DF-sense: multi-user acoustic sensing for heartbeat monitoring with dualforming

Lei Wang, Tao Gu, Wei Li, Haipeng Dai, Yong Zhang, Dongxiao Yu, Chenren Xu, Daqing Zhang

DF-Sense: Multi-user Acoustic Sensing for Heartbeat Monitoring with Dualforming

Lei Wang[SP], Tao Gu[M], Wei Li[P], Haipeng Dai[N], Yong Zhang[C]
Dongxiao Yu[D], Chenren Xu[PK], Daqing Zhang[PIK]
[S]Soochow University [P]Peking University [M]Macquarie University [N]Nanjing University
[C] Shenzhen Institute of Advanced Technology, CAS [D]Shandong University [I]Institut Polytechnique de Paris
[K]Key Laboratory of High Confidence Software Technologies (Peking University), Ministry of Education

Abstract

Acoustic sensing for heartbeat monitoring has become a prevailing research topic in wireless sensing. Existing acoustic sensing systems have two limitations––limited sensing range, and heartbeat monitoring for a single user only, hindering the large-scale deployment of applications. In this paper, we present DF-Sense, a *D*ual *F*orming based multi-user acoustic *Sensing* system for heartbeat monitoring in home settings. Specifically, we design a novel sensing signal-to-noise ratio (SSNR) enhancement model, namely *Dualforming*, based on the constructive superposition across multiple subcarriers and microphones, and further build the quantitative relationship between critical factors and SSNR enhancement to optimize sensing performance. To enable *Dualforming*, we propose a novel MUltiple Subtle SIgnal Classification (*MUS*^{2}*IC*) method to identify multiple subjects with subtle motions. We implement DF-Sense using commercial acoustic devices and conduct extensive experiments in a home setting. Results show that DF-Sense achieves high precision measurement of instantaneous heart rate within the range of 10 m, which is sufficient for most daily space requirements, and is able to monitor heartbeat for up to 6 subjects in a 2-D space simultaneously.

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing**.

Keywords

Beamforming, Acoustic Sensing, Heartbeat

ACM Reference Format:
Lei Wang, Tao Gu, Wei Li, Haipeng Dai, Yong Zhang, Dongxiao Yu, Chenren Xu, Daqing Zhang. 2023. DF-Sense: Multi-user Acoustic Sensing for Heartbeat Monitoring with Dualforming. In *ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '23), June 18–22, 2023, Helsinki, Finland*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3581791.3596867

MobiSys '23, June 18–22, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0110-8/23/06…$15.00
https://doi.org/10.1145/3581791.3596867

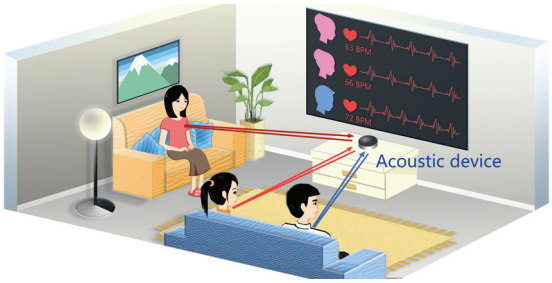


Figure 1: Example of an application scenario for DF-Sense

1 Introduction

Monitoring heartbeat in a home setting plays a crucial role in assessing people's health conditions. Existing solutions leverage on wearable sensors such as Electrocardiograph (ECG) and photoplethysmography (PPG) devices, to record small electrical changes or blood volume changes of the cardiac cycle to derive heart rhythm. Wearable systems can achieve reasonable high accuracy, but wearing sensors in a long-term basis may cause discomfort to users. Contact-free heartbeat monitoring leveraging on computer vision [9] and radio frequency (RF) signals [1, 4, 8, 23, 39, 43–45] have been intensively studied in recent years. Vision-based systems extract heartbeat features from video images captured by camera, however, they require good lighting condition and may raise privacy concerns. RF-based systems typically use Frequency Modulated Continuous Wave (FMCW) [1, 8] or Ultra Wide Band (UWB) [4, 23, 43, 45] to enable heartbeat measurement, but most of them require dedicated and expensive devices.

Acoustic sensing offers a promising solution to contact-free heartbeat monitoring, leveraging on its high sensing granularity and the wide availability of acoustic devices. Several FMCW-based acoustic sensing systems have been proposed to monitor heartbeat [22, 42], but two common limitations exist. First, acoustic signals attenuate rapidly over distance (i.e., path loss), as a result, the sensing range for heartbeat monitoring in existing systems is limited for the large-scale deployment of acoustic sensing applications [22, 33, 42]. In addition, existing acoustic sensing systems focus on single user only. When multiple subjects appear in the space, received acoustic signals will be mixed and entangled, making it difficult to distinguish between multiple users. To address the limitations in existing systems, in this paper, we present DF-Sense, a *Dualforming* based acoustic sensing system for multi-user heartbeat monitoring in home settings. DF-Sense can be deployed on any acoustic

devices/appliances such as TV and portable speaker system with microphone (abbreviated as mic), as shown in Fig. 1. Essentially, we propose *Dualforming*, a novel signal processing technique leveraging on both space-domain beamforming and frequency-domain beamforming, to significantly enhance SSNR [41] at multiple subject locations in a 2-D space. Note that we leverage the metric SSNR to quantify the sensing capacity of our system, which is defined as the ratio of the power of the target reflection and that of the noise. Our system chooses Orthogonal Frequency Division Multiplexing (OFDM) as the basic scheme since it evenly distributes the receiver power over multiple subcarriers. By knowing the locations of multiple users, including path lengths and incident angles relative to the transceiver, *Dualforming* derives the phase corresponding to the subcarrier of each mic in the array, then creates a constructive superposition by compensating for the phase differences across multiple subcarriers and mics. Theoretical results show that it improves the SSNR by MN times, where M and N are the numbers of the mics and subcarriers, respectively. Consequently, *Dualforming* can substantially extend the acoustic sensing range. In addition, since the SSNRs at non-target locations will be deconstructively reduced, we can efficiently sense heartbeat for multiple users at different locations without mutual interference. Using heartbeat as an example, DF-Sense achieves an average heartbeat measurement error of 0.92 Beats per Minute (BPM) in a range of 10 m, and is able to monitor heartbeat for up to 6 subjects in a 2-D space simultaneously. Though promising, two critical issues remain to address. Firstly, there is a big gap between signal superposition and the desired SSNR enhancement. The enhancement effect can be quantitatively measured by four factors––bandwidth, frequency interval between adjacent subcarriers, mic array width and mic spacing. How to correlate the superposition across multiple subcarriers and mics with the SNR enhancement becomes crucial. This is because, given the phase difference between subcarriers, it is difficult to achieve the desired SSNR enhancement by directly superimposing the signals across multiple subcarriers on each mic. Without understanding the correlation between the measurement factors and the enhanced performance, we may have to resort to a trial-and-error approach which is time-consuming and costly. We use *Dualforming* to establish the quantitative model between each of the factors and the SSNR enhancement, providing the theoretical basis for parameter tuning to achieve the desired SSNR enhancement.

However, *Dualforming* requires prior knowledge of subjects' locations. To detect multiple subjects, inspired by the MUSIC algorithm, we propose *MUS^{2}IC* to detect the target path of a subject in distance. By extending the MUSIC algorithm which is based on signal samples of the same instant, *MUS^{2}IC* is built on received signals from multiple continuous instants. With the dynamic signals from continuous frames, the rank of the correlation matrix in the MUSIC algorithm can be increased for the detection of multiple subjects. Meanwhile, the environmental dynamics consist of subjects' subtle motions only. Thus, the subtle motion of each subject can be easily singled out. *MUS^{2}IC* has two superior advantages: 1) able to detect and distinguish multiple subjects simultaneously; 2) able to distinguish subjects with subtle motions (e.g., respiration and heartbeat) from other static objects such as desks, walls, and the floor.

By enhancing SSNR, DF-Sense can sense subtle motions, however, many household appliances may generate similar tiny movements during operation such as fridge and microwave oven. We separate human subjects from non-human motions based on the observation that human respiration is a quasi-periodical movement with a frequency range between 0.16 and 0.33 Hz, while non-human motions usually have a higher frequency range. In addition, how to extract heartbeat patterns from mixed signals which are dominated by the stronger respiration components is critical. A low-range band-pass filter can be applied, but the human body may have some tiny body motions which introduce noise to the frequency range of the heartbeat. We adopt Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMAN) to decompose mixed signals into multiple separated modes, corresponding to each pattern of heartbeat, respiration and other noise, respectively.

The main contributions of this work are highlighted as follows:

- To the best of our knowledge, DF-Sense appears to be the first acoustic sensing system to enable monitoring heartbeats for multiple subjects in a room-scale.
- We design a novel signal processing technique named *Dualforming* to enhance SSNR based on the constructive superposition across multiple subcarriers and mics, and further build the quantitative relationship between critical factors and the SSNR enhancement to optimize sensing performance.
- We propose a novel method named *MUS^{2}IC*, built on received signals from multiple continuous instants, to locate and identify multiple users with subtle motions in distance.
- We conduct extensive experiments using commercial acoustic devices to evaluate DF-Sense. Results show that DF-Sense achieves high precision measurement of instantaneous heartbeat within a range of 10 m, which is sufficient for most daily space requirements, and is able to monitor heartbeat for up to 6 subjects in a 2-D space simultaneously.

2 DF-Sense Design

2.1 Dual Forming Model

Space-domain beamforming [6, 7] combines signals in an antenna array to enhance the SSNR of signals at a particular direction (signals at other directions experience constructive interference). It thus can be used to improve sensing capability. It is also capable of sensing multiple users by enhancing the SSNR of multiple targets in different directions concurrently. However, we discover two limitations when applying this scheme to multi-user sensing. As shown in Fig. 2, when two subjects (e.g., user 1 and user 2) are located at different distances in a similar direction, space-domain beamforming fails to enhance relevant SSNR due to phase ambiguity. Second, performance enhancement is directly related to the size of the antenna array, i.e., a wider array antenna can achieve fine-grained sensing at a longer distance. However, a wider antenna array may not be suitable for home environments.

To address these limitations, we propose *Dualforming* to enhance the SSNR of the signals coming from a specific distance and direction for multi-user fine-grained acoustic sensing. *Dualforming* essentially leverages both space-domain beamforming and frequency-domain beamforming. Similar to antenna array in space-domain beamforming, frequency-domain beamforming combines

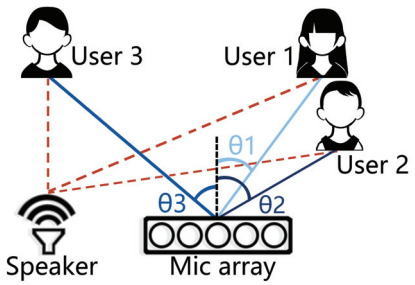


Figure 2: Multi-user model

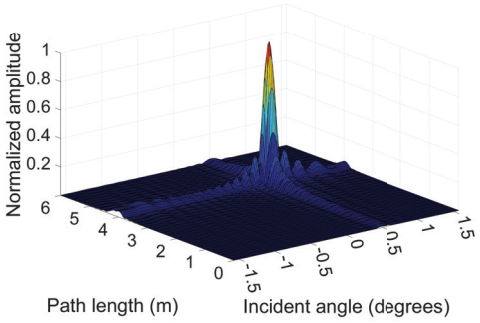


Figure 3: CFR amplitude with *Dualforming* in 2-D space.

OFDM measurements at multiple frequencies in a constructive way to enhance the SSNR of the signal from a certain distance. We elaborate on the *Dualforming* model in details.

Assume that a mic array consists of M mics, and adjacent mics are uniformly separated by the distance of Δd. An object is far away from the mic array with a path length of $d_l(t)$. The incident angle of $\theta_l(t)$ from the first mic and the path length relevant to the m-th mic can be further approximated as $d_l(t) + (m-1)\Delta d \sin(\theta_l(t))$ since the incident angles of object from the first and m-th mic are approximately equal. Then, the mathematical representation of Channel Frequency Response (CFR) at time t for the m-th mic and frequency f_n can be presented as:

$$H_m(f_n,t) = \alpha_{m,n}(l)\exp(-j2\pi f_n \frac{d_l(t)+(m-1)\Delta d\sin\theta_l(t)}{c}) + \epsilon_m(f_n,t), \quad (1)$$

where $\alpha_{m,n}(l)$ is the representation of amplitude attenuation corresponding to l-th path at m-th mic and frequency of f_n, c is the speed of the sound, and $\epsilon_m(f_n,t)$ denotes the corresponding additive white gaussian noise (AWGN). For brevity of description, we replace $\alpha_{m,n}(l)\exp(-j2\pi f_n \frac{d_l(t)+(m-1)\Delta d\sin\theta_l(t)}{c})$ with $A_m(f_n,t)$ (i.e., $A_m(f_n,t) = \alpha_{m,n}(l)\exp(-j2\pi f_n \frac{d_l(t)+(m-1)\Delta d\sin\theta_l(t)}{c})$). Note that $A_m(f_n,t)$ represents the CFR of the target reflection. Since amplitude attenuation is highly related to path length and mic position in the array has little impact on path difference, we assume that $\alpha_{m,n}(l)$ is approximate to $\alpha(l)$, which is determined by path length. For M mics and N frequencies, we express $M \cdot N$ CFR representations regarding the target reflection in the vector form as $\mathbf{A}(t) = [\hat{A}_1(t), \hat{A}_2(t), \ldots, \hat{A}_{M\cdot N}(t)]^{\mathrm{T}}$, where $\hat{A}_k(t) = H_{\lfloor k/N\rfloor+1}(f_{k-N\cdot\lfloor k/N\rfloor}, t)$ is the k-th element in the CFR vector. Similarly, let $W_{m,n}(\theta_w, d_w) = \exp(-j2\pi[f_n(m-1)\Delta d\sin(\theta_w)/c + (n-1)\Delta f d_w/c])$ denote the phase-shifted complex weight for the CFR signals at m-th mic and frequency of f_n, where d_w and θ_w are the weighted path length and incident angle of the l-th path relative to the first mic. Accordingly, we have the weighted vector of $\mathbf{W}(\theta_w, d_w) = [\hat{W}_1(\theta_w, d_w), \hat{W}_2(\theta_w, d_w), \ldots, \hat{W}_{M\cdot N}(\theta_w, d_w)]^{\mathrm{T}}$, where $\hat{W}_k(\theta_w, d_w) = W_{\lfloor k/N\rfloor+1, k-N\cdot\lfloor k/N\rfloor}(\theta_w, d_w)$ is the k-th element in the weighted vector. Combining CFR signals across M mics and N frequencies, we express the synthesized CFR signals regarding the target reflection as:

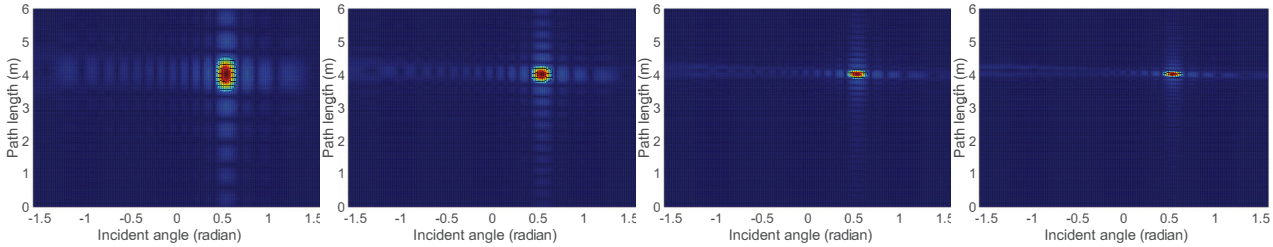
$$\begin{aligned} A_w(t) &= \mathbf{W}(\theta_w,d_w)^{\mathrm{H}}\cdot\mathbf{A}(t) \\ &= \textstyle\sum_{m=1}^{M}\sum_{n=1}^{N} H_m(f_n,t)\cdot W^*_{m,n}(\theta_w,d_w) \\ &= \alpha(l)e^{-j2\pi f_1\frac{d_l(t)}{c}}\textstyle\sum_{n=1}^{N} e^{-j2\pi(n-1)\Delta f\frac{d_l(t)-d_w}{c}}\cdot \\ &\quad \textstyle\sum_{m=1}^{M} e^{-j2\pi f_n(m-1)\frac{\Delta d(\sin\theta_l(t)-\sin\theta_w)}{c}}, \end{aligned} \quad (2)$$

where $\mathbf{W}(\theta_w, d_w)^{\mathrm{H}}$ is the conjugate transpose of $\mathbf{W}(\theta_w, d_w)$ and $W^*_{m,n}(\theta_w, d_w)$ is the conjugate of $W_{m,n}(\theta_w, d_w)$. The changes in the path length due to heartbeat are minimal, thus, we consider $d_l(t)$ to be an approximately constant for a specific user. The expected power of CFR can be further expressed as:

$$\begin{aligned} \mathbb{E}(|A_w(t)|^2) &= \alpha^2(l)\cdot\left|\textstyle\sum_{n=1}^{N}\exp(-j2\pi(n-1)\Delta f\frac{d_l(t)-d_w}{c})\right|^2\cdot \\ &\quad \left|\textstyle\sum_{m=1}^{M}\exp(-j2\pi f_n(m-1)\frac{\Delta d(\sin\theta_l(t)-\sin\theta_w)}{c})\right|^2 \\ &= M^2N^2\alpha^2(l)\left|\mathrm{sinc}(N\pi\Delta f\frac{d_l(t)-d_w}{c})\right|^2\cdot \\ &\quad \left|\mathrm{sinc}(M\pi f_n\frac{\Delta d(\sin\theta_l(t)-\sin\theta_w)}{c})\right|^2. \end{aligned} \quad (3)$$

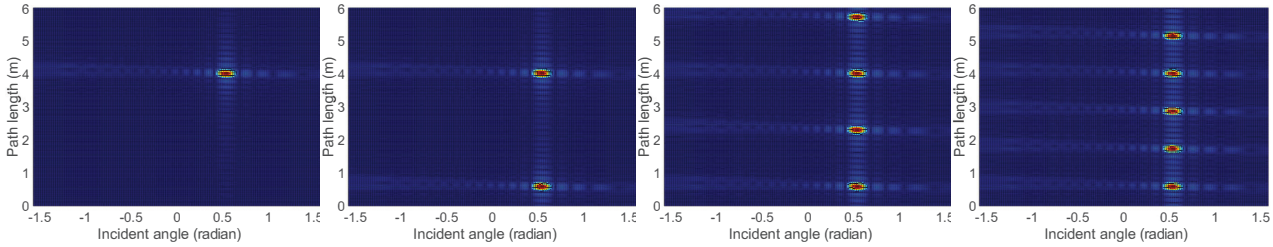
Eq. (3) shows that the expected power of CFR for path l can be maximized when the weighted path length and incident angle are equal to those of the l-th path (i.e., $\theta_w = \theta_l(t)$ and $d_w = d_l(t)$). Conversely, *Dualforming* has an evident degradation with the value increase of $|\theta_w - \theta_l(t)|$ or $|d_w - d_l(t)|$. We give the simulation result in Fig. 3 to illustrate CFR amplitude change at different paths in the 2-D space. The subject is assumed to be located at $(d_l(t), \theta_l(t)) = (4, \frac{\pi}{6})$. The weighted path length and incident angle are set to be equivalent to those in the ground truth. Then, we observe that the CFR amplitude at the target path can be substantially enhanced while the destructive effect of the noisy multipath can be effectively suppressed, which demonstrates the advantage of *Dualforming*.

Similar to the CFR representation $\mathbf{A}(t)$, we also use the vector $\boldsymbol{\epsilon}(t)$ to represent the noises on M mics and N subcarriers, where $\boldsymbol{\epsilon}(t) = [\hat{\epsilon}_1(t), \hat{\epsilon}_2(t), \ldots, \hat{\epsilon}_{M\cdot N}(t)]^{\mathrm{T}}$ and $\hat{\epsilon}_k(t) = \epsilon_{\lfloor k/N\rfloor+1}(f_{k-N\cdot\lfloor k/N\rfloor}, t)$ is the k-th element in the noise vector. Along with the synthesis process of CFR regarding the target reflection, the noise is re-formed as $\epsilon_w(t) = \mathbf{W}(\theta_w, d_w)^{\mathrm{H}}\cdot\boldsymbol{\epsilon}(t)$. Thus, the expected power of noise can be expressed as:



(a) Bandwidth = 0.5 kHz (b) Bandwidth = 1 kHz (c) Bandwidth = 2 kHz (d) Bandwidth = 3 kHz

Figure 4: The relationship between the bandwidth and the enhancement range in the path length.



(a) Freq. spacing = 25 Hz (b) Freq. spacing = 100 Hz (c) Freq. spacing = 200 Hz (d) Freq. spacing = 300 Hz

Figure 5: Relationship between frequency spacing and grating lobes in the path length

$$\begin{aligned}\mathbb{E}(|\epsilon_w(t)|^2) &= \mathbb{E}((\mathbf{W}(\theta_w, d_w)^{\mathrm{H}}\epsilon(t))(\mathbf{W}(\theta_w, d_w)^{\mathrm{H}}\epsilon(t))^*) \\ &= \mathbb{E}(\sum_{i=1}^{MN} \left|\hat{W}_i(\theta_w)\right|^2 |\hat{\epsilon}_i(t)|^2) + \mathbb{E}(\sum_{i=1}^{MN} \\ &\sum_{j=1,j\neq i}^{MN} (\hat{W}_i(\theta_w, d_w)\hat{W}_j^*(\theta_w, d_w)\hat{\epsilon}_i(t)\hat{\epsilon}_j^*(t))\end{aligned} \tag{4}$$

Given that $\left|\hat{W}_i(\theta_w)\right|^2$ is equal to 1 and $\hat{\epsilon}_i(t)$ is the independent and identically distributed AWGN with mean of 0 and variance of σ^2, we have $\mathbb{E}(\sum_{i=1}^{MN} \left|\hat{W}_i(\theta_w)\right|^2 |\hat{\epsilon}_i(t)|^2) = MN\sigma^2$ and $\mathbb{E}(\hat{\epsilon}_i(t)\hat{\epsilon}_j^*(t)) = \mathbb{E}(\hat{\epsilon}_i(t))\mathbb{E}(\hat{\epsilon}_j^*(t)) = 0$ for $\forall j \neq i$. Thus, Eq. (4) can be transformed into $MN\sigma^2$. Combining Eq. (3) and Eq. (4), the maximum SSNR of the enhanced signal can be expressed as:

$$SSNR_{max} = \frac{MN\alpha^2(l)}{\sigma^2} \tag{5}$$

In contrast, for a single mic and subcarrier, the SSNR is only $\frac{\alpha^2(l)}{\sigma^2}$, which theoretically demonstrates that *Dualforming* improves the SSNR by up to MN times comparing to the case with a single mic and subcarrier.

2.2 Quantization of Enhancement

Based on Eq. (3), we know intuitively that the signal from a specific location can be substantially enhanced in a 2-D plane. Further understanding of the impact of various practical factors on enhancement is necessary to guide our design. In this section, we focus on quantifying the relationship between typical factors (i.e., bandwidth, frequency interval, mic spacing, and mic array width) and the SSNR enhancement.

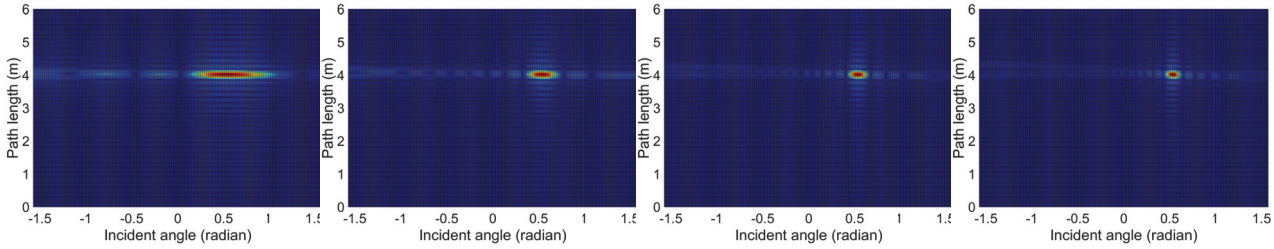
2.2.1 Bandwidth According to the first part (i.e., $\left|\mathrm{sinc}(N\pi\Delta f\frac{d_l(t)-d_w}{c})\right|^2$) in Eq. (3), the bandwidth (i.e., $B = N \cdot \Delta f$) is inversely related to the path length range for SSNR enhancement. As Fig. 4 shows, when we improve the bandwidth, there is a clear narrowing trend in the path length range for SSNR enhancement. To further quantify the relationship between effective enhancement range and bandwidth, we introduce Half Power Beam Range (HPBR) to express path length range, in which the amplitude of the radiation pattern decreases by less than 50% from the peak of the main beam. Therefore, the HPBR is regarded as the path length range for effective sensing. Since Eq. (3) consists of two equally important components regarding the path length and incident angle, the HPBR, expressed as $|d_l(t) - d_w|_{max}$ (maximal value of $|d_l(t) - d_w|$ over time t), should satisfy the requirement as follows:

$$\mathrm{sinc}(\pi B\frac{|d_l(t) - d_w|_{max}}{c})^2 = \lambda. \tag{6}$$

where λ is set as $\sqrt{50}\%$, which indicates that the power decrease by 50%. Hence, the optimal bandwidth can be calculated as:

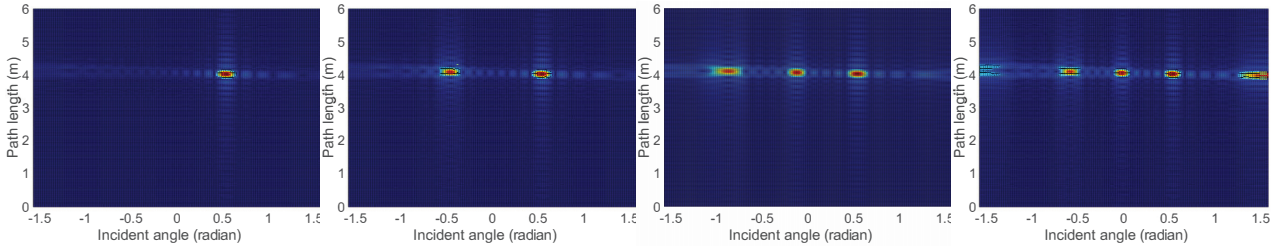
$$B \approx \frac{0.32c}{|d_l(t) - d_w|_{max}}. \tag{7}$$

This means that the SSNR of signals in HPBR will be enhanced, and those outsides will be suppressed by selecting the bandwidth of B. In our system, since the path change caused by heartbeat due to round-trip is about 0.2 − 1 mm (the chest displacement due to



(a) Mic array width = 5 cm (b) Mic array width = 10 cm (c) Mic array width = 15 cm (d) Mic array width = 20 cm

Figure 6: Relationship between the mic array width and the enhancement range along different directions



(a) Mic spacing = 1 cm (b) Mic spacing = 2 cm (c) Mic spacing = 3 cm (d) Mic spacing = 4 cm

Figure 7: Relationship between the spacing and the grating lobes along various directions

heartbeat is 0.1 − 0.5 mm), we set HBPR to 1 mm and the bandwidth is theoretically calculated as 11 kHz. However, the resultant bandwidth requires subject to keep perfectly still, which may not be practical for the elderly or patients. According to our observation, when the human body remains still, the chest moves slowly at different degrees within 2.5 cm. Thus, we set HPBR to 5 cm (due to round-trip), then set the bandwidth to 2.2 kHz. To make the signal inaudible, we set the frequency range of the system from 17 to 19.2 kHz.

2.2.2 Frequency Interval Frequency interval has a direct impact on the enhancement. As Fig. 5 shows, when frequency interval increases, multiple peaks of beams appear obviously along the direction of path length although only one peak corresponds to the subject. The additional beams are often called *grating lobes* on the path. We now explain the quantitative relationship between frequency interval and grating lobes. In Eq. (2), we observe that multiple peaks along the path will come into being when we have:

$$2\pi\Delta f\frac{d(t)-d_w}{c} \quad = \quad k_1\cdot 2\pi, \tag{8}$$

where k_1 is an integer number and $d(t)$ is the length of any path in the environment. As the hotspot illustrated in Fig. 5(a), the main beam is identified when k equals to 0. Similarly, a grating lobe exists when there is a non-zero k_1 that satisfies Eq. (8). Obviously, the interfering objects at the grating lobe position will seriously degrade the enhancement. To avoid grating lobes, we should have:

$$\Delta f \quad < \quad \frac{c}{|d(t)-d_w|_{max}}. \tag{9}$$

In this way, Eq. (8) is satisfied only when k_1 is 0, indicating that only one main beam corresponds to the subject. For example, to ensure no grating lobes within the path length difference of 12 m from the subject, the frequency interval should be less than 28.6 Hz.

Furthermore, as discussed in Section 2.1, the enhanced SSNR is correlated with the number of subcarriers N. With the bandwidth determined in Section 2.2.1, the number of subcarriers equals to $N = \frac{B}{\Delta f}$, which indicates that a lower frequency interval will give rise to a higher SSNR. Ideally, SSNR should be improved as much as possible. However, a lower frequency interval will lead to a lower sampling rate of CFR, also defined as $f_s' = \Delta f$, which may result in ambiguity in detecting signal changes according to Nyquist–Shannon Sampling Theorem[21]. For Δf, the maximum frequency of the dynamic components should be lower than $\frac{\Delta f}{2}$ to avoid aliasing. In this way, the sampling rate f_s' should be high enough to detect environmental vibration, including electric appliances and vital signs. In our system, we set the frequency interval to 25 Hz, which is sufficient to cover the heartbeat information. Meanwhile, the SSNR at the bandwidth of 2.2 kHz will be nearly 88x stronger than that at the single frequency.

2.2.3 Mic Array Width According to the latter part (i.e., $\left|\mathrm{sinc}(M\pi f_n\frac{\Delta d(\sin\theta_l(t)-\sin\theta_w)}{c}\right|^2$) in Eq. (3), the mic array width (i.e., $L = M\cdot\Delta d$) is inversely related to the range of incident angle for SSNR enhancement.

As Fig. 6 shows, with mic array width increases, there is a clear narrowing trend in the range of incident angle for SSNR enhancement. Similar to bandwidth, we introduce another term that is

commonly used in beamforming [18], namely Half Power Beam Width (HPBW), to further quantify the relationship between the effective enhancement range of incident angle and the mic array width. HPBW denotes a range of incident angle in which the amplitude of the radiation pattern decreased by less than 50% from the peak of the main beam, which is regarded as the range of incident angle for effective sensing. Since Eq. (3) consists of two equally important components regarding the path length and incident angle, $|\sin\theta_l(t) - \sin\theta_w|_{max}$ (maximal value of $|\sin\theta_l(t) - \sin\theta_w|$over time t) should satisfy the following conditions similar to Eq. (6):

$$\text{sinc}(L\pi f_n \frac{|\sin\theta_l(t) - \sin\theta_w|_{max}}{c}) \quad = \quad \sqrt{50\%}. \tag{10}$$

Then, the mic array width can be approximated as:

$$L \quad \approx \quad \frac{0.32c}{f_n\,|\sin\theta_l(t) - \sin\theta_w|_{max}}. \tag{11}$$

According to L'Hospital's rule [28], we have $|\sin\theta_l(t) - \sin\theta_w| \approx |(\theta_l(t) - \theta_w)\cos\theta_l(t)|$. Since $\theta_l(t) \in [-\frac{\pi}{3}, \frac{\pi}{3}]$ covers most of the sensing range, we have $|\sin\theta_l(t) - \sin\theta_w| \geq \frac{1}{2}\,|(\theta_l(t) - \theta_w)|$. Besides, we have $f_1 < f_2 < \ldots < f_N$ since the frequency of n-th subcarrier is given by $f_n = f_1 + (n-1)\Delta f$. Therefore, to ensure the enhancement over N subcarriers, the mic array width should be set as follows:

$$L \quad \approx \quad \frac{0.64c}{f_1\,|\theta_l(t) - \theta_w|_{max}}, \tag{12}$$

where $|\theta_l(t) - \theta_w|_{max}$ is defined as HPBW. From Eq. (12), we infer the optimal mic array width to enable effective sensing within the given HPBW. In our system, we set the mic array width to 0.135 m to effectively sense an object within an incident angle difference of 0.1 radians from a given location.

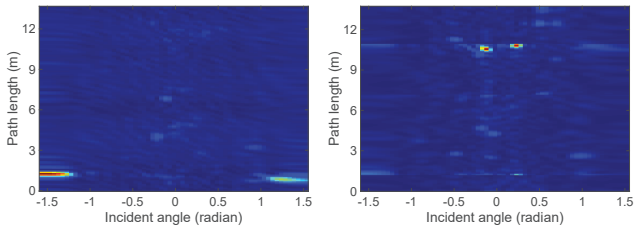
2.2.4 Mic Spacing Similar to the effect of frequency interval on grating lobes on the path, there may also exist *grating lobes* along with the incident direction, which are associated with the mic spacing, i.e., Δd. As shown in Fig. 7, when the mic spacing increases, multiple peaks of beams arise, and the number is on the rise. Considering the latter part of Eq. (2), the peak of beam appears when:

$$2\pi f_n \frac{\Delta d\,(\sin\theta_l(t) - \sin\theta_w)}{c} \quad = \quad k_2 \cdot 2\pi, \tag{13}$$

where k_2 is an integer number. Obviously, the main lobe appears when k_2 equals to 0, and grating lobe appears when k_2 is non-zero. To avoid grating lobes along with the incident directions, the mic spacing should be set to satisfy the requirement as follows:

$$\Delta d \quad < \quad \frac{c}{(|\sin\theta_l(t) - \sin\theta_w| \cdot f_n)}. \tag{14}$$

Since both $\theta_l(t)$ and θ_w belong to $[-\frac{\pi}{3}, \frac{\pi}{3}]$, we derive that $|\sin\theta_l(t) - \sin\theta_w| \leq \sqrt{3}$. In addition, due to $f_1 < f_2 < \ldots < f_N$, we finally have $\Delta d < \frac{c}{\sqrt{3}f_N}$. For the maximum frequency of 19.2 kHz, we set the mic spacing to 0.9 cm to avoid spatial signal ambiguity.



(a) Standard MUSIC (b) MUS^{2}IC

Figure 8: Spectrogram comparison between standard MUSIC and *MUS^{2}IC*.

2.3 Localization for Subjects with Subtle Motions

The objective of *Dualforming* is to enhance the SNR of the long-range target path for fine-grained activity sensing. However, without the target's physical location (i.e., incident angle and path length), it is challenging to achieve the desired enhancement in the presence of the abundant multipath reflections for the indoor environment. To localize subjects, it is important to identify and differentiate the target path with subtle motion from the interfering paths. The MUSIC algorithm has been used to identify and estimate the 2-D location of target [11, 14]. However, it requires target to be either a source of signals or with large-scale movements such as walking, to introduce intermittently appearing strong reflections. Furthermore, they have limited capability of identifying a small number of high-energy targets. Unlike the prior studies, we need to differentiate multiple passive targets with subtle motions from rich indoor multipath simultaneously, which is more challenging. Inspired by the MUSIC algorithm, we propose a novel method, namely *MUS^{2}IC*, to identify and accurately estimate the incident angle and path length of subjects with subtle motions. In what follows, we describe the process of the MUSIC algorithm and then present our scheme.

2.3.1 Estimation with MUSIC The basic idea of the MUSIC algorithm is that the specific path length and incident angle will introduce different amount of phase differences across different subcarriers and mics at the receiver. Assume that there are M mics and N subcarriers at the receiver, where the frequency interval between adjacent subcarriers is Δf, and the adjacent mic spacing is Δd. For a subject located with the incident angle of θ and path length of d, the introduced phase difference of two consecutive mics can be express in the form of complex exponent as $\Phi_\theta = e^{-j2\pi f_n \frac{\Delta d \sin\theta}{c}}$, where f_n is the frequency of n-th subcarrier. Similarly, the phase difference of two adjacent mics can be represented as $\Omega_d = e^{-j2\pi\Delta f \frac{d}{c}}$. Thus, for an incident angle of θ and path length of d, steering vector, which is denoted as *virtual joint array* consisting of phase differences introduced at M mics and N subcarriers, is given by:

$$\begin{aligned}\mathbf{a}(\theta, d) = \quad & [1, \ldots, \Omega_d^{N-1}, \Phi_\theta, \ldots, \Phi_\theta\Omega_d^{N-1}, \ldots, \Phi_\theta^{M-1}, \\ & \ldots, \Phi_\theta^{M-1}\Omega_d^{N-1}]^{\mathrm{T}}.\end{aligned} \tag{15}$$

When there are K incident signals which can be represented as $\mathbf{S}(t) = [s_i(t), s_i(t), \ldots, s_K(t)]^{\mathrm{T}}$ at the first mic and subcarrier, the

received signal at the joint array can be further written as:

$$\mathbf{X}(t) = \mathbf{A}\mathbf{S}(t) + \mathbf{N}(t), \tag{16}$$

where $\mathbf{N}(t)$ is a MN-by-1 noise vector and $\mathbf{A} = [\mathbf{a}(\theta_1, d_1), \mathbf{a}(\theta_2, d_2), \ldots, \mathbf{a}(\theta_K, d_K)]^{\mathrm{T}}$.

MUSIC essentially decomposes the covariance matrix $\mathbf{X}(t)\mathbf{X}(t)^H$ (H denotes the conjugate transpose operation) into multiple eigenvectors corresponding to the signal space and noise space, all of which are orthogonal to each other. Accordingly, MUSIC arranges the eigenvectors in an ascending order of eigenvalues and then uses the eigenvectors corresponding to the smallest $(MN - K)$ eigenvalues to form the noise vector subspace $\mathbf{U}_{\mathbf{N}}(t) = [\mathbf{u}_1(t), \mathbf{u}_2(t), \ldots, \mathbf{u}_{MN-K}(t)]^{\mathrm{T}}$. Due to the orthogonality of signal and noise space, the spectrum function related to incident angle θ and path length d can be expressed as:

$$P_M(\theta, d) = \frac{1}{\mathbf{a}(\theta, d)^H \mathbf{U}_{\mathbf{N}}(t)\mathbf{U}_{\mathbf{N}}(t)^H \mathbf{a}(\theta, d)}. \tag{17}$$

Theoretically, subjects can be clearly distinguished by looking for the peak in the spectrogram.

2.3.2 Limitations of MUSIC The MUSIC algorithm is not able to measure strong light-of-sight (LOS) paths or strong reflective objects when LOS is relatively weak. However, when a subject is further away from the transceiver or always keeps still, the prior method will be disabled. For example, we ask two subjects to sit 4.5 m away from the transceiver and 1 m away from each other and breathe normally. As shown in Fig. 8(a), the peak indicating a strong LOS path can be clearly identified, while the peak referring to subjects is completely buried into the background noise. We analyze two causes that lead to such ambiguity. First, received signals vector $\mathbf{X}(t)$ in the MUSIC algorithm comes from the same timing points with a size of $(MN, 1)$. According to linear algebra [26], the rank of $\mathbf{X}(t)\mathbf{X}(t)^H$ (i.e., $R[\mathbf{X}(t)\mathbf{X}(t)^H]$) follows the relationship as:

$$R[\mathbf{X}(t)\mathbf{X}(t)^H)] \leq R[\mathbf{X}(t)]. \tag{18}$$

It indicates that only one non-zero eigenvalue corresponds to the strongest incident signal, and the remaining zero eigenvalues correspond to the noise. This implies that MUSIC has the capacity of detecting the only strongest path. Second, the subject reflection is weak and can be easily drowned into stronger reflections in the environment. Although Dynamic-MUSIC [14] distinguishes the target path based on an observation that the large-scale mobile reflection (with strong reflection) may disappear intermittently. However, this scheme fails when the subject keeps still with only subtle motions (e.g., respiration and heartbeat).

2.3.3 Estimation with MUS^{2}IC To address the limitation in distinguishing the subject with the subtle motion from the stronger static signals such as the LOS and other static reflections (e.g., walls, doors, tables, etc.), we propose *MUS^{2}IC* to discover the time-varying correlations between receiving signals among *joint array*. The key idea of *MUS^{2}IC* is to track the principal dynamic components of change from the signal at multiple consecutive instants.

We describe the process of *MUS^{2}IC* in detail as follows. First, we obtain the CFR streams at the *virtual joint array* from multiple consecutive instants and form a new MN-by-L matrix $\mathbf{X}$, where $\mathbf{X} = [\mathbf{X}(t_1), \mathbf{X}(t_2), \ldots, \mathbf{X}(t_L)]$. The interval of the CFR streams is set to 1 s to ensure sufficient dynamic components can be captured.

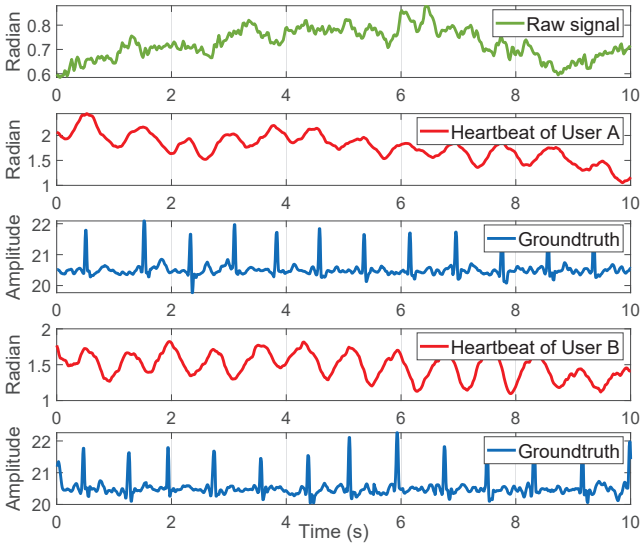


Figure 9: Illustration for heartbeat monitoring when holding breath

Meanwhile, the number of samples L should be greater than that of the subjects to ensure accurate identification of multiple subjects. Second, we concentrate on removing the influence of the static path component from each CFR stream. In detail, we calculate the constant offset, representing the static path component, by averaging each row of samples in matrix $\mathbf{X}$ (i.e., each CFR stream). Then, by subtracting the constant offset from each CFR stream, we obtain the residual matrix $\mathbf{X}'$, which is dominated by the principal components indicating the dynamic changes. Third, similar to steps in the MUSIC algorithm, we take the Eigen analysis on the new correlation matrix $\mathbf{X}'\mathbf{X}'^H$ and derive the corresponding noise vector subspace $\mathbf{U}'_{\mathbf{N}}(t) = [\mathbf{u}'_1(t), \mathbf{u}'_2(t), \ldots, \mathbf{u}'_{MN-K}(t)]^{\mathrm{T}}$. Finally, we derive the spectrum function of incident angle and path length as:

$$P'_M(\theta, d) = \frac{1}{\mathbf{a}(\theta, d)^H \mathbf{U}'_{\mathbf{N}}(t)\mathbf{U}'_{\mathbf{N}}(t)^H \mathbf{a}(\theta, d)}. \tag{19}$$

To explore the capability of *MUS^{2}IC* in detecting multiple subjects, we process the same signals in Fig. 8(a) with *MUS^{2}IC*, and the results are shown in Fig. 8(b). Compared to Fig. 8(a), we observe that the strong LOS path is severely suppressed, while in contrast, the weak reflection paths representing the two volunteers far away from the transceiver are clearly identified. This clearly indicates the benefits from the superiority of *MUS^{2}IC* in extracting small-scale change in chest displacement caused by respiration as well as eliminating the strong static components. Beyond that, these two close-by subjects do not interfere with each other, further demonstrating the capacity of *MUS^{2}IC* in localizing multiple subjects with subtle motions.

3 Heartbeat Extraction

3.1 User Identification

MUS^{2}IC can detect multiple subjects with subtle motions, however, it may also detects vibration of home appliances such as fridge

or microwave oven. To address this issue, we analyze frequency distribution to avoid environmental influence. Human respiration is a quasi-periodical movement with dominant frequency components in a range of 0.16 ∼ 0.33 Hz, and the frequency components of appliances are usually much higher. Based on this observation, we divide the sum of the power within the frequency range of 0.16 ∼ 0.33 by the sum of the frequency of 0.1 ∼ 5 Hz. When the ratio exceeds 0.6, the pattern should correspond to the subject. Additionally, there may be interference from appliances that share similar frequencies. However, mechanical vibrations exhibit higher stability and distinctive vibration amplitudes, allowing for differentiation from human vital signs.

3.2 Sensing Heartbeat

Using *MUS²IC* to calculate the 2-D location of subject, DF-Sense can theoretically employ the *Dualforming* model to quantitatively boost the SSNR of the area near a subject to improve sensing performance. To explore the capability of DF-Sense in sensing subtle heartbeat, we first carry out a benchmark experiment. We use the same setup as in Fig. 8. We ask two subjects to breathe normally while keeping still. Due to displacement change induced by respiration, we obtain the incident angle and path length of two subjects from *MUS²IC*. We then ask them to hold their breath while remaining still. Consequently, the SNR in the area around the subjects can be boosted using the *Dualforming* technique, as shown in Fig. 9. We observe that the original signal waveform fluctuates irregularly while *Dualforming* outputs the heartbeat patterns corresponding to each subject, respectively, demonstrating its capability of heartbeat monitoring.

3.3 Heartbeat Pattern Extraction

To discover heartbeat pattern, intuitively we can apply a low-range band-pass filter to CFR measurements since respiration and heartbeat differ in frequency. Respiration and heartbeat have a frequency range of $(0.16 - 0.33)$ Hz and $(1 - 2)$ Hz, respectively, thus a band-pass filter in the passband frequency range of $(0.8 - 2.2)$ Hz may be a feasible solution. This indeed works when subject keeps absolutely still without any body movements. In reality, human body may have tiny body motions such as leaning forward/backward and shrugging shoulders which will introduce noise to the frequency range of $(0.1 - 2)$ Hz. We observe that the above three components in CFR measurements overlap in frequency, therefore, a simple band-pass filter cannot effectively filter out the noise.

In DF-Sense, we apply Empirical mode decomposition (EMD) [10], an adaptive technique for analyzing non-stationary signals, to effectively decompose CFR streams into a series of independent time-frequency components (i.e., various modes corresponding to different motion patterns), which are called intrinsic mode function (IMF). However, it gives rise to a potential issue––"mode mixing" in EMD. Multiple oscillations with disparate scales may arise in the same mode or the oscillation with similar scales may appear in different modes, creating difficulty in extracting heartbeat patterns. To address the mode mixing problem, Ensemble EMD (EEMD) [37] adds different realizations of white Gaussian noise. However, it may result in instability in the number of modes and incompleteness of the decomposition due to the white noise. By extending EEMD, Complete Ensemble Empirical mode Decomposition with Adaptive

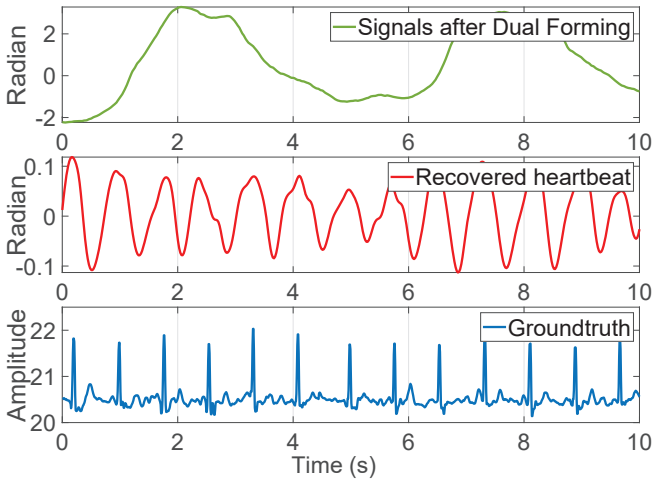


Figure 10: Recovered heartbeat signal via ICEEMDAN

Noise (CEEMDAN) [29] has been proposed to solve the instability of the number of modes and achieve the completeness. We adopt an extended method––Improved CEEMDAN (ICEEMDAN) [5] to extract heartbeat patterns in DF-Sense.

Let operator $E_k(.)$ denote the k-th IMF obtained from EMD, ε_k denotes the SSNR coefficient, $M(.)$ denotes the operator that produces a local average on the signal, $\cdot$ and $n_i(t)$ denotes the i-th realization of the Gaussian white noise. Then, heartbeat pattern extraction undergoes the following steps. (1) Add the Gaussian white noise on the enhanced CFR signal $H_w(t)$ and repeat this process I times to get a realization sequence: $\{H_i(t) = H_w(t) + \varepsilon_0 E_0(n_i(t)), i = 1, \ldots, I\}$. Apply EMD (the detailed procedure can be found in [10]) on the sequence to compute the local mean of I realizations, then the first residue can be achieved as: $r_1(t) = \sum_{i=1}^{i=I} M(H_i(t))$. (2) Compute the first IMF mode as: $d_1(t) = H_w(t) - r_1(t)$. (3) Compute the second residue by performing EMD on the realization of $r_1(t) + \varepsilon_1 E_2(n_i(t))$ as $r_2(t) = \sum_{i=1}^{i=I} M(r_1(t) + \varepsilon_1 E_2(n_i(t)))$. Then, the second IMF mode can be expressed as: $d_2(t) = r_1(t) - r_2(t)$. (4) Compute the k-th residue as $r_k(t) = \sum_{i=1}^{i=I} M(r_{k-1}(t) + \varepsilon_{k-1} E_k(n_i(t)))$ when $k = 3, \ldots, K$. (5) Compute the k-th IMF mode as $d_k(t) = r_{k-1}(t) - r_k(t)$. (6) Iterate the process from Step 4 to 5 for the next k. The coefficient $\varepsilon_k = \varepsilon \cdot \mathrm{std}(r_k(t))$ is set to allow adjusting the SNR at each iteration.

To verify the effectiveness of ICEEMDAN, we conduct an experiment by asking a subject to sit and breathe normally at a distance of 4.5 m from the transceiver. Fig. 10 plots the raw heartbeat waveform obtained from the ECG device, and the heartbeat patterns extracted by ICEEMDAN, and the results show that the heart rate is identical to the ground truth.

After the decomposition process, we perform peak detection by localizing the extreme points from the decomposed CFR sequence. Note that the interval of two detected neighboring peaks should be in the range of 0.5 ∼ 1 s, which corresponds to the cardiac cycle, to avoid a false alarm. The instantaneous heart rate can be

(a) Device (b) Hall (c) Meeting room

Figure 11: Experimental settings

further estimated accordingly. Note that to preserve the potential loss features of chest motion, respiration, and anomalous heartbeat patterns, the IMF component that corresponds to the heartbeat can be replaced with a zero sequence, enabling the recovery of loss features using ICEEMDAN.

4 Evaluation

4.1 Experimental Setup

We implement DF-Sense using an array of 16 MEMS mics (InvenSense ICS-40730) and loudspeaker PHILIPS SPA710, as shown in Fig. 11(a). The spacing between the adjacent mics is 0.9 cm, and the system works with the bandwidth from 17 to 19.2 kHz, which is inaudible to humans [24]. It is important to note that the present system operates on a commercial audio system featuring a 48 kHz sampling rate. However, the techniques can be readily adapted to other audio systems with a higher sampling frequency. The mic array and speaker are placed close to each other so that the round-trip paths to remote subjects have approximately the same length. The received signal is sent to a laptop (DELL XPS 15 9500) through WiFi for further processing using MATLAB. The ground truth of heartbeat signal is collected with a 3-lead Electrocardiogram (ECG) (Heal Force PC-80B) that is used for recording and measuring the actual heart rhythm. If not specified, all experiments are conducted in the hall, as shown in Fig. 11(b), and the environmental noise level keeps at 40 dBA. We recruit ten university students with ages from 21 to 32 years to participate in our experiments. In all experiments, all subjects are required to breathe normally. For each subject, we collect samples for 2 minutes and repeat ten times to reduce the impact of random errors. The subject stays at a fixed distance of 5 m and an incident angle of 0 ° relative to the transceiver throughout each experiment. Note that all experiments are conducted upon approval of the institutional review board (IRB) at our institute.

4.2 System Performance

Monitoring at different distances: *DF-Sense achieves an average heartbeat measurement error of less than 0.92 BPM when the subject is less than 10 m away from the transceiver.* As shown in Fig. 12, we observe that the measurement error stays relatively low and stable when the distance is less than 4 m. For a longer distance (i.e., ≥ 5 m), there is a slight increase in the measurement error. This is probably due to the degradation of energy at a longer distance. In reality, a range of 10 m may be more than sufficient for most home environments. In addition, we compare DF-Sense with the state-of-the-art space-domain beamforming method [31, 36]. As shown in Fig. 12, beamforming achieves similar low errors at a distance of less than 3 m, but dramatically decreases in performance as the distance increases, demonstrating the superiority of DF-Sense over space-beamforming.

Monitoring at different incident angles: *DF-Sense achieves an average heartbeat measurement error of less than 0.92 and 1.14 BPM when the subject is at a distance of 3 m and 5 m from the transceiver, respectively.* In this experiment, we ask the subject to stay at a fixed distance of 3 and 5 m, and repeat the measurement several times with various angles of incidence ranging from 0 ° to 60 °. As shown in Fig. 13, DF-Sense accurately extracts heart rate with an angle range of 100 ° (50 ° for each side). The degradation of heartbeat signal at a large angle of incidence is entailed by both a weaker signal reflected from the human body, as well as the directionality limit of the mic array (i.e., signals close to the normal have the highest gain).

Monitoring at different spacing: *DF-Sense achieves an average error of less than 1.18 BPM in measuring heartbeat when the radical spacing between two subjects is greater than 0.5 m.* In this experiment, we ask two subjects to sit in an approximate line along the transceiver and the subject in front does not block the signal received by the latter subject. We keep the front subject stationary and then move the other subject radially to increase the spacing between them. Fig. 14 compares *Dualforming* with beamforming in terms of accuracy under various spacings. We observe that *Dualforming* accurately extracts the heartbeat of each subject while beamforming ends up with unacceptable measurement errors.

Monitoring at various incident angle differences: *DF-Sense achieves an average measuring error of less than 1.21 when the incident angle difference between two subjects is greater than 15 °.* To explore the performance when subjects have various incident angle differences at different distances, in this experiment, we ask two subjects to sit at a distance of 3 and 5 m, respectively. The experiment is repeated with different angles from 0 ° up to 30 ° with a step of 5 °. Fig. 15 shows that the angle difference of less than 10 ° (≤ 10 °) incurs a considerable error in measurement for both distances. In comparison, a higher difference (≥ 15 °) results in an acceptable accuracy for both distances and tends to be stable after 20 °. Note that at the distance of 3 m, the angle difference of 15 ° corresponds to a spacing of about 0.8 m, which is acceptable for most users.

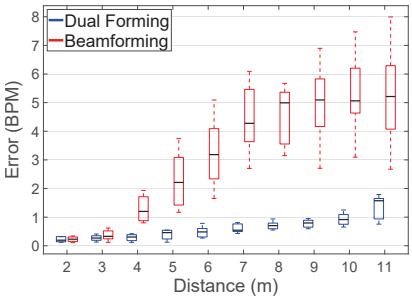


Figure 12: Accuracy at different distances

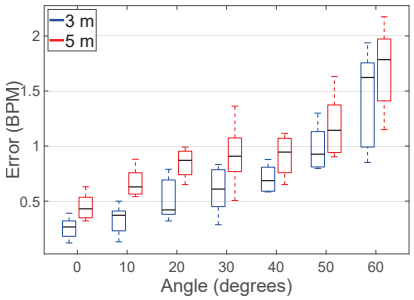


Figure 13: Accuracy at different angles of incidence

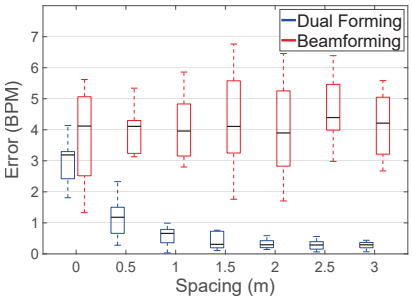


Figure 14: Accuracy at various radical spacings

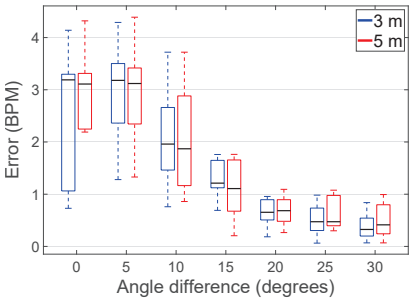


Figure 15: Accuracy along various angle differences of incidence

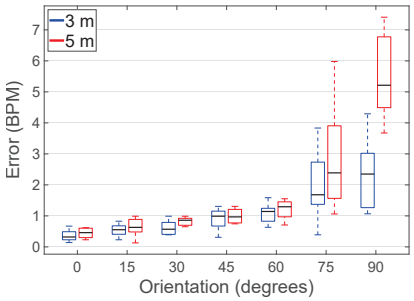


Figure 16: Accuracy at different orientations

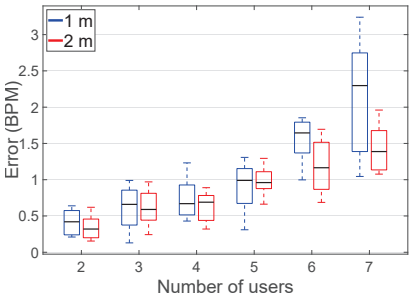


Figure 17: Accuracy at different numbers of users

Monitoring at different orientations: *DF-Sense achieves an average measuring error of less than 1.29 BPM when the user orientation range is within 120 ° (60 ° for each side).* Fig. 16 shows the heartbeat measurement error when subject is at different orientations toward the transceiver. Results show that our system has a good orientation tolerance of up to 120 °, although the measurement error slightly increases with orientation increases. We also observe that an unacceptable error occurs when the orientation is greater than 150 ° (75 ° for each side). This is reasonable since the dynamic reflection from the body side is relatively weak so that *MUS²IC* may fail to localize the subject accurately. The tiny displacement change of the chest caused by the heartbeat from the body side will lead to a lower SNR, making it more difficult to extract the heartbeat pattern from the original signal. Overall, the orientation tolerance of DF-Sense should be sufficient in practice.

Monitoring multiple users: *DF-Sense achieves excellent performance in monitoring heartbeats of multiple subjects at the same time.* To evaluate the capacity of DF-Sense in monitoring heartbeats of multiple subjects, in this experiment, we ask a number of subjects to randomly sit at different positions in a room. In addition, the spacing of adjacent subjects is set to 1 m and 2 m, respectively, in order to study the impact of spacing. Fig. 17 shows that DF-Sense supports simultaneous measurements of up to 5 subjects and 6 subjects for the spacings of 1 m and 2 m, respectively. Number of subjects exceeding 6 may result in performance degradation. This is due to heartbeat signals drowned into the accumulation of dynamic signals caused by irrelevant activities. We also observe that the performance under the spacing of 1 m is slightly worse than that under the spacing of 2 m. This is reasonable since the mutual interference at a closer range is stronger than that at a longer range. In our future work, we will develop an advanced denoising algorithm that minimizes mutual interference, enabling the inclusion of a larger number of subjects simultaneously.

Accuracy in localization: *MUS²IC achieves a median localization error of 0.025 m in 1-D scenario and 0.047 m in 2-D scenario.* In this experiment, we evaluate the localization accuracy of *MUS²IC* in 1-D and 2-D scenarios. The localization under the 1-D scenario refers to the measurement of path length. We use the Euclidean distance as a localization metric. For each scenario, we repeat the experiment at different locations 500 times. Fig. 18 shows the cumulative distribution function (CDF) of localization errors. For both scenarios, the third quartile falls below 0.08 m, demonstrating that *MUS²IC* is capable of accurately locating subject.

4.3 System Robustness

Impact of appliance: *DF-Sense is robust to interference from household appliances.* To evaluate the impact of appliances, in this experiment, we place five different types of household appliances (i.e., microwave oven, refrigerator, fan, juicer, and washer) 1 m and 2 m away from the subject, respectively. As shown in Fig. 19, DF-Sense performs well in the presence of microwave, refrigerator, and juicer, but experiences minor error increases for fan and washer. This may be due to non-negligible dynamic reflection caused by the large displacement of the fan blades and the washer. Overall, DF-Sense

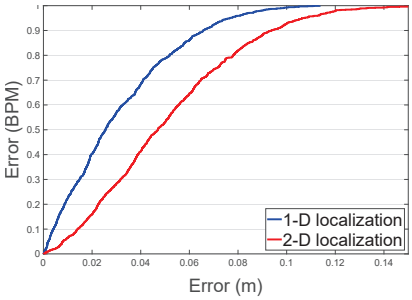


Figure 18: CDF of localization errors

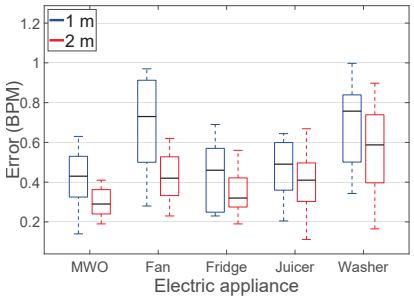


Figure 19: Effect of different types of home appliances

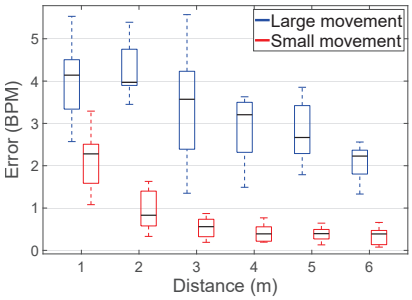


Figure 20: Impact of close-by movements

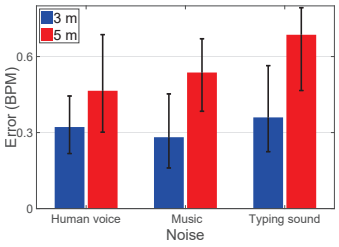


Figure 21: Impact of environmental noise

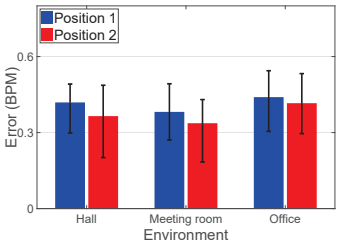


Figure 22: Impact of various environments

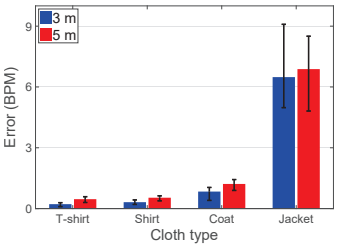


Figure 23: Effect of different types of clothes

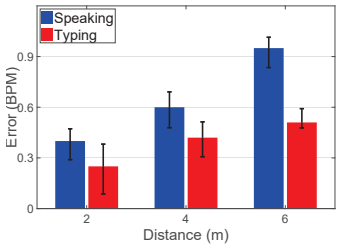


Figure 24: Effect of typing and speaking

achieves a median error of less than 0.75 BPM for all appliances at 1 m and 2 m away.

Impact of other movements: *DF-Sense achieves robust heartbeat monitoring when there exists small movement at more than 2 m from the subject.* To evaluate the impact of movements from nearby subjects, in this experiment, we ask another participant to perform both small and large movements at various distances from the subject. Note that the term "small movemen" describes small-amplitude movements (e.g., performing hand gestures), and "large movemen" describes large-amplitude movements (e.g., walking, raising legs). As shown in Fig. 20, while the small movement has little impact, large movement results in a severe error even if the subject is 6 m away. When the subject is 2 m away, the average measurement error is 0.83 BPM, and the error decreases with distance.

Impact of environmental noise: *DF-Sense is capable of monitoring heartbeat under environmental noises.* In this experiment, we simulate three types of noise, i.e., human voice, background music, and keyboard typing, to evaluate the impact of environmental noise. We ask a subject to sit 3 m and 5 m away from the transceiver. The noise source is played at 3 m away from the subject. Fig. 21 shows that the measurement errors in presence of different noise are low. This is reasonable since the frequency range of the ultrasound signals used in our system is much higher than that of the audible sound.

Impact of indoor environments: *DF-Sense is robust to different indoor environments and achieves an average error of less than 0.43 BPM for all environments.* To evaluate the robustness of DF-Sense in different environments, in this experiment, we select three indoor environments––hall, meeting room, and office. Fig. 11(b) and Fig. 11(c) show the sample experimental environments. For each environment, we ask a subject to sit at two positions––the corner and the middle of the room. Fig. 22 shows that different environments have a negligible impact on the performance of our system.

Impact of clothing type: *DF-Sense is robust to different types of clothes.* To evaluate the impact of clothing, in this experiment, we ask a subject to wear different clothes (i.e., T-shirt, business shirt, coat, and jacket) and sit 3 m and 5 m away from the transceiver, respectively. As shown in Fig. 23, DF-Sense achieves an average measurement error of less than 1.22 BPM for both distances except for wearing a jacket. We observe that with the slight increase in clothes thickness, the performance gradually degrades for both distances. This may be due to small chest displacement overwhelmed and blocked by the thick texture of clothing.

Impact of typing and speaking: DF-Sense is robust to physical activities, such as speaking and typing on a keyboard. To comprehensively investigate the effects of these two activities, we conduct experiments in which users were asked to perform both activities at varying distances from the transceiver end. We allow the users to simulate tapping movements over the air. The results, shown in Figure 24, indicate that DF-Sense achieves an average measurement error of less than 0.95 and 0.51 BPM for speaking and typing, respectively, at a distance of up to 6 m. Note that the system's performance in the speaking scenario is slightly weaker than in the typing scenario. The difference is mainly due to the spatial proximity between the mouth and the heart in the 2-D space, which results in mutual interference.

Platform	MUS^{2}IC	Dual Forming	Heart Extraction	Total
Laptop	19.69 ms	5.15 ms	1.55 ms	26.39 ms
Raspberry Pi 3B+	25.06 ms	8.12 ms	3.34 ms	36.52 ms

Table 1: Time consumption

4.4 System Latency

DF-Sense achieves small latency and it is suitable for real-time monitoring. In this experiment, we measure the processing time of DF-Sense on two types of platforms––a Raspberry Pi 3B+ and a laptop with an i7-10750H CPU and 16 GB memory. DF-Sense processes one frame consisting of 1920 samples under a sampling rate of 48 kHz. The process involves three steps––*MUS^{2}IC*, *Dualforming*, and heartbeat extraction, executed in order. As the duration of each audio frame is 40 ms, ideally, the processing latency has to be less than 40 ms to enable real-time processing. As shown in Tab. 1, DF-Sense achieves a total latency of 26.39 ms and 36.52 ms on average for the laptop and Raspberry Pi, respectively. Note that the *MUS^{2}IC* algorithm is triggered to obtain the subject location within a few seconds on system startup. Therefore, only *Dualforming* and heart extraction are involved in regular monitoring.

5 Related Work

Fine-grained acoustic sensing: Acoustic signals have the advantage of sensing fine-grained movements. AcuTe [2] and PACE [3] leverage propagation features to accurately compute time and distance. Mao et al. [15] develop an FMCW-based acoustic tracking system, achieving an accuracy of 5 mm. VSkin [27] enables an acoustic-based solution to detect finger tapping movement with high accuracy. FingerIO [20] tracks subcentimeter-level finger motion using acoustic signals. Strata [40] tracks a fine-grained finger by selecting the corresponding channel tap. LLAP [35] tracks finger movement direction and distance by exploring the phase changes of ultrasound signals with centimeter-level accuracy. In addition to finger gesture sensing, many studies leverage acoustic signals to sense respiration [19, 25, 32, 34, 38]. However, all these sensing systems work in a short range.

Long-range acoustic sensing: Many efforts have been made to extend the acoustic sensing range. Mao et al. [17] enable hand tracking within 4.5 m, achieving an average error of centimeters. DeepRange [16] develops a ranging system with a single speaker and mic, and the ranging error is within 1 cm at 4 m from the transceiver. FM-Track [12] enables tracking of a hand-sized target with an error of about 4 cm. However, the ranging resolution of these approaches is at centimeter-level, which cannot be applied for subtle motions such as heartbeat and respiration. Two recent studies, i.e., RespTracker [30] and LASense [13], increase the sensing range for respiration monitoring from 2 m to 6 m. However, they fail to monitor multiple users at different distances.

Contact-free heartbeat monitoring: Many efforts have been devoted to contact-free approaches, including vision-based [9], RF-based [1, 4, 8, 23, 39, 43–45]. Vision-based approaches use video signals collected by camera to infer the change of blood oxygen in response to heartbeat [9]. However, this approach needs good lighting and may raise privacy concerns. RF-based approaches, including FMCW radar [1, 8], millimeter wave (mmWave) radar [39], and UWB radar[4, 23, 43, 45], have shown their advantages in monitoring heartbeat. However, these solutions require dedicated and expensive devices [22, 42]. In addition, the effective sensing range for heartbeat monitoring in existing systems is insufficient for large-scale deployment of acoustic sensing application [22, 33, 42], and existing systems focus on single user only.

6 Discussion and Future Work

Array shape: Our system exhibits versatile deployment capabilities, accommodating not only linear arrays but also arrays of various shapes. For instance, in the case of a circular array, the phase of each subcarrier microphone is influenced by three critical factors: the yaw angle, pitch angle, and path length. To achieve precise calculation of an object's position in 3-D space using the *MUS^{2}IC* algorithm, we can redefine the virtual joint array to incorporate these three parameters. Subsequently, we can calculate the phase differences induced by these factors in different microphones, facilitating the constructive superposition of CFR and enabling Dual Forming. We will evaluate the system performance with a circular array in our future work.

Number of microphones: The incident angle resolution is inversely proportional to the number of microphones, while the distance resolution for the targeted region is directly affected by the number of subcarriers. Theoretically, the effective mitigation of interference from sources with distinct distances can be achieved even when the number of microphones is limited, provided that a sufficient number of subcarriers is available. Subsequently, our plan is to empirically investigate the performance of the system with different numbers of microphones to corroborate the inference.

Occlusion issue: Acoustic waves experience a significant decay when propagating through thick obstacles, the system performance may be drastically reduced in an occluded scenario.

Therefore, in this scenario, we may resort to other wireless signals such as WiFi, LoRa, and mmWave which we leave for our future work.

Future application potential: The fundamental principle of our system lies in employing multiple subcarriers and receivers to facilitate the enhancement of SSNR. This strategy can be seamlessly extended to other similar sensing modalities, including WiFi and Long-term Evolution (LTE) signals. We believe that the proposed methods can benefit sensing technologies that utilize OFDM-based signals in the future.

7 Conclusion

In this paper, we present DF-Sense, a *Dualforming* based multi-user acoustic sensing system for heartbeat monitoring in a range of 10 m in 2-D space. We take a systematic approach to address the challenges associated with high-resolution acoustic sensing in distance, and sense heartbeat for multiple subjects in a 2-D scenario. Evaluation results show that DF-Sense is capable of sensing heartbeat accurately and robustly in different home scenarios.

Acknowledgments

This research is supported by NSFC (Grant No. 62102006, 62022005, 62272010 and 62061146001, A3 Project 62061146001).

References

[1] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C Miller. 2015. Smart homes that monitor breathing and heart rate. In *Proceedings of ACM CHI*.

[2] Chao Cai, Zhe Chen, Henglin Pu, Liyuan Ye, Menglan Hu, and Jun Luo. 2020. AcuTe: Acoustic Thermometer Empowered by a Single Smartphone. In *Proceedings of ACM SenSys*.

[3] Chao Cai, Henglin Pu, Peng Wang, Zhe Chen, and Jun Luo. 2021. We hear your pace: Passive acoustic localization of multiple walking persons. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–24.

[4] Zhe Chen, Tianyue Zheng, Chao Cai, and Jun Luo. 2021. MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing. In *Proceedings of ACM MobiCom*.

[5] Marcelo A Colominas, Gastón Schlotthauer, and María E Torres. 2014. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control* 14 (2014), 19–29.

[6] Yongjiu Du, Ehsan Aryafar, Joseph Camp, and Mung Chiang. 2014. iBeam: Intelligent client-side multi-user beamforming in wireless networks. In *Proceedings of IEEE INFOCOM*.

[7] Chao Feng, Xinyi Li, Yangfan Zhang, Xiaojing Wang, Liqiong Chang, Fuwei Wang, Xinyu Zhang, and Xiaojiang Chen. 2021. RFlens: metasurface-enabled beamforming for IoT communication and sensing. In *Proceedings of ACM MobiCom*.

[8] Chen-Yu Hsu, Aayush Ahuja, Shichao Yue, Rumen Hristov, Zachary Kabelac, and Dina Katabi. 2017. Zero-effort in-home sleep and insomnia monitoring using radio signals. *Proceedings of the ACM on Interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–18.

[9] Jingjing Hu, Yunze He, Jie Liu, Min He, and Wenjin Wang. 2012. Illumination robust heart-rate extraction from single-wavelength infrared camera using spatial-channel expansion. In *Proceedings of IEEE EMBC*.

[10] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*.

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

[12] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of ACM Sensys*.

[13] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. 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 (2022), 1–27.

[14] Xiang Li, Shengjie Li, Daqing Zhang, Jie Xiong, Yasha Wang, and Hong Mei. 2016. Dynamic-music: Accurate Device-free Indoor Localization. In *Proceedings of ACM UbiComp*.

[15] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-precision Acoustic Motion Tracking. In *Proceedings of MobiCom*.

[16] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. 2020. DeepRange: Acoustic Ranging via Deep Learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.

[17] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-based room scale hand motion tracking. In *Proceedings of ACM MobiCom*.

[18] Andreas F Molisch, Vishnu V Ratnam, Shengqian Han, Zheda Li, Sinh Le Hong Nguyen, Linsheng Li, and Katsuyuki Haneda. 2017. Hybrid Beamforming for Massive MIMO: A Survey. *IEEE Communications Magazine* 55, 9 (2017), 134–141.

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

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

[21] Emiel Por, Maaike van Kooten, and Vanja Sarkovic. 2019. Nyquist–Shannon sampling theorem. *Leiden University* 1 (2019), 1.

[22] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In *Proceedings of IEEE INFOCOM*.

[23] Muhammad Salman Raheel, James Coyte, Faisel Tubbal, Raad Raad, Philip Ogunbona, Christopher Patterson, and Dana Perlman. 2019. Breathing and heartrate monitoring system using IR-UWB radar. In *Proceedings of IEEE ICSPCS*.

[24] A Rodríguez Valiente, A Trinidad, JR García Berrocal, C Górriz, and R Ramírez Camacho. 2014. Extended high-frequency (9–20 kHz) audiometry reference thresholds in 645 healthy subjects. *International journal of audiology* 53, 8 (2014), 531–545.

[25] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of ACM MobiCom*.

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

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

[28] Angus E Taylor. 1952. L'Hospital's rule. *The American Mathematical Monthly* 59, 1 (1952), 20–24.

[29] María E Torres, Marcelo A Colominas, Gaston Schlotthauer, and Patrick Flandrin. 2011. A complete ensemble empirical mode decomposition with adaptive noise. In *Proceedings of IEEE ICASSP*.

[30] Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. 2021. RespTracker: Multi-user Room-scale Respiration Tracking with Commercial Acoustic Devices. In *Proceedings of IEEE INFOCOM*.

[31] Anran Wang, Dan Nguyen, Arun R Sridhar, and Shyamnath Gollakota. 2021. Using smart speakers to contactlessly monitor heart rhythms. *Communications biology* 4, 1 (2021), 1–12.

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

[33] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–24.

[34] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.

[35] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-Free Gesture Tracking Using Acoustic Signals. In *Proceedings of ACM MobiCom*.

[36] Zhi Wang, Fusang Zhang, Siheng Li, and Beihong Jin. 2021. Exploiting Passive Beamforming of Smart Speakers to Monitor Human Heartbeat in Real Time. In *Proceedings of IEEE GLOBECOM*. IEEE.

[37] Zhaohua Wu and Norden E Huang. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis* 1, 01 (2009), 1–41.

[38] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of ACM MobiSys*.

[39] Zhicheng Yang, Parth H Pathak, Yunze Zeng, Xixi Liran, and Prasant Mohapatra. 2016. Monitoring vital signs using millimeter wave. In *Proceedings of ACM MobiHoc*.

[40] Sangki Yun, Yichao Chen, Huihuang Zhang, Lili Qiu, and Wenguang Mao. 2017. Strata: Fined-Grained Device-Free Tracking Using Acoustic Signals. In *Proceedings of ACM MobiSys*.

[41] Youwei Zeng, Jinyi Liu, Jie Xiong, Zhaopeng Liu, Dan Wu, and Daqing Zhang. 2021. Exploring Multiple Antennas for Long-range WiFi Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–30.

[42] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can" Hear" Your Heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–24.

[43] Shujie Zhang, Tianyue Zheng, Zhe Chen, and Jun Luo. 2022. Can We Obtain Fine-grained Heartbeat Waveform via Contact-free RF-sensing?. In *Proceedings of IEEE INFOCOM*.

[44] Mingmin Zhao, Fadel Adib, and Dina Katabi. 2016. Emotion recognition using wireless signals. In *Proceedings of ACM MobiCom*.

[45] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. 2020. V2iFi: In-vehicle vital sign monitoring via compact RF sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–27.