly, it significantly improves the rank of $\mathbf{X}\mathbf{X}^H$ based on the difference in signal at different times so that we can distinguish multiple objects in the 1D scenario simultaneously. Secondly, it can accurately distinguish the target with subtle motions (*e.g.,* respiration, heartbeat) from other static reflections such as desk, wall, and the floor. This is because the the increase in rank is mainly due to the change of subtle motion at different times. It means that the target path with subtle motions will be highlighted, while the static reflection will be suppressed. Figure 7b shows that the peak corresponding to target is significantly highlighted comparing with Figure 7a based on the Continuous-MUSIC scheme. Note that since the LOS path has the strongest energy, it still has the highest peak, we can locate the target's path by from the second peak.



(a) Heartbeat of user A



(b) Heartbeat of user B

Figure 9: Enhanced CFR in 2D scenario

6 DUAL-FORMING DESIGN

In this section, we study the high-resolution sensing within large range in the 2D scenario. We first extend the *Continuous-MUSIC* technique to identify multiple objects by combing both the multiple subcarriers and microphones. Second, we propose the model to enhance the SNR corresponding to the reflection from multiple objects. Third, we give an in-depth analysis of the enhancement model factors that are different from those in the 1D scenario.

6.1 Dual Continuous-MUSIC

In the 2D scenario, the prior subcarrier array will be extended to the sensor array comprising of N subcarriers and M microphones. This means that the traveling path will introduce a distinct phase shift depending both on the TOF and Angle of Arrival (AOA). Consider a path with TOF of τ and AOA of θ, the steering vector will be formed by the AOA and TOF over N subcarriers and M microphones, which is given by: $\mathbf{a}(\tau,\theta) = [1, e^{-j2\pi\Delta f\tau}, \ldots, e^{-j2\pi\Delta f(N-1)\tau} \cdot e^{-j2\pi(M-1)f_N\Delta d\sin\theta/c}]^T$. Note that the i-th element in the new steering vector can be represented as: $e^{-j2\pi\Delta f(n-1)\tau} \cdot e^{-j2\pi(m-1)f_n\Delta d\sin\theta/c}$, where $m = \lfloor i/N \rfloor$ and $n = i \mod N$ and $f_n = f_1 + (n-1)\Delta f$. The steering vector $\mathbf{a}(\tau,\theta)$ is with size of MN and the the spacing between two adjacent microphones is Δd. Thus, for the same K incident signals, the received signals with time t will be converted to:

$$\mathbf{X}'(t) \quad = \quad \mathbf{A}'\mathbf{S}(t) + \mathbf{N}(t) \tag{17}$$

where $\mathbf{A}' = [\mathbf{a}(\tau_1,\theta_1), \ldots, \mathbf{a}'(\tau_2,\theta_2), \mathbf{a}(\tau_K,\theta_K)]$. Similar to the Continuous-MUSIC algorithm in the 1D scenario, we take the eigen analysis on the new correlation matrix $\mathbf{X}'\mathbf{X}'^H$, of which $\mathbf{X}'$ is constructed by continuous received signals (*i.e.,* $\mathbf{X}' = [\mathbf{X}'(t_1), \mathbf{X}'(t_2), \ldots, \mathbf{X}'(t_l)]$). Finally, we derive the spectrum function of the traveling delay and incident angle as:

$$P(\tau,\theta) \quad = \quad \frac{1}{\mathbf{a}^H(\tau,\theta)\mathbf{E}'_{N-K}\mathbf{E}'^H_{N-K}\mathbf{a}(\tau,\theta)} \tag{18}$$

where the matrix $\mathbf{E}'_{N-K}$ consists of $(N-K)$ eigenvectors corresponding to the smallest eigenvalues of the correlation matrix $\mathbf{X}'\mathbf{X}'^H$.

6.2 Dual-Forming Model

Given the estimated TOF τ and AOA θ of the target path, we firstly need to understand the phase relationship between different subcarriers and microphones. Assuming that the CFR for n-th subcarrier at m-th microphone is expressed as $CFR_{m,n}$. Similar to 1D scenario, the phase shift of $CFR_{m,n}$ relative to $CFR_{1,1}$ corresponding to target path can be described as: $e^{j\Delta ph_{m,n}} = e^{-j2\pi\Delta f(n-1)\tau} \cdot e^{-j2\pi(m-1)f_n\Delta d\sin\theta/c}$. Thus, for n-th subcarrier at m-th microphone, we compensate the CFR with the constructive offset $e^{-j\Delta ph_{m,n}}$ and add up all the improvements as follows:

$$CFR_e^{2d} \quad = \quad \sum_{m=1}^{M}\sum_{n=1}^{N} CFR_{m,n} \cdot e^{-j\Delta ph_{m,n}} \tag{19}$$

Model Verification: To evaluate the capacity of the *Dual Continuous-MUSIC* technique and *Dual-Forming* model, we do a small benchmark experiment. Detailedly, we ask two volunteers to sit 4.5 m away from the transceiver with a separation of 1.5 m. Figure 8a shows that it is difficult to differentiate the user's path from other stronger static paths using standard 2D-MUSIC algorithm [24] with only respiration and heartbeat. In contrast, *Dual Continuous-MUSIC* algorithm can highlight the user's path and suppress other strong static paths meanwhile, as shown in Figure 8b.

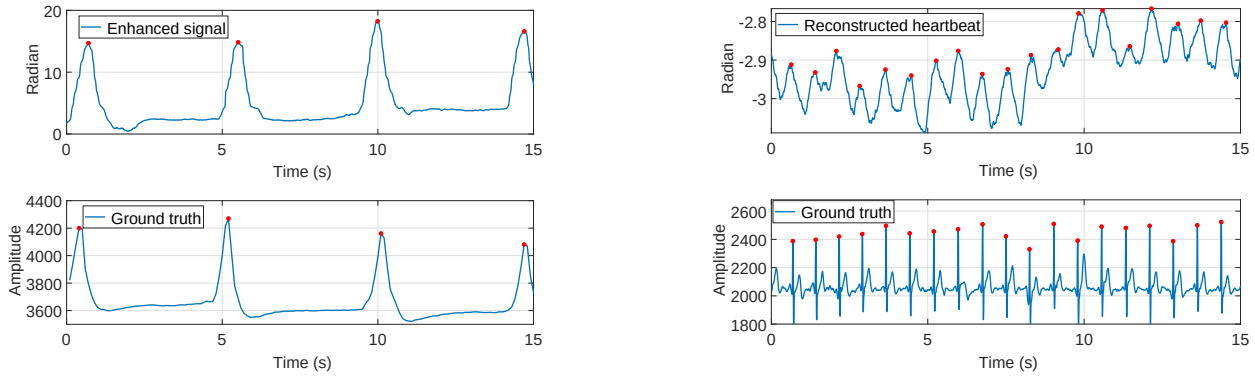
After identifying two users' paths, we asked both users to hold their breath while enhancing the two paths separately. Figure 9 shows that our system can not only separating different users but also monitoring the their heartbeats with super-resolution.

6.3 Factors Analysis

Suppose the true TOF and AOA are τ_0 and θ_0, respectively. The CFR_e^{2d} will be further describe as:

$$\begin{aligned} CFR_e^{2d} = CFR_{1,1}\{[M\textstyle\sum_{n=1}^{N} e^{j2\pi(n-1)\Delta f(\tau-\tau_0)}] + \\ \textstyle\sum_{n=1}^{N}[\sum_{m=1}^{M} e^{-j2\pi(m-1)f_n\Delta d(\sin\theta-\sin\theta_0)/c}]\} \end{aligned} \tag{20}$$

The formula can be analyzed in two aspects. For the first part of the formula, as with the 1D scenario, we can adjust the enhanced TOF range by specific bandwidth and frequency spacing. For the latter part, we analyze the enhanced AOA



(a) Waveform of the enhanced signal

(b) Reconstructed heartbeat signal

Figure 10: Vital signal reconstruction

range through the microphone array width and the microphone spacing. We firstly discuss the effect of the microphone array width $M\Delta d$ on the enhanced AOA range. For the subcarrier f_n, the part $\sum_{m=1}^{M} e^{-j2\pi(m-1)f_n\Delta d(\sin\theta-\sin\theta_0)/c}]$ will be reduced to $M\text{sinc}[\pi f_n M\Delta d(\sin\theta-\sin\theta_0)/c]$. Similar to the beamforming technique [28], we derive the Half Power Beam Width (HPBW) to meet the following requirement:

$$|\pi f_n M\Delta d(\sin\theta-\sin\theta_0)/c| \quad = \quad 1.4 \tag{21}$$

Additionally, since $\pi/2 \le \theta, \theta_0 \le \pi/2$, we have $|\sin\theta-\sin\theta_0| < |\theta-\theta_0|$. Then, we have $|\theta-\theta_0| < 0.45c/(f_n M\Delta d)$. This indicates that the enhanced AOA range $|\theta-\theta_0|$ depends on both the microphone array width $M\Delta d$ and the f_n. Thus, for the given AOA range $|\theta-\theta_0|$, we should have:

$$M\Delta d \quad > \quad 1.4c/(\pi f_n\,|\theta-\theta_0|) \tag{22}$$

Due to $f_1 < f_2 < \ldots < f_N$, we need to have $M\Delta d > 1.4c/(\pi f_1\,|\theta-\theta_0|)$ to ensure the enhancement on selected N subcarriers,

Next, we focus on describing the effect of the microphone spacing Δd. Similar to the effect of frequency spacing on the grating lobes in the enhanced TOF, there may also exist the gating lobes for AOA. Due to the space limit, we omit the inference process and directly give the requirement for avoiding the the gating lobes:

$$\Delta d \quad < \quad c/(|\sin(\theta)-\sin(\theta_0)|\cdot f_n) \tag{23}$$

Due to $|\sin(\theta)-\sin(\theta_0)| \le 2$ and $f_1 < f_2 < \ldots < f_N$, we finally have $\Delta d \le c/(2f_N)$.

To summarize, we set the same bandwidth and frequency spacing as in the 1D scenario to enhance the the SNR in the physical range. Additionally, we set the microphone array width as 14.5 cm with the microphone spacing of 0.9 cm. This setting can enhance the SNR within the angle range around 3.5° from the target and avoiding the grating lobes from −90° to 90°.

7 VITAL SIGNAL MONITORING

In this section, we focus on estimating respiration rate and reconstructing heartbeat from the enhanced CFR measurements.

As illustrated in Figure 10a, the plot shows the amplitude 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 three components, *i.e.,* , respiration, heartbeat. Since the signals in the respiration is much stronger than that in the heartbeat, we can easily estimate the respiration rate based on the detected peaks.

Since the heartbeat is a quasi-periodic movement of which the fundamental frequency is in the range of 0.8 ∼ 2 Hz, we first perform FFT on the enhanced CFR measurement to obtain the spectrum of the signal. Second, we search for the prominent peak pointing to the heart rate in the spectrum of CFR measurements. Third, we employ an adaptive IIR comb notch filter [29], which is normally used to filter out the fundamental frequency and the corresponding harmonic components, on the CFR measurements. Detailedly, the order of the notching filter is set as $Sampling\ rate/Heart\ rate$. The heartbeat is, thus, reconstructed by taking the difference between the CFR measurements and the filtered result, as shown in Figure 10b. Similar to calculating the respiration rate, we perform peak detection and estimate the heart rate by calculating the interval of the adjacent peaks.

8 IMPLEMENTATION

As shown in Figure 11a, we implement LoEar on one acoustic device, *i.e.,* , linear microphone array, and we just employ one of the microphones to receive signals. We separately adopt smartphone to evaluate LoEar for respiration monitoring and smart speaker for heartbeat sensing. The signals are transmitted and received at a sampling rate of 48 kHz with frequency narrowly ranging in $17-21$ kHz that is inaudible to most people [30]. The received signal is further sent to

(a) Devices (b) Scenario

Figure 11: Experimental setup

the laptop (DELL XPS 15 9500) through WiFi for the subsequent real-time processing using MATLAB. The ground truth of the respiration signal is captured using the Vernier respiration belt [31] that records the pressure change of the belt regarding the chest. Additionally, the ground truth of heartbeat signal is captured using 3-lead Electrocardiogram (ECG) monitor (Heal Force PC-80B) [32] that is intended for measuring and recording the ECG waveform and heart rate.

9 EVALUATION

9.1 Experimental Setup

We evaluated the system's performance of measuring vital signs, namely heartbeat rate and respiration rate, in 1D scenarios and 2D scenarios. The experiments were conducted in a square room with a size of $7m \times 7m$. The environmental noise level stays at -80 dbA for frequencies below 5000 Hz and -90 dbA for frequencies ranging from 15 kHz to 22 kHz. For the trans-receiver, the speaker is placed 1 m away on the side of the microphone array so that they both point to the user with the strongest directional gain.

9.2 Experiments in 1D Scenario

User distance: We evaluated the accuracy of our system using 1D augmentation at different distances ranging from 1 m to 5 m with a step of 0.5 m. The results are shown in Fig. 12. When the user is at a close distance (*i.e.,* , $\leq$ 1.5 m), the accuracy of heartbeat reconstruction is not as good as that of larger distances. This is because that, at a small distance, minor actions and movements will also be augmented and magnified so that the weak heartbeat vibration signal is distorted. The results show that 1D augmented vital signal reconstruction works well ranging from 2 m to 5 m, with a minor performance degradation at the longer end.

User clothing: As sound absorption rate varies with garment materials, we also evaluated the performance when users have different types of clothes. A user was asked to repeat the measurement with different clothes on at a distance of 3 m and a AoA of 0 °. As shown in Fig. 13, the sweater has the largest impact on heartbeat accuracy. This is obvious as the sweater is porous, the holes bring huge impedance for sound waves to get through. Note that the heartbeat signal of a user wearing a coat can still be accurately measured despite the thickness. The results show that the system works well with common types of clothes.

Minimal user spacing: We explored the maximum ToA resolution of our system under 1D scenarios, which indicates the minimum spacing of users for the system to distinguish their signals. Two users were guided to sit close to each other at an absolute distance of 3 m to the device, with a fixed difference of 10 ° on their AoA. The system tried to reconstruct the respiration and heartbeat signals as the distance was gradually reduced from 100 cm with a fixed step of 20 cm until the two signals became indistinguishable. The results are shown in Fig. 14. While still capable of restoring the respiration pattern at an inter-distance of 20cm, the restored heartbeat signal becomes significantly contaminated by interference from the other user. This is due to the destructive superposition of the two waves as well as the limited field of view of the hardware. Nonetheless, the experiment shows that the system is able to restore the respiration and heartbeat signal at a minimal inter-distance of 20 cm and 40 cm respectively.

9.3 Experiments in 2D Scenario

User distance: Analogous to the 1D scenario, we also explored how the user distance affects the system performance under 2D enhancing scenarios. The user stays at a fixed angle of 0 ° throughout this experiment. As shown in Fig. 15, the system achieves a relatively small error in both the respiration rate and the heartbeat rate at a distance ranging from 3 m to 6.5 m, which suffices most rooms. At longer distances (*i.e.,* , $\geq 7m$), the heartbeat accuracy starts to reduce significantly, while the respiration rate still stays in a reasonable interval. In general, our system is able to measure the heartbeat signal with negligible error at a distance up to 6.5 m. We did not test the upper distance limit of the respiration rate, but judging from the result we believe the respiration can be accurately reconstructed in a distance up to 8 m.

User angle: Similar to user distance, we also evaluated the impact of user angle. The user stayed at a fixed distance of 5 m, and repeated the measurement several times with various angles ranging from 0 ° to 60 ° as in Fig. 16. The results show that LoEarcan robustly measure respiration rate with an angle range of 120 ° (60 ° for each side). For the heartbeat signal, this range is 100 ° (50 ° for each side). The degradation of the heartbeat signal at a large angle is entailed by both a weaker signal reflected from the human body, as well as the directionality limit of the microphone array (*i.e.,* signals close to the normal have the highest gain).

Minimal AoA difference: To evaluate the performance when users have a small AoA difference, two users were

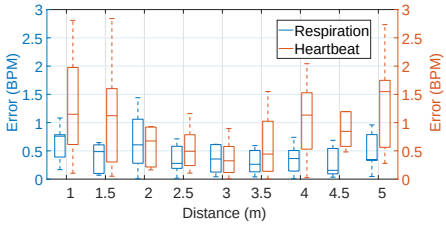


Figure 12: User distance

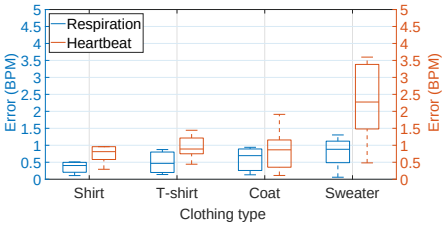


Figure 13: User clothing

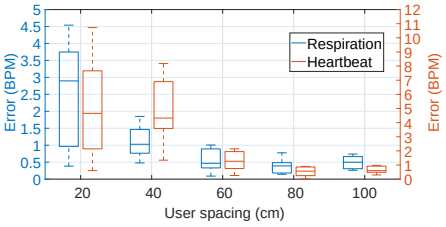


Figure 14: User spacing

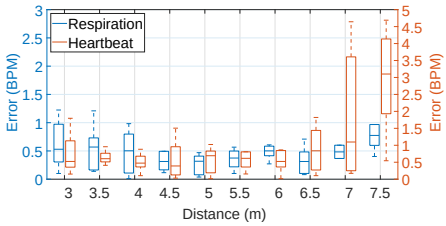


Figure 15: User distance in 2D

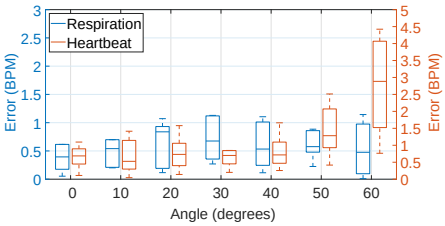


Figure 16: Angle in 2D

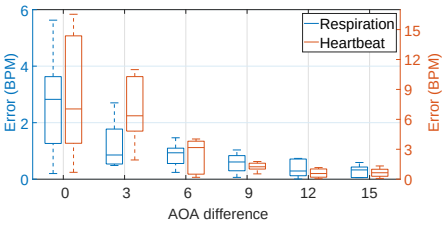


Figure 17: AOA difference

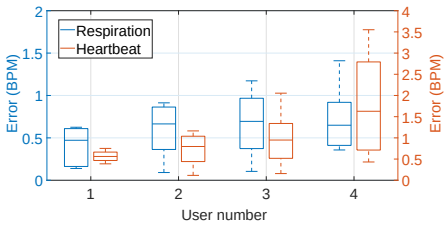


Figure 18: User number

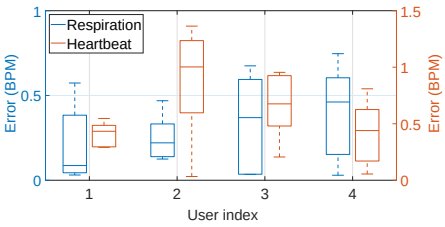


Figure 19: User diversity

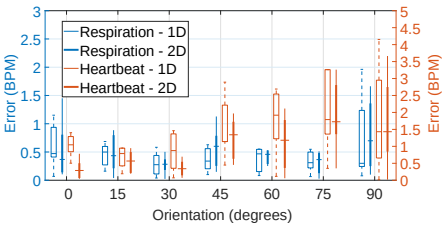


Figure 20: Orientation in 1D and 2D

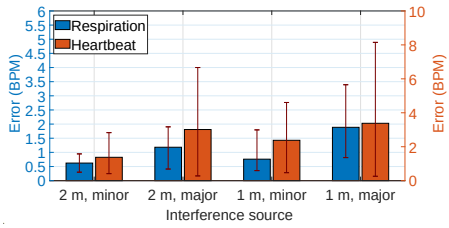


Figure 21: Interference

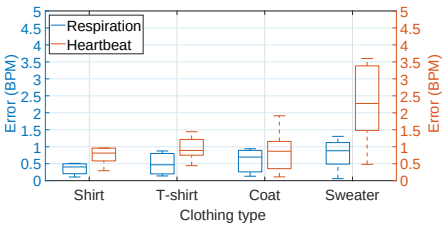


Figure 22: Device type

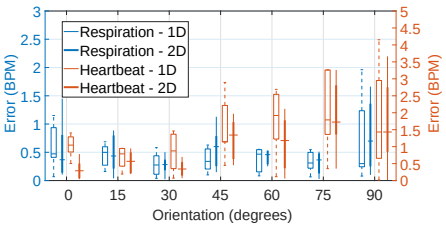


Figure 23: Algorithm

asked to sit at a fixed distance of 5 m. The experiment was repeated several times with different AoA from 15 ° down to 0 ° with a step of 3 °. Fig. 17 shows the result. The performance degrades sharply when the AoA difference gets below 3 °. At a distance of 5 m, which is usually the size of a room, this minimal angle is equal to a distance of 8 cm. This proves that our system can work in most cases where users are randomly scattered inside a room without obstruction.

User number: To determine the maximum user capacity of the system, users were asked to sit at a distance of 4 m with a 15 ° angle interval. The system measured the respiration and heartbeat patterns of the users simultaneously. We ran the experiment multiple times, each time with one more user in play. Fig. 18 shows the result. The system can support a simultaneous measurement of up to 4 people, although for 4 people the heartbeat measurement exhibits an error of up to 4 bpm, which becomes non-ideal. This result shows that our system is capable of supporting common family-scale respiration and heartbeat measurement.

User diversity: We measured the respiration and heartbeat rate of 4 different users at a distance of 4 m and an angle of 0 °. The results in Fig. 19 show that the system has a slightly worse accuracy on user 2, which attributes to thicker clothing. The overall result shows that the performance remains

consistent on different users, which proves the generality of the system in measuring the vital signals.

User orientation: We also tested the system performance when users have different orientations or facing directions. These measurements were conducted at a distance of 5 m and an AoA of 0 °. The user changed his orientation towards the microphone array during this experiment. Results in Fig. 20 show two things: First, respiration reconstruction is more immune to user orientation compared to the heartbeat. This is reasonable as relatively speaking, the amplitude of chest movement changes not as much as the heartbeat signal does when viewing from a side angle. The bodyside causes more reduction to the heartbeat signal relatively. Second, 2D enhancing improves extra accuracy compared to 1D solely, which is especially shown at medium angles such as 45 ° or 60 °. Generally speaking, our system has good orientation tolerance of up to 120 ° in total.

Interference: We explored how the existence of a nearby noise source can affect the performance of the system. We asked a user to take the measurement, while another user nearby tried to make interference by making minor (*i.e.,* shaking legs) and major (*i.e.,* walking back and forth) movements at different locations. Results are shown in Fig. 21. The respiration rate stays robust when a major noise source is 1 m away from the target user. However, in both distances, major movement brings a significant effect to the heartbeat rate measurement. This is reasonable as the reflected heartbeat signal is so weak that it can be easily drowned out by the airflow caused by the major movement noise.

Device type: We also verified the generality of the enhancement model on different microphone-enabled and speaker-enabled devices. The results are shown in Fig. 22. **TODO:** Report the result.

Algorithm: **TODO:** Report

10 RELATED WORK

Recent works related to LoEar are 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.* [33] presents an indoor ultrasonic location tracking system with the localization error around 1 *m*. AAMouse [34] uses the Doppler shift to estimate the velocity and the hand moving distance with the error of 1.4 *cm*. Multiwave [35] leverages the Doppler Effect to translate hand movements into user interface commands. AudioGest [36] and Soundwave [37] presents a device-free system that can sense the hand in-air movement around user's devices. Recent works [10, 38] enable highly accurate hand gesture recognition based on the deep learning technique. Beyond sensing hand gestures, there have been significant works focus on sensing large scale movements. CovertBand [39] tracks individuals' locations and daily activities both within a room using acoustic signals. AcousticID [40] 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 have the superiority in sensing subtle movements. Many prior works have been devoted to subtle motion sensing using acoustic signals. FingerIO [2] tracks subtle finger motion using the reflected acoustic signals. LLAP [4] senses the finger movement direction and distance by using the phase changes of the sound signals with tracking accuracy of 3.5 *mm*. CAT [5] develops a FMCW system achieving accuracy of 5 *mm*. Sun *et al.* [12] proposes acoustic-based tapping scheme to detect finger tapping movement with high accuracy. Strata [3] 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 [14–16] and heartbeat [17, 18]. However, all these sensing systems work in a small range. Instead, our system pushes the limited range of the high-resolution sensing to the room scale using commercial acoustic devices.

Long-range Acoustic Sensing: There have been a few attempts to extend the acoustic sensing range. FM-Track [11] enables tracking hand-sized target with error of 4 *cm*. Mao *et al.* [7] realizes hand motion tracking within 4.5 *m* and achieves $1.2 - 3.7$ *cm* error. DeepRange [6] 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 comparison, our system realizes monitoring subtle motions (*e.g.,* respiration and heartbeat) using single speaker and microphone with the range of 5 *m*. Furthermore, the range can be pushed to 7 *m* with microphone array, and multiple users' subtle motions can be sensed when they are at different locations in 2D space.

11 DISCUSSIONS

While this paper presents a novel high-resolution sensing system for the long range and demonstrates its sensing capacity in various scenarios, there are several limitations to be addressed in the future. Firstly, both *Continuous-MUSIC* and *Dual Continuous-MUSIC* techniques require that all irrelevant objects should keep static before detecting the target. This is because all dynamic reflections will be highlighted in

the profile as shown in Figure 8b. Otherwise, the irrelevant objects may be falsely detected instead of the target. This issue limits the further deployment of our system, and we will consider addressing this challenge in future work.

Secondly, *Carrierforming* also works in a relatively quiet environment, where the significantly environmental dynamics should be far away from the target. For example, when a person is walking around the target (*e.g.,* $\leq$ 1 m), it is practically impossible to sense the subtle motions of 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 is separated from surrounding reflections, and we take this issue as future work.

Thirdly, the experiment shows that the performance of system on the mobile phone is worse than that on the smart speaker. The main reason may be that the transmitted sound power of mobile phone is rather lower. To further explore the universal capacity of the system on mobile phones, we will test more brands of phones on sale in future.

Take the speaker array as the future work.

12 CONCLUSION

In this paper, we developed LoEar, an acoustic-based system which pushes the range limit of the high-resolution sensing. We are the first to focus on increasing the high-resolution sensing range using single microphone and speaker without considering machine learning technologies. To this end, we propose a systematic approach to address the challenges associated with the high-resolution sensing for the long range. Furthermore, we extend the sensing in the 1D scenario to the 2D scenario to simultaneously sense multiple targets with different locations. We demonstrated that the presented system has a strong capacity in sensing subtle motions in different scenarios. We believe the system will significantly benefit the prospective long-range acoustic sensing.

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, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of ACM CHI*, 2016.

[3] 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.

[4] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of ACM MobiCom*, 2016.

[5] Wenguang Mao, Jian He, and Lili Qiu. Cat: high-precision acoustic motion tracking. In *Proceedings of MobiCom*, 2016.

[6] 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.

[7] 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.

[8] Anran Wang and Shyamnath Gollakota. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of ACM CHI*, pages 1–11, 2019.

[9] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of ACM MobiCom*, 2018.

[10] 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.

[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] Ke Sun, Wei Wang, Alex X Liu, and Haipeng Dai. Depth aware finger tapping on virtual displays. In *Proceedings of ACM MobiSys*, 2018.

[13] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of ACM MobiSys*, 2015.

[14] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. Contactless infant monitoring using white noise. In *Proceedings of MobiCom*, 2019.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] Franz Hlawatsch and François Auger. *Time-frequency analysis*. Wiley Online Library, 2008.

[21] 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.

[22] 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.

[23] Ralph Schmidt. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation*, 34(3):276–280, 1986.

[24] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using wifi. In *Proceedings of ACM SIGCOMM*, pages 269–282, 2015.

[25] Jie Xiong and Kyle Jamieson. Arraytrack: A fine-grained indoor location system. In *Proceedings of USENIX NSDI*, 2013.

[26] 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.

[27] Gilbert Strang, Gilbert Strang, Gilbert Strang, and Gilbert Strang. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.

[28] Andreas F Molisch, Vishnu V Ratnam, Shengqian Han, Zheda Li, Sinh Le Hong Nguyen, Linsheng Li, and Katsuyuki Haneda. Hybrid beamforming for massive mimo: A survey. *IEEE Communications Magazine*, 55(9):134–141, 2017.

[29] 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.

[30] 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.

[31] https://www.vernier.com/product/go-direct-respiration-belt/.

[32] http://www.healforce.com/en/html/products/portableecgmonitors/healthcare-equipment-portable-ECG-monitors-PC-80B.html.

[33] Patrick Lazik and Anthony Rowe. Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In *Proceedings of ACM SenSys*, 2012.

[34] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proceedings of ACM MobiSys*, 2015.

[35] Corey R Pittman and Joseph J LaViola Jr. Multiwave: Complex hand gesture recognition using the doppler effect. In *Proceedings of ACM GI*, 2017.

[36] 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.

[37] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: Using the doppler effect to sense gestures. In *Proceedings of ACM CHI*, 2012.

[38] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing*, 2020.

[39] 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.

[40] 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.