

Article

Remote Multi-Person Heart Rate Monitoring with Smart Speakers: Overcoming Separation Constraint

Thu Tran *, Dong Ma and Rajesh Balan

School of Computing and Information Systems, Singapore Management University, Singapore 178902, Singapore; dongma@smu.edu.sg (D.M.); rajesh@smu.edu.sg (R.B.)

* Correspondence: ndttran.2019@phdcs.smu.edu.sg

Abstract: Heart rate is a key vital sign that can be used to understand an individual's health condition. Recently, remote sensing techniques, especially acoustic-based sensing, have received increasing attention for their ability to non-invasively detect heart rate via commercial mobile devices such as smartphones and smart speakers. However, due to signal interference, existing methods have primarily focused on monitoring a single user and required a large separation between them when monitoring multiple people. These limitations hinder many common use cases such as couples sharing the same bed or two or more people located in close proximity. In this paper, we present an approach that can minimize interference and thereby enable simultaneous heart rate monitoring of multiple individuals in close proximity using a commonly available smart speaker prototype. Our user study, conducted under various real-life scenarios, demonstrates the system's accuracy in sensing two users' heart rates when they are seated next to each other with a median error of 0.66 beats per minute (bpm). Moreover, the system can successfully monitor up to four people in close proximity.

Keywords: heart rate monitoring; acoustic-based sensing; smart speakers; multi-person tracking; spatial localization; FMCW



Citation: Tran, T.; Ma, D.; Balan, R. Remote Multi-Person Heart Rate Monitoring with Smart Speakers: Overcoming Separation Constraint. *Sensors* **2024**, *24*, 382. <https://doi.org/10.3390/s24020382>

Academic Editors: Ming Yan, Chunguo Li and Chien Aun Chan

Received: 29 November 2023

Revised: 31 December 2023

Accepted: 5 January 2024

Published: 8 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Heart rate is one of the key indicators used to evaluate individuals' overall health. For example, changes in heart rate can be used to assess the state of the nervous system [1] and are used as a stress indicator [2]. In addition, rapid resting heart rate has been suggested as a risk factor for cardiovascular mortality [3], and heart rate dynamics are used to infer sleep stages, as these dynamics are more conspicuous during the later Rapid Eye Movement (REM) period [4]. Although traditional cardiac monitoring approaches such as electrocardiogram (ECG) can achieve high accuracy, these are contact-based approaches that are expensive, uncomfortable to wear for prolonged periods, and cumbersome to set up and use. Thus, they are not suitable for home monitoring or for patients with skin allergies or burn injuries, where skin-contact sensors are not feasible.

Recent advancements in remote sensing have suggested various ways to leverage radio frequency (RF) signals to monitor heart rate without any sensors or probes attached to the skin. These include Frequency Modulated Continuous Wave (FMCW) radar [5–7], WiFi [8,9], and millimeter wave [10,11]. In addition, acoustic signals [12–15] have been used to extract respiration and heart rate in a contactless manner. These approaches use a speaker to emit inaudible high-frequency waves, typically above 18 kHz, use microphones to capture the reflected signal after it bounces back from targets in the nearby areas, then analyze the reflected signal to extract valuable information such as the respiration rate and HR as well as the distance and the angle of the target relative to the transceiver.

Even with extensive prior work in remote heart rate sensing, heart rate detection has focused mostly on single sensing. Only a few solutions target the sensing of multiple heart

rates [6,16], and all of these only work under conditions that restrict their practical adoption. In particular, [6] requires the subjects to be 1 to 2 m separated away from each other when using RF signals; [12,16] require at least 40 to 50 cm separation as well as a 10° angular difference between each subject to achieve acceptable performance using acoustic signals. This leads to the infeasibility of tracking multiple people in real-life scenarios, such as sitting side by side or lying next to each other on the same bed, as shown in Figure 1.

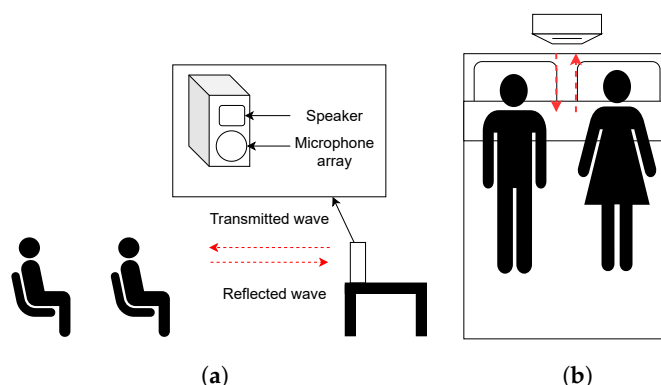


Figure 1. Practical scenarios of multi-person heart rate monitoring: (a) two people sitting in line and (b) Two people sharing a bed.

In this paper, we aim to achieve multiple heart rate monitoring in such practical scenarios using a commodity smart speaker, the MiniDSP UMA-8-SP USB mic array [17], which has the same layout as the Amazon Echo Dot [18]. A smart speaker is considered as an appealing platform for contactless and acoustic-based heart rate monitoring for two key reasons. First, smart speakers have become increasingly prevalent in home environments, where they provide various voice-based services. Second, commercial smart speakers usually incorporate a microphone array design to deliver high-quality audio services. These microphone arrays offer high-resolution signals for active acoustic sensing that have been demonstrated to improve heart rate detection performance [13,19].

However, detecting and differentiating heart rates poses challenges when multiple people are in close proximity, as their acoustic reflection signals interfere with one another due to increased multipath interference. In this paper, we present an acoustic-based system that can extract multiple heart rates as well as their location with no separation requirement. To achieve this, we first separate users at different distances by processing the reflected FMCW signals. For each particular distance, we apply a Fast Fourier Transform (FFT) to extract the frequency of their corresponding heart rate. Then, we propose an algorithm to eliminate the interference and amplify the heart rate signal. Next, to further identify users' spatial information, we leverage the microphone array equipped in the smart speaker and apply beamforming to obtain their azimuth angles.

To assess the effectiveness of our approach, we conducted a study approved by the Institutional Review Board (IRB) under various realistic conditions. Using data collected from ten couples, our system shows the possibility of accurate heart rate detection when two individuals are positioned in close proximity, with a median error of 0.66 beats per minute (bpm). Moreover, we demonstrate the scalability of our technique by successfully identifying the heart rates and locations of four individuals seated next to each other.

2. Related Work

2.1. Contact-Based Heart Rate Monitoring

Contact-based heart rate monitoring approaches typically require sensors attached to the human body during the measurement process. These sensors are usually ECG or photoplethysmography (PPG). ECG is the traditional and “gold standard” contact-based technique to measure cardiac signals [20,21]. When performing ECG, electrodes are attached to the patient's skin at several spots, such as the chest or arms to record the

electrical impulses of the heart when they travel across the electrodes. Both the strength and the timing of the pulses are monitored. Although ECG is highly accurate and widely used to diagnose heart-related diseases, this method is not suitable for home monitoring as it requires well-trained technicians and is done merely in clinical settings. The most common alternative monitors for home use are pulse meters or wrist bands [22,23]. They typically use PPG sensors, which contains a light source and a photodetector. The light source shines a green light onto the skin, and the changes in the light reflected back from the skin are monitored by the photodetector to extract associated heart pulses. Although this method is more convenient than ECG, it is still a contact-based approach, and is inapplicable to people with skin allergies or burn injuries.

2.2. RF-Based Heart Rate Monitoring

The past few years have witnessed growth in the number of studies that monitor vital signs, including respiration and heart rate, in a contactless manner through the use of RF waves. For example, well-studied RF-based approaches including broad-band FMCW radar [6,7], WiFi signals [8,9], RFID [24], and millimeter wave [11] have all shown accurate respiration rate detection even in the presence of more than one person. In addition to respiration sensing, [25–27] have demonstrated the possibility of heart rate detection using these radar technologies. However, because heartbeat-induced chest displacement is subtle and always drowned out by respiration, only a few papers [6,11,28] have provided results for heart rate monitoring of more than one person; all of these papers require the subjects to be separated by a minimum of 1 to 2 m, making it infeasible to monitor two people sitting next to each other or sharing a bed. Furthermore, such RF-based systems require dedicated hardware, which generally leads to high cost.

2.3. Acoustic-Based Heart Rate Monitoring

Recently, the rise in popularity of acoustic-enabled mobile devices such as smartphones and smart speakers has introduced a potential alternative to RF-based heart rate measurement. Both RF and acoustic signals employ similar techniques to monitor respiration and heart rate. For instance, [29,30] developed an approach that utilizes the in-built speaker and microphone in smartphones and smart speakers to monitor the respiration rate in a single user. Similarly, [15] proposed a method for remotely monitoring heart rate using smartphones, while [13,14] suggested an approach using a smart speaker prototype with a single speaker–microphone pair. In order to expand the sensing range of these systems, [19] incorporated multiple microphones and a beamforming algorithm. However, due to signal interference, these studies remain unable to simultaneously monitor multiple people. Two recent studies [12,16] introduced novel beamforming algorithms that address this interference issue, enabling the monitoring of multiple heart rates; nevertheless, these algorithms require individuals to be separated by 40 to 50 cm and to be at least 10° apart. In our proposed approach, we first utilize a heatmap to identify users' distances and heart rates, then apply beamforming on multiple microphones using the known distances and heart rates. This allows us to obtain the localization of the users while eliminating the separation requirement. A summary of the differences between our proposed approach and existing studies is presented in Table 1.

Table 1. Related works on heart rate monitoring using acoustic-based approaches. Each study's median error is reported at its corresponding distance. In [13–15,19], the focus was on single users, while in [12,16] the authors monitored multiple users, for which the separation requirement is listed under *Separation*.

Study	Median Error (bpm)	Multiple Sensing	Separation	Distance (m)
[15]	0.6	No	Not applicable	0.3
[14]	0.75	No	Not applicable	0.2
[19]	0.6	No	Not applicable	0.6
[13]	1	No	Not applicable	0.4–0.6
[16]	0.8	Yes	40 cm and 10°	3
[12]	1.18	Yes	50 cm	3
Our approach	0.9	Yes	Not required	3

3. FMCW Background and Key Challenge

3.1. FMCW Background

In our proposed system, the speaker emits a sequence of FMCW chirps that are continuously modulated in frequency over a predefined time period. The reflected chirp is received by the microphone array, and the time difference between the transmitted and the reflected chirps indicates the range R between the transceiver and the object. For example, the blue line in Figure 2 shows the transmitted chirp with the frequency varying according to the sweep time T . The frequency of one single chirp is denoted as $f(t) = f_0 + \frac{B}{T}t$, where f_0 and B indicate the start frequency and the bandwidth, respectively.

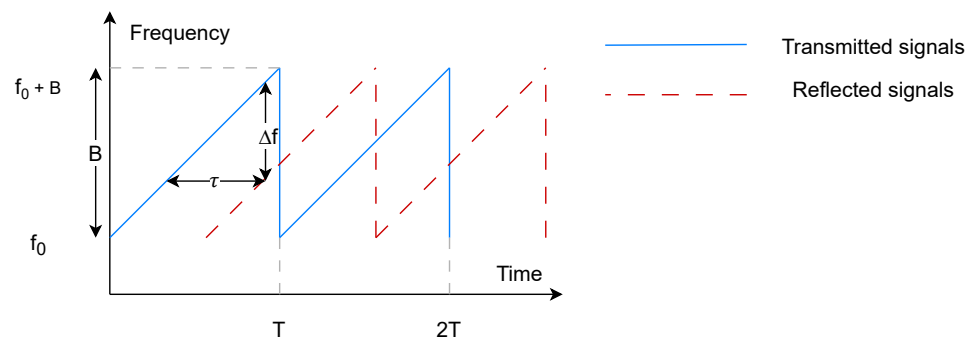


Figure 2. Transmitted and reflected FMCW signals.

Therefore, a single chirp is represented as

$$x(t) = \cos\left(2\pi \int f(t)dt\right) = \cos\left(2\pi\left(f_0t + \frac{Bt^2}{2T}\right)\right). \quad (1)$$

The reflected chirp from a target to the receiver is the delayed version of the transmitted chirp, as shown in Figure 2, and is expressed as

$$x'(t) = \alpha \cos\left(2\pi\left(f_0(t - \tau) + \frac{B(t - \tau)^2}{2T}\right)\right), \quad (2)$$

where α and τ refer to the signal amplitude attenuation factor and time delay, respectively. With a moving target, we have $\tau = \frac{2r(t)}{c} = \frac{2(R+vt)}{c}$, where c is the speed of sound 343 m/s, v is the speed of the moving subject, and $r(t)$ is the distance of the moving subject by time. For a static object, $\tau = \frac{2R}{c}$; to compute the range R , we multiply the transmitted signal $x(t)$ by the received signal $x'(t)$. The mixed signal $x_m(t)$ is then represented as follows:

$$\begin{aligned}
x_m(t) &= x(t) \cdot x'(t) \\
&= \frac{\alpha}{2} \left[\cos \left(2\pi \left(f_0 \tau - \frac{B(\tau^2 - 2t\tau)}{2T} \right) \right) + \cos \left(2\pi \left(f_0(2t - \tau) + \frac{B(2t^2 - 2t\tau + \tau^2)}{2T} \right) \right) \right]. \quad (3)
\end{aligned}$$

The mixed signal consists of two terms. By taking the derivative of the phase by t , we have the frequency of the first term, which is a constant $\Delta f = \frac{B}{T} \tau = \frac{2BR}{cT}$. This implies that every distance R maps to a specific frequency Δf . The second term is a function of t with high frequency, and can be removed by a low-pass filter. In the end, after the multiplication and low pass filter, we have

$$x_{mf}(t) = \frac{\alpha}{2} \cdot \exp \left[j2\pi \left(f_0 \tau - \frac{B(\tau^2 - 2t\tau)}{2T} \right) \right]. \quad (4)$$

By transforming the frequency of $x_{mf}(t)$, we have

$$R = \frac{cT}{2B} \Delta f. \quad (5)$$

Given a typical audio sensing bandwidth of $B = 5k$ Hz (usually from 18 kHz to 23 kHz), according to [30], the resolution of R is $\delta R \geq \frac{cT}{2B} \delta f = \frac{cT}{2B} \cdot \frac{1}{T} = \frac{343}{2 \cdot 5000} = 3.43$ cm.

Although this resolution is sufficient to monitor centimeter-level human breathing, it is much lower than the heartbeat-induced chest displacement Δd , which is approximately 0.1–0.5 mm. Therefore, heart rate has been measured using phase-based methods [6,13–15]. It has been demonstrated that the minute chest displacement Δd can cause phase change in $x_{mf}(t)$ up to 18.9° . Specifically, $x_{mf}(t)$ can be expressed as follows.

$$\begin{aligned}
x_{mf}(t) &= \frac{\alpha}{2} \cdot \exp \left[j \left(2\pi f_0 \tau - \frac{\pi B \tau^2}{T} + \frac{2\pi t \tau B}{T} \right) \right] \\
&\approx \frac{\alpha}{2} \cdot \exp \left[j \left(2\pi f_0 \tau + \frac{2\pi t \tau B}{T} \right) \right] \\
&= \frac{\alpha}{2} \cdot \exp \left[j \left(\frac{4\pi f_0}{c} r(t) + 2\pi \Delta f t \right) \right] \\
&= \frac{\alpha}{2} \cdot \exp \left[j \left(\frac{4\pi f_0}{c} \Delta d + 2\pi \Delta f t \right) \right]
\end{aligned} \quad (6)$$

As such, if the chest displacement caused by a heartbeat is 0.5 mm, the phase change is calculated as

$$\frac{4\pi f_0}{c} \Delta d = \frac{4\pi \cdot 18,000 \cdot 0.0005}{343} = 0.105\pi = 18.9^\circ. \quad (7)$$

Similar to prior works [6,29], in this paper we use distance to separate users. As each Δf indicates one R , we generate these frequency bins (which can be converted into distance bins) by applying FFT on x_{mf} , then analyzing the phase at each distance bin to determine the heart rate.

3.2. Key Challenge: Signal Interference

The reflected signals at the microphone array undergo interference when more than one person is present. Figure 3 is a heart rate–distance heatmap generated from the reflected signals, showing the heart rates and distances of two people lying down next to each other; the device is placed above their heads, with brighter points indicating higher power. There are two heart rates of 72 and 67 bpm annotated in the figure; however, they are not visibly recognizable due to interference effects.

As shown in the figure, two types of effects are observed, which we name the *distance effect* and *frequency effect*. The distance effect happens when an object or person is located at a certain distance; it makes all the frequency power at that distance higher, as the frequency power is proportional to the power of the reflection signal. This signal is shown as bright

columns in Figure 3). In the heatmap, it can be observed that this effect happens mainly at 0.5 to 1 m, which is the location of the two users.

The frequency effect, depicted as bright rows in Figure 3, is caused by multipath reflections bouncing off walls or furniture and arriving at the receiver with increased arrival time. This multipath effect has been well studied in the literature [31,32]. Despite the longer travel time, these reflections carry the same frequency information and are linearly correlated to the directly reflected wave, resulting in equivalent frequencies spanning a wide range of distances. Multipath signals have lower power than true reflections due to power loss over distance. We note that while these two effects can occur with a single person, they become more severe with multiple people, as interference is more likely. Our solution to this problem is proposed in Section 4.

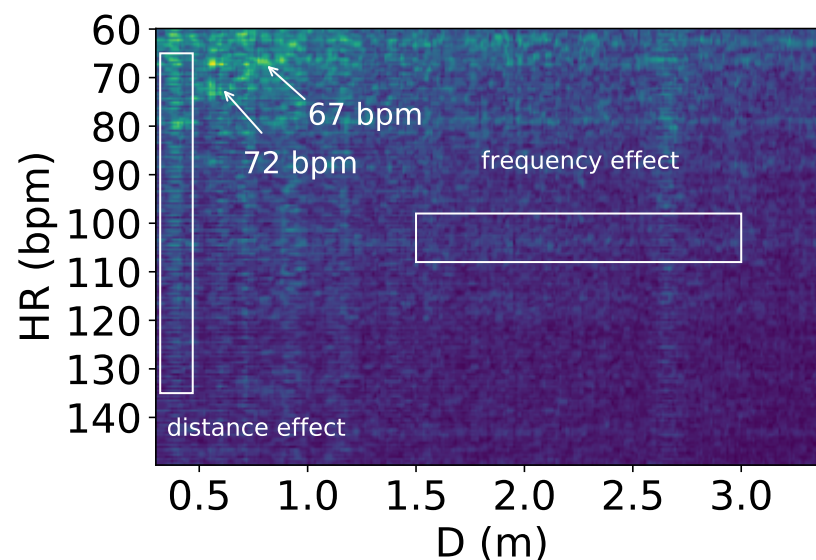


Figure 3. Heart rate–distance heatmap showing heart rates and interference from the reflected signals. The x-axis represents the distance D from the users to the device. In the figure, the two people with heart rates of 72 and 67 bpm located in front of the device cannot be distinguished visually from the heatmap.

4. System

4.1. System Overview

Our proposed system uses a commercial speaker and circular seven-microphone array with the same microphone layout and sensitivity as a commodity Amazon Echo Dot [18] (MiniDSP UMA-8-SP USB mic array [17]) as an FMCW transceiver. The speaker emits FMCW signals and the microphone array captures the signal reflected by the user, allowing all heart rates, distances, and angles of users to be identified. As shown in Figure 4, our system has four main modules:

- **Signal Processing:** This module processes raw reflected signals received by the microphones, removes noise and frequencies outside the range of f_0 and $f_0 + B$, and performs a mix operation on each chirp to generate the heart rate–distance heatmap.
- **Interference Removal:** The generated heatmap produced in the previous step is prone to distance and frequency interference effects. This interference is canceled through a two-step algorithm in order to highlight the heart rate signals.
- **Blob Detection:** Next, users' heart rates and distances are detected by applying a blob detection algorithm to the heatmap.
- **Beamforming:** Finally, beamforming is applied to detect each user's azimuth angle based on their distance and heart rate.

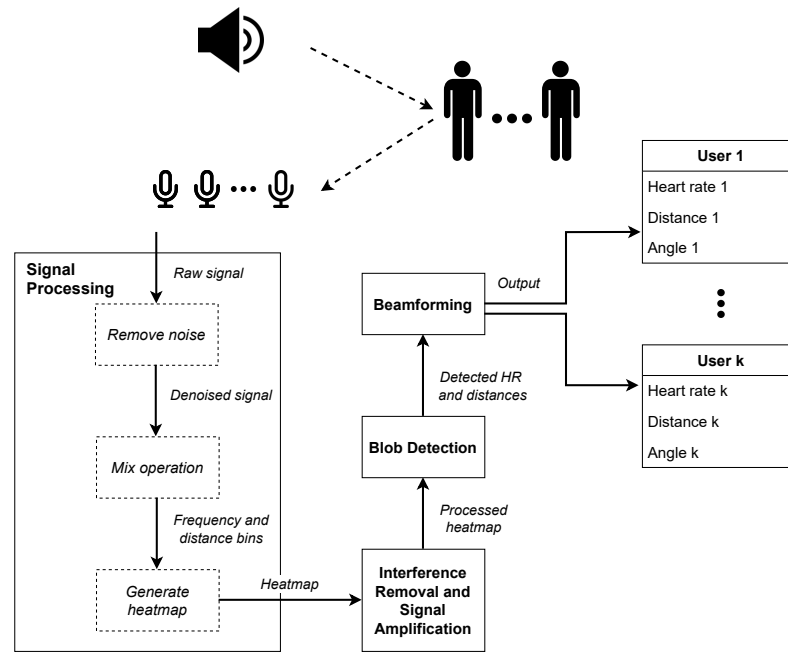


Figure 4. Overview of system for detecting the heart rates of k users.

4.2. Signal Processing

Remove Noise: The raw signal received by each microphone is filtered using a bandpass filter between f_0 to $f_0 + B$ to remove ambient noise such as human laughter and music, which are far lower than the operating FMCW frequency $f_0 \geq 18k$ Hz [33], as such noise has little impact on the system.

Mix Operation: When the raw signal has been processed, a mix operation is applied to each received chirp by multiplying them by the transmitted chirp (Equation (3)). As explained in Section 3, Δf is proportional to the distance of the target and is obtained by performing FFT on one chirp. The frequency bins after FFT are converted into distance bins using Equation (5).

In the FFT, the frequency resolution is $\frac{F_s}{N}$, where F_s is the sampling rate of the signal and N is the number of datapoints. Our system employs a sampling rate of $F_s = 48k$ Hz and a chirp length of $T = 0.04$ s. Therefore, the frequency resolution is $\frac{F_s}{T \cdot F_s} = \frac{48,000}{0.04 \cdot 48,000} = 25$ Hz, which is converted to a *distance resolution* of 3.43 cm that is sufficient to differentiate users even when they are next to each other.

Generate Heatmap: The heart rate can be extracted using the phase changes of each distance bin over time. More particularly, we consider all distance bins within the device's working range to find the bin that contains the heartbeat signal. To illustrate this, we collected data with a subject located 1.08 m away from the speaker. Figure 5a shows the amplitude changes of the distance bins ranging from 0.58 m to 2 m. It can be observed that the signal at 1.08 m shows a periodic pattern from the user's breathing and heartbeats, while no vital signs can be seen at 0.58 m and 2 m. Figure 5b plots the frequencies for the 0.58 m, 1.08 m, and 2 m distance bins. These figures show that the bin with the highest amplitude, 0.58 m, is not guaranteed to contain heartbeat signals. This is because the distance effect (as the multipath) can cause the reflected signal to be stronger than the direct reflected signal. This phenomenon has been investigated in prior work [31,32].

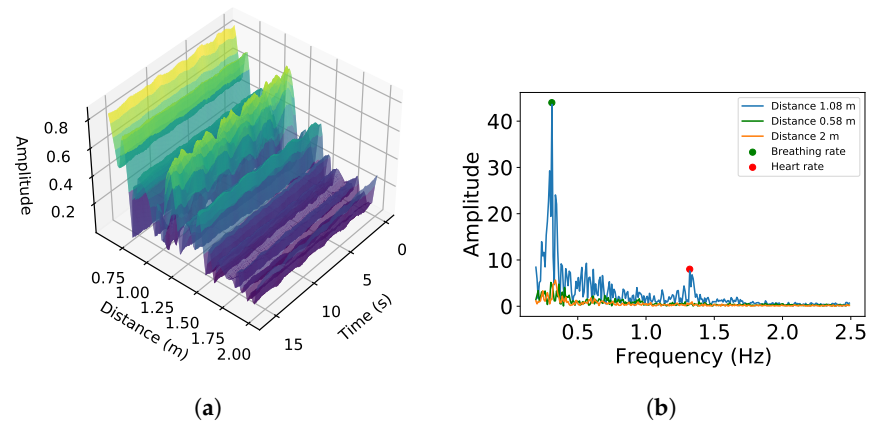


Figure 5. Amplitude changes and frequency domain of distances: (a) FFT amplitude changes by distance and (b) breathing and heart rate obtained by applying FFT at 1.08 m.

Because we need to analyse the phase change of each distance bin, a chirp length of $T = 0.04$ s leads to a $F_{sh} = \frac{1}{0.04} = 25$ Hz sampling rate for the heart rate signal. Therefore, the frequency resolution of this FFT for heart rate is computed as $\frac{F_{sh}}{N} = \frac{25}{25 \cdot 60} = \frac{1}{60} \approx 0.0167$ Hz, where 60 (s) is the window length of the signal to which FFT is applied. This frequency resolution is equivalent to $0.0167 \cdot 60 = 1$ bpm. To enhance this resolution we zero-pad the signal with 4096 samples prior to the FFT. This interpolation can yield a higher display resolution, in this case, $F_{sh} = \frac{25}{25 \cdot 60 + 4096} \approx 0.004$ Hz ≈ 0.3 bpm. We note that while this approach does not truly improve the frequency resolution, it provides a smoother FFT output, which allows peak-picking algorithms to achieve better accuracy.

Finally, this module produces a heart rate–distance heatmap. If more than one microphone is used, the heatmaps from each microphone are stacked into a single map by averaging their amplitude.

4.3. Interference Removal and Heart Rate Signal Amplification

The stacked heart rate–distance heatmap generated by the last step is affected by noise from the distance and frequency effects (see Section 3.2).

To remove these unwanted effects, we apply L1-normalization, a normalization technique that modifies the dataset values to ensure that the sum of the absolute values in each row always adds up to 1, to all the rows (i.e., heart rates) and then all the columns (i.e., distances). L1-normalization balances out those cases when entire columns or rows have similar amplitude while preserving the relative ratio of rows and columns that contain heartbeat signals. Next, we apply Gaussian smoothing to the heatmap to highlight the heart rates and remove noise. The algorithm is described in Algorithm 1.

Algorithm 1: Remove interference and amplify heart rate signals

Input: Heatmap S , with n rows and m columns

Output: Interference-free and amplified heatmap S

1 **Function** RemoveInterferenceAndAmplify(S, n, m):

 /* Step 1

*/

2 **for** $i \leftarrow 0$ to $n - 1$ **do**

3 $S[i, :] = \text{Normalize}(S[i, :])$

 /* Step 2

*/

4 **for** $i \leftarrow 0$ to $m - 1$ **do**

5 $S[:, i] = \text{Normalize}(S[:, i])$

6 **return** S

Figure 6b,c shows these steps to remove the effects, while Figure 6d shows the Gaussian smoothing used to highlight the heart rates.

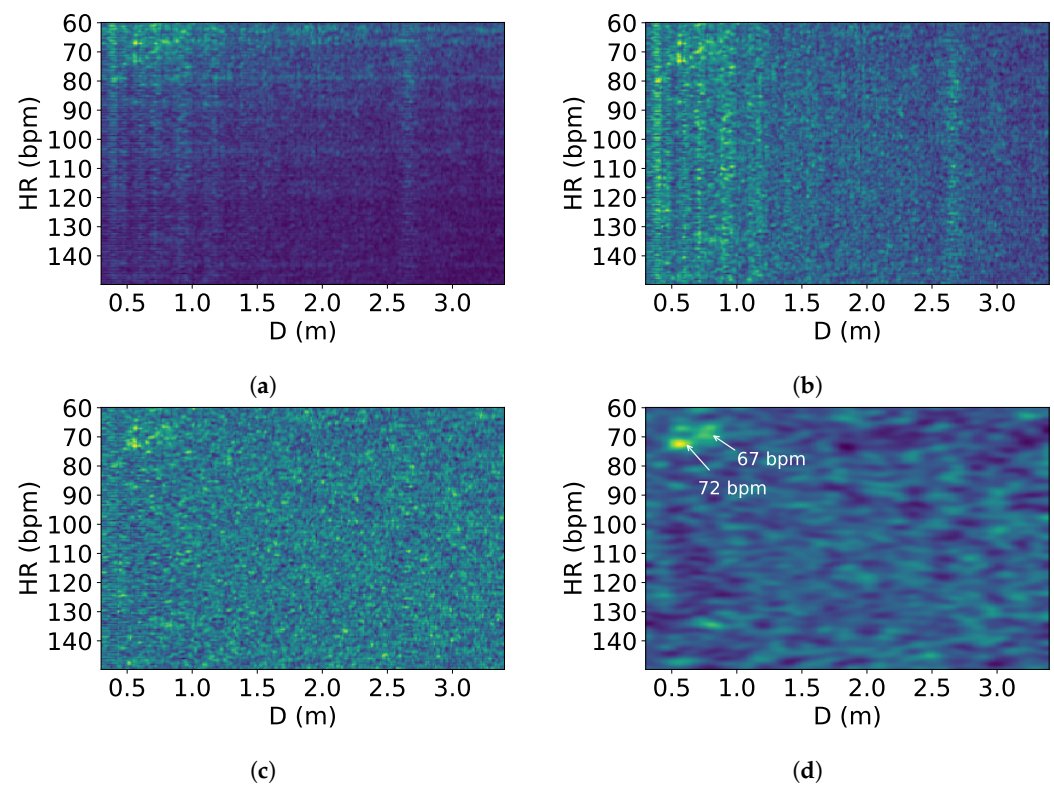


Figure 6. Interference removal: (a) original heatmap S; (b) heatmap S after step 1; (c) heatmap S after step 2; (d) smoothed heatmap S.

4.4. Blob Detection

A blob is a set of adjacent pixels that share common traits such as brightness or color. Because people can occupy many points in both distance and frequency, we detect the top brightest blobs in the heatmap instead of the top highest peaks. In this module, the input image is the heatmap from the previous step and the blobs to be detected are the elliptical bright spots on the image's dark background that indicate users' distances and heart rates. We apply the Laplacian of Gaussian (LoG) [34] as the primary blob detection method. With the blobs detected, we find the top k brightest blobs by calculating each blob's mean value in the frequency range of 0.8 Hz to 2.5 Hz, with k being the number of users. Figure 7 shows detected distances and heart rates when there are three people in front of the device.

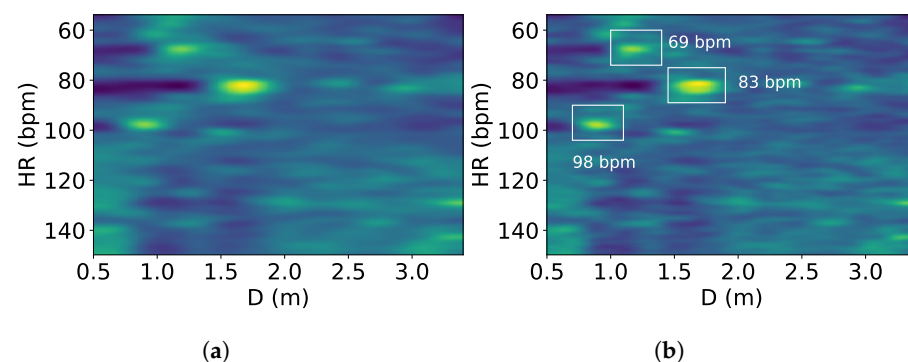


Figure 7. Three heart rates of 69 bpm, 83 bpm, and 98 bpm: (a) heatmap with three people and (b) the three brightest blobs.

4.5. Beamforming

To further identify users' locations in space, we apply digital beamforming on a circular microphone array [35] (Figures 8 and 9a) to obtain their azimuth angles from the known heart rates and distances extracted in the last module. According to Equation (6), we can rewrite x_{mf} using the n th distance bin and i th chirp:

$$x_{mf}(i, n) = \frac{\alpha}{2} \cdot \exp \left[j \left(\frac{4\pi f_0}{c} r(f(n) + g(i)) + \frac{2\pi\tau B}{T} (f(n) + g(i)) \right) \right]$$

where f and g are the function that linearly converts the n^{th} distance bin and i^{th} chirp, respectively, into time t in seconds. Specifically, $g(i) = T \cdot (i - 1)$, $i \geq 1$ and $f(n) = 2 \frac{\text{ToDistance}(n)}{c}$, $n \geq 1$, where $\text{ToDistance}(n)$ is a function to convert distance bin n to distance in meters.

Based on the azimuth angle θ provided by the circular microphone array in Figure 8, we project the source onto the x-y plane to obtain $\varphi = \frac{\pi}{2}$ with channel l ; then, x_{mf} can be expressed as

$$x_{mf}(l, i, n) = \frac{\alpha}{2} \cdot \exp \left[j \left(\frac{4\pi f_0}{c} r(f(n) + g(i)) + \frac{2\pi\tau B}{T} (f(n) + g(i)) + 2\pi \frac{r_0 \cdot \cos(\theta - \Theta(l))}{c} \right) \right] \quad (8)$$

where r_0 is the radius of the circular array, θ is the azimuth angle of the target, and $\Theta(l) = \frac{2\pi}{L}l$ is the relative angle at microphone l . When the target is static, we have

$$x_{mf}(l, i, n) = \frac{\alpha}{2} \cdot \exp \left[j \left(\frac{4\pi f_0}{c} R + \frac{2\pi\tau B}{T} (f(n) + g(i)) + 2\pi \frac{r_0 \cdot \cos(\theta - \Theta(l))}{c} \right) \right]. \quad (9)$$

Because we obtained distances and heart rates in the last step, we can represent $x_{mf}(l, i, n)$ as $x_{n,l}(i)$ when fixing the distance and channel. To obtain the heartbeats across different angles for a given distance, beamforming is performed over $L = 6$ microphones:

$$y_n(i, \theta) = S^H(\theta) X_n(i) + W(i)$$

where $S(\theta) = [s_1(\theta), \dots, s_L(\theta)]$ is the steering vector towards angle θ (with $s_l = \exp(j2\pi \frac{r_0 \cdot \cos(\theta - \Theta(l))}{c})$), $X_n(i) = [x_{n,1}(i), \dots, x_{n,L}(i)]$, and $W(i)$ is the Gaussian white noise. In our implementation, n includes a range of distances covering all users and y_n is the average of all values of n .

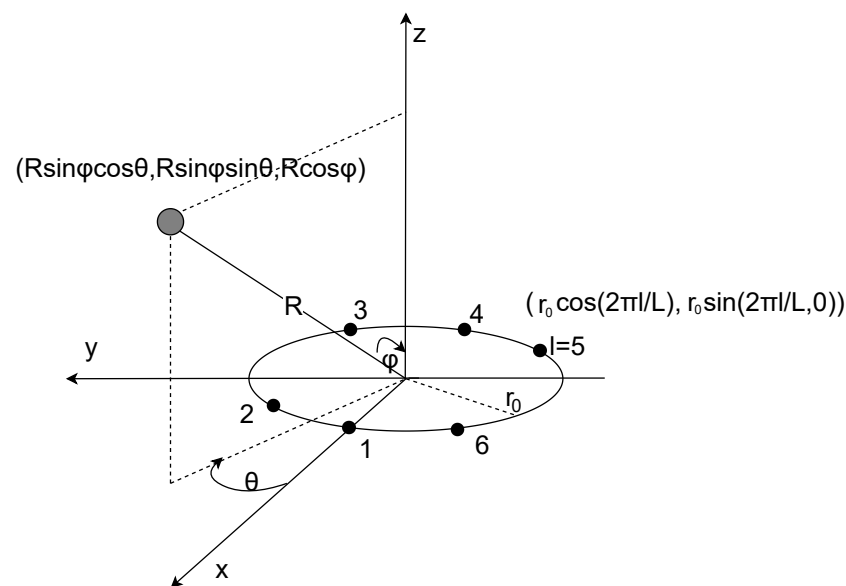


Figure 8. Source and circular microphone array with $L = 6$.

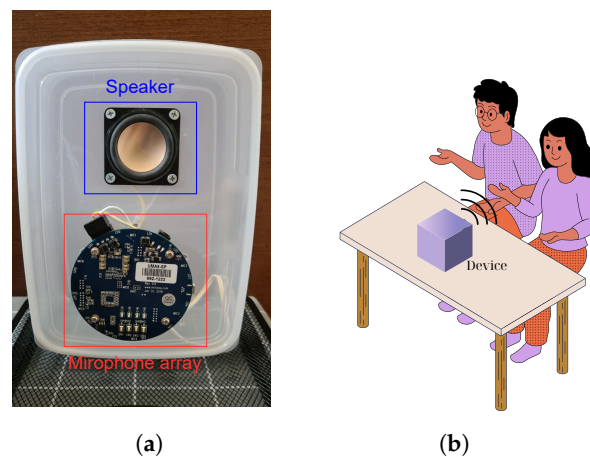


Figure 9. Device and example showing the experimental setup: (a) the device and (b) one of the experimental setups.

5. Results

In this section, we report the detailed performance of the proposed system in various realistic scenarios, including sitting, lying down with different postures, in the presence of ambience noise, and with more than two users.

5.1. Experimental Setup

We prototyped our system using an off-the-shelf seven-microphone circular array [17] connected to a speaker (PUI Audio AS05308AS-R), as shown in Figure 9a. This prototype has the same microphone layout and sensitivity as the widely used Amazon Echo Dot [18]. Table 2 provides information on all of the parameters used in our experiment. In addition, we employed Polar H10 ECG sensors [36] to collect heart rates for use as ground truth. Figure 10 illustrates the heartbeats extracted from our system and corresponding heartbeats collected by the ECG sensors. The metric used to evaluate our system was the heart rate (bpm), which we compared against heart rates captured by ECG sensors worn by participants. Figure 9b describes one of our experimental setups for two people sitting next to each other with no separation requirement.

Table 2. Parameter settings for the experiments.

Parameter	Value
Chirp frequency	18 kHz to 23 kHz
Bandwidth	5 kHz
Chirp length	0.04 s
Maximum tracking distance	$\frac{cT}{4} = 3.43$ m
Sampling rate	48 kHz
Sound pressure	45 dB(A) at 0.3 m

To evaluate our system, we recruited ten couples (nine males and eleven females in the age range of 19 to 57 and with a median age of 26) to evaluate the impact of different parameters under various daily life scenarios. All experiments were approved by the IRB. We conducted further experiments with sets of three and four participants to verify that our proposed system can accurately monitor the heart rates of more than two people in close proximity.

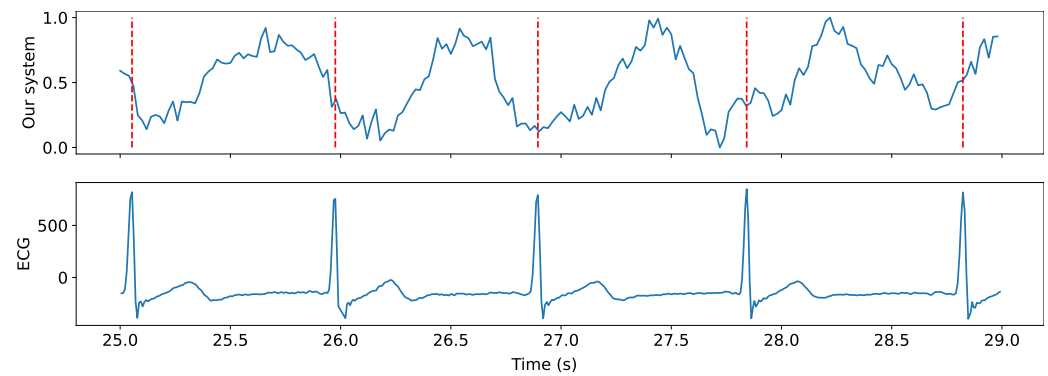


Figure 10. Extracted heartbeats of an individual located at 1 m and ground truth from ECG. The signal extracted from our system then undergoes FFT to obtain the heart rate in bpm.

5.2. Overall Performance

Initially, we assessed the system's performance in everyday scenarios involving two individuals located in front of the device at an arbitrary distance from 0.5 to 1 m. Each recording session lasted for 2 min per couple, resulting in approximately 20 min of total recording time and a total of 1124 datapoints. As depicted in Figure 11a, the observed resting heart rates ranged from 57 to 96 bpm. The grey dashed lines represent two standard deviations. Additionally, Figure 11b illustrates that the achieved median error for heart rate estimation was 0.66 bpm and 1.67 bpm at the 90th percentile.

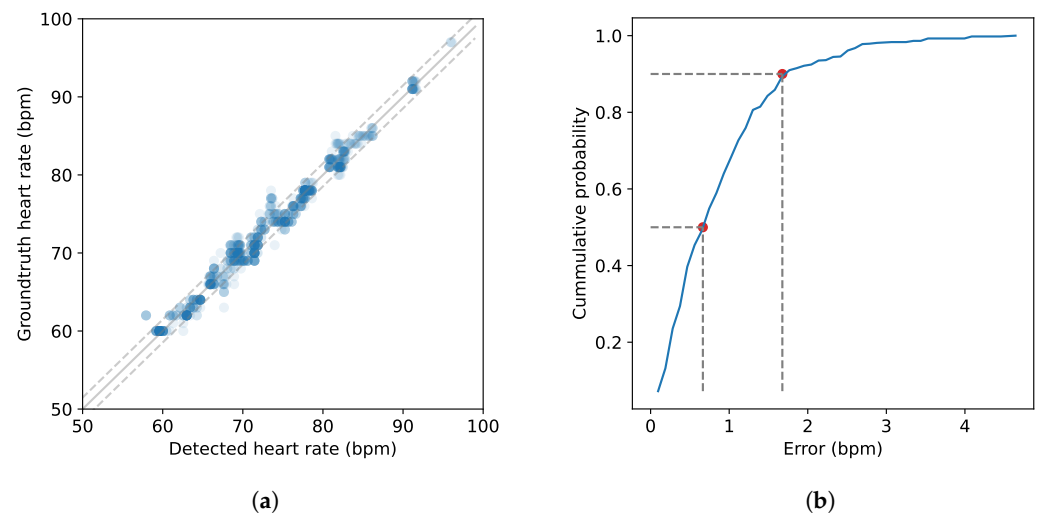


Figure 11. Overall evaluation of the system: (a) detected and ground truth heart rates in bpm and (b) cumulative distribution function of the error.

5.3. Impact of Distance

We asked ten couples to evaluate the impact of distance on the system's performance when sitting next to each other in front of the device at various distances ranging from 0.5 to 3 m, with a step size of 0.5 m. The measurements were taken at each distance for 2 min while the couples are asked to remain stationary. Figure 12 shows median errors below 1 bpm when the distance was shorter than 2 m, with errors of 0.9, 0.71, 0.79, and 0.81 bpm at 0.5, 1, 1.5 and 2 m, respectively. Due to power loss, the median error increased slightly when the participants were located further from the device, with errors of 1.2 and 0.93 bpm at 2.5 and 3 m, respectively.

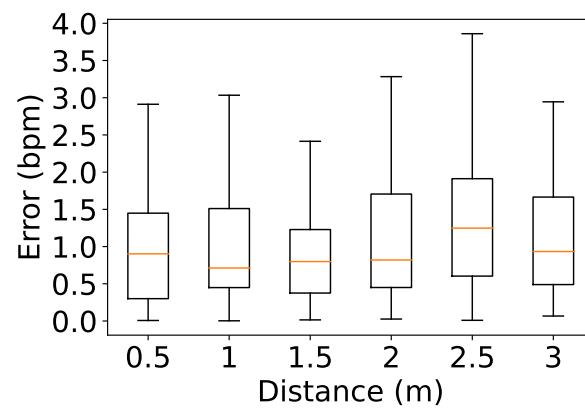


Figure 12. Impact of distance.

5.4. Impact of Angle

To assess the impact on the system of the angle between the users and the device, participants were asked to sit still at a fixed distance of 2 m with an angle φ changing from 0° to 15° and 30° , as shown in Figure 13. In Figure 14, it can be seen that the lowest median error when participants were seated at 0° was 0.65 bpm, while the highest median error was 0.93 bpm at 30° . We note that although the users were closer together when φ was lower, this did not decrease the accuracy very much. In fact, the error is mainly caused by the weak reflection when users are not directly facing the device, which leads to higher φ .

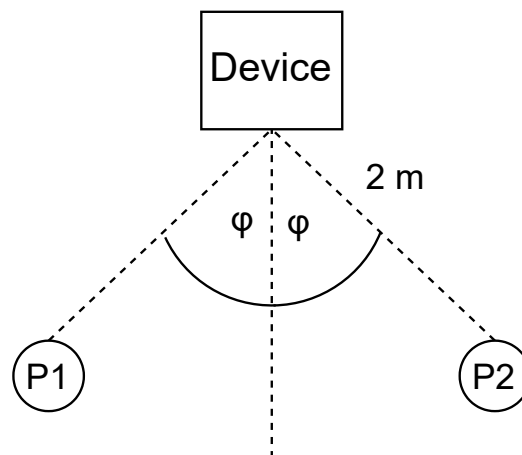


Figure 13. Results for users sitting at different angles; P1 and P2 refer to the two participants.

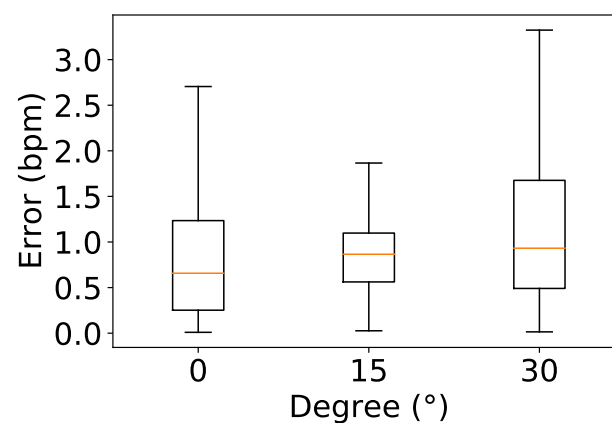


Figure 14. Impact of angle.

5.5. Impact of Ambient Noise

We investigated the impact of ambient noise by conducting the measurements in the presence of loud music. Each couple was asked to sit in front of the device at a distance of 2 m. The speaker was placed next to them while it played songs at 50 and 75 dB(A). These sound levels are comparable to normal conversation and road noise, respectively. Figure 15 shows the system performance under these two noise levels and when the average noise level when the room was quiet at 25 dB(A). There was an increase in median error with higher sound pressure, especially when the sound pressure exceeded the sound pressure of the device itself, with the median error increasing to 1.5 bpm at 75 dB(A) compared to 0.8 bpm at 25 dB(A).

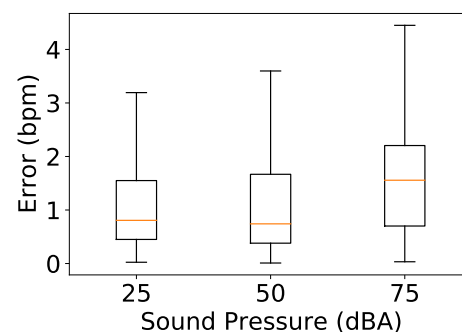


Figure 15. Impact of noise.

5.6. Lying Down with Different Postures

Different postures can lead to varying levels of accuracy. We conducted a user study involving couples lying down in the same bed in four common real-world postures: lying face-up, lying face-down, lying on the right side, and lying on the left side. To ensure that the participants could not block each others' signals when performing different postures, we positioned the device 0.5 to 1 m away from their heads, similar to Figure 1b. The results are shown in Figure 16.

In Figure 16, it can be seen that the best accuracy of 0.38 bpm was achieved with both participants lying face-up, while the lowest (1.4 bpm) was found with both lying face-down. This is because the latter posture weakens the body displacement caused by the heart. Motion signals were slightly reduced for users lying on their right and left sides as compared to facing up, with the median error increasing to 0.72 and 0.68 bpm, respectively.

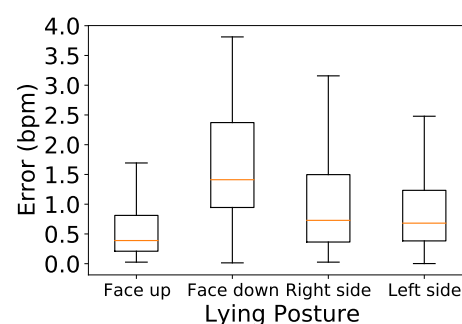


Figure 16. Impact of posture.

5.7. Lying Down with Blanket

We further evaluated our system with people lying face-up in bed covered by a blanket. As shown in Figure 17, the highest accuracy was achieved when both people were not covered by the blanket, with a median error of 0.53 bpm. Because thick cloth attenuates the signal, with the blanket there was a slight increase in the error to 0.64 bpm.

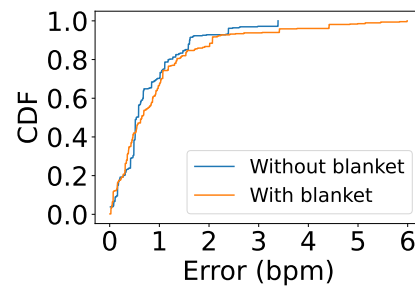


Figure 17. Impact of blanket.

5.8. Impact of Movement

Instant body movements such as posture changes, talking, or phone swiping do not lead to the same rhythm as heartbeats. Hence, such movements should have little impact on measurement. To investigate the performance of our system when users are perform such sudden movements, we asked each couple to sit next to each other while they reading a text on their phones and scrolling down or up once in a while. Participants were asked to naturally change their posture if needed; in fact, such short-time motions sometimes cause noise in the same frequency range as the heart rate, as body movements lead to much larger phase changes compared to subtle heartbeat motions. To deal with this, we applied a heuristic method that assumes the correct heart rate will change very little over time while the noise will disappear in the subsequent time intervals. Despite a slight difference in the error distribution, Figure 18 shows comparable median errors between users sitting still and users performing sudden movements, with 0.9 and 1 bpm, respectively.

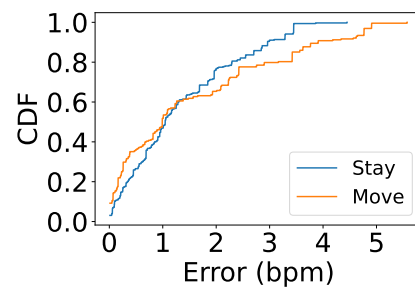


Figure 18. Impact of movement.

5.9. Impact of Number of Targets

To evaluate our solution with up to four people, we asked groups of two, three, and four people to sit side by side for a 5-min trial at a distance of 2 m. As shown in Figure 19, the errors in all four cases matched the errors in Figure 12 for two people sitting at a distance of 2 m. Thus, it can be concluded that the accuracy of the system is not impacted even when increasing the number of users to four.

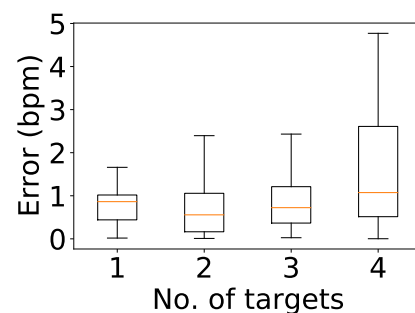


Figure 19. Impact of number of targets.

5.10. Impact of Number of Microphones

To assess the system performance under different microphone arrays, we ran the collected data with two, four, and seven microphones in the array. For the array with two microphones, we selected the ones located at 120° and 240°; for the array with four microphones, we selected the ones located at 60°, 120°, 240°, and 300° (see Figure 8); and for the array with seven microphones, all of the microphones in the array were operating. Figure 20 shows that the median heart rate error was below 1 bpm in all three cases (0.91, 0.82, and 0.85 bpm for two, four, and seven microphones, respectively). However, regarding the detectable time, which refers to the time during which all users' heart rates are visible in the heatmap, the two-microphone array had the lowest rate at 86%, compared to the seven-microphone one at 100%. This is because additional microphones improve the system's ability to capture reflected signals from more directions.

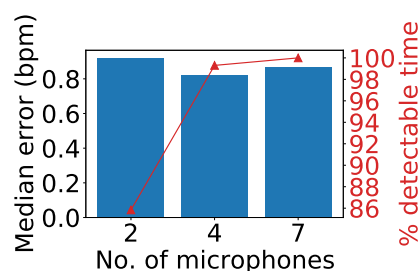


Figure 20. Impact of number of microphones.

5.11. Heart Rate Monitoring with Smartphone

To examine the approach on another platform, we implemented our system on a Samsung Galaxy S20 Plus smartphone. Due to the phone's design, it was only possible to use the single speaker–microphone pair at the bottom of the phone. Because only one microphone could be utilized, the angle information of the targets is not available. However, it remains possible to track the users by distance. We asked two volunteers to sit at distances of 2.7 m and 2.9 m in front of the smartphone. Figure 21 demonstrates that our system can achieve sensing of multiple heart rates at distances up to 3 m when deployed on a commercial smartphone with only one speaker–microphone pair; on the other hand, the sole existing smartphone-based approach [15] reports a maximum monitoring range of only 0.3 m.

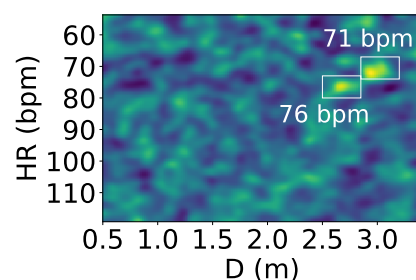


Figure 21. Heart rate detection by smartphone.

6. Discussion

The following limitations may apply to our proposed method:

- Prone to Rhythmic Movement:** Our approach can be susceptible to the impact of body movement, which is a known challenge for handling motion noise in acoustic-based methods. Because we assume that the user position falls within a frequency of 0.8 to 2.5 Hz (i.e., the normal heartbeat frequency range), any other modulation within this frequency range that does not originate from the human heart, although very unlikely, will confuse the system. As a result, although our system works well

when the motion frequency is beyond the usual range of the heart rate (e.g., if the user shakes their head during measurement), it is suggested that users remain stationary during measurement to minimize possible noise that could fall within the heart rate frequency. In addition, we assumed that the device did not vibrate and that the background within the device's working range contained no motion within the heartbeat frequency range. It is understood that voluntary or involuntary movement during signal acquisition typically reduces the fidelity of heart rate tracking [13,14].

- **Lack of Evaluation of Standing Postures:** While we conducted an extensive evaluation of our proposed system in various real-life scenarios, settings involving standing users were not included in our evaluation, as standing postures were not evaluated in any prior works in the literature. Therefore, we focused our evaluation on settings that were comparable with existing research. In fact, the standing setting is a challenge due to the difficulty of users maintaining stationary positions while in natural situations. For example, it is uncommon for an individual to remain completely still while standing for extended periods; a natural standing posture often involves walking or jogging. On the contrary, sitting or lying down naturally allows for more stationary positions. As a result, natural standing postures pose a great challenge in extracting subtle heartbeat signals, as we expect there to be significant motion noise. Consequently, we chose to leave the evaluation of different standing postures to future work.
- **Performance of Heart Rate Detection:** One assumption in our approach is that the normal heart rate falls within the range of 0.8 and 2.5 Hz, which is known as the normal heart rate range [37]. As such, any heart rate below 0.8 Hz may not be correctly detected, as the second harmonic of the respiration signal falls into this range and has significantly higher amplitude than the heartbeat.

7. Conclusions

In this paper, we present a remote approach to monitor the heart rates of multiple individuals using a commercial smart speaker with no separation requirement. Our proposed method removes interference and amplifies heart rates using a seven-microphone array on a smart speaker. This approach is able to separate heartbeat signals even when multiple users are sitting next to one another or lying down. Through our user study in various practical sitting and lying scenarios, the proposed approach is demonstrated to be highly accurate in these situations, with a median error of only 0.66 bpm. We believe that this approach can provide insightful inputs to other works, such as sleep stage classification, stress detection, and emotion classification.

Author Contributions: Conceptualization, T.T., D.M. and R.B.; methodology, T.T.; software, T.T.; validation, T.T.; formal analysis, T.T.; writing—original draft, T.T.; writing—review and editing, D.M. and R.B.; visualization, T.T.; funding acquisition, R.B.; data curation, T.T.; resources, supervision, and project administration, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant (Project ID: 22-SIS-SMU-051). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

Institutional Review Board Statement: IRB Approval has been approved under Category 2B: Expedited Review. The IRB approval number is IRB-21-168-A123(1021). Approval period from 12 October 2021 to 11 October 2022.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to privacy issues.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Acharya, U.R.; Joseph, K.P.; Kannathal, N.; Lim, C.M.; Suri, J.S. Heart rate variability: A review. *Med Biol. Eng. Comput.* **2006**, *44*, 1031–1051. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Taelman, J.; Vandeput, S.; Spaepen, A.; Van Huffel, S. Influence of mental stress on heart rate and heart rate variability. In Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering, Antwerp, Belgium, 23–27 November 2008; pp. 1366–1369.
3. Greenland, P.; Daviglus, M.L.; Dyer, A.R.; Liu, K.; Huang, C.F.; Goldberger, J.J.; Stamler, J. Resting heart rate is a risk factor for cardiovascular and noncardiovascular mortality: The Chicago Heart Association Detection Project in Industry. *Am. J. Epidemiol.* **1999**, *149*, 853–862. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Snyder, F.; Hobson, J.A.; Morrison, D.F.; Goldfrank, F. Changes in respiration, heart rate, and systolic blood pressure in human sleep. *J. Appl. Physiol.* **1964**, *19*, 417–422. [\[CrossRef\]](#)
5. Yang, Y.; Yuan, Y.; Zhang, G.; Wang, H.; Chen, Y.C.; Liu, Y.; Tarolli, C.G.; Crepeau, D.; Bukartyk, J.; Junna, M.R.; et al. Artificial intelligence-enabled detection and assessment of Parkinson’s disease using nocturnal breathing signals. *Nat. Med.* **2022**, *28*, 2207–2215. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Adib, F.; Mao, H.; Kabelac, Z.; Katabi, D.; Miller, R.C. Smart homes that monitor breathing and heart rate. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 18–23 April 2015; pp. 837–846.
7. Yue, S.; He, H.; Wang, H.; Rahul, H.; Katabi, D. Extracting multi-person respiration from entangled rf signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 86. [\[CrossRef\]](#)
8. Liu, J.; Wang, Y.; Chen, Y.; Yang, J.; Chen, X.; Cheng, J. Tracking vital signs during sleep leveraging off-the-shelf wifi. In Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Hangzhou, China, 22–25 June 2015; pp. 267–276.
9. Wang, H.; Zhang, D.; Ma, J.; Wang, Y.; Wang, Y.; Wu, D.; Gu, T.; Xie, B. Human respiration detection with commodity wifi devices: Do user location and body orientation matter? In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, NY, USA, 12–16 September 2016; pp. 25–36.
10. Meng, Z.; Fu, S.; Yan, J.; Liang, H.; Zhou, A.; Zhu, S.; Ma, H.; Liu, J.; Yang, N. Gait recognition for co-existing multiple people using millimeter wave sensing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 849–856.
11. Yang, Z.; Pathak, P.H.; Zeng, Y.; Liran, X.; Mohapatra, P. Monitoring vital signs using millimeter wave. In Proceedings of the 17th ACM International Symposium on Mobile ad hoc Networking and Computing, Paderborn, Germany, 5–8 July 2016; pp. 211–220.
12. Wang, L.; Gu, T.; Li, W.; Dai, H.; Zhang, Y.; Yu, D.; Xu, C.; Zhang, D. DF-Sense: Multi-user Acoustic Sensing for Heartbeat Monitoring with Dualforming. In Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services, Helsinki, Finland, 18–22 June 2023; pp. 1–13.
13. Wang, A.; Nguyen, D.; Sridhar, A.R.; Gollakota, S. Using smart speakers to contactlessly monitor heart rhythms. *Commun. Biol.* **2021**, *4*, 319. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Zhang, F.; Wang, Z.; Jin, B.; Xiong, J.; Zhang, D. Your Smart Speaker Can “Hear” Your Heartbeat! *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 161. [\[CrossRef\]](#)
15. Qian, K.; Wu, C.; Xiao, F.; Zheng, Y.; Zhang, Y.; Yang, Z.; Liu, Y. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In Proceedings of the IEEE INFOCOM 2018—IEEE Conference on Computer Communications, Honolulu, HI, USA, 15–19 April 2018; pp. 1574–1582.
16. Wang, L.; Li, W.; Sun, K.; Zhang, F.; Gu, T.; Xu, C.; Zhang, D. LoEar: Push the Range Limit of Acoustic Sensing for Vital Sign Monitoring. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 145. [\[CrossRef\]](#)
17. UMA-8-SP USB mic Array. Available online: <https://www.minidsp.com/products/usb-audio-interface/uma-8-sp-detail> (accessed on 4 January 2024).
18. Amazon Echo Dot Smartspeaker. Available online: <https://www.amazon.com/All-New-Amazon-Echo-Dot-Add-Alexa-To-Any-Room/dp/B01DFKC2SO> (accessed on 4 January 2024).
19. Wang, Z.; Zhang, F.; Li, S.; Jin, B. Exploiting Passive Beamforming of Smart Speakers to Monitor Human Heartbeat in Real Time. In Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021; pp. 1–6.
20. Israel, S.A.; Irvine, J.M.; Cheng, A.; Wiederhold, M.D.; Wiederhold, B.K. ECG to identify individuals. *Pattern Recognit.* **2005**, *38*, 133–142. [\[CrossRef\]](#)
21. Li, C.; Zheng, C.; Tai, C. Detection of ECG characteristic points using wavelet transforms. *IEEE Trans. Biomed. Eng.* **1995**, *42*, 21–28. [\[PubMed\]](#)
22. Nardelli, M.; Vanello, N.; Galperti, G.; Greco, A.; Scilingo, E.P. Assessing the quality of heart rate variability estimated from wrist and finger ppg: A novel approach based on cross-mapping method. *Sensors* **2020**, *20*, 3156. [\[CrossRef\]](#)
23. Temko, A. Accurate heart rate monitoring during physical exercises using PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2016–2024. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Chen, L.; Xiong, J.; Chen, X.; Lee, S.I.; Zhang, D.; Yan, T.; Fang, D. LungTrack: Towards contactless and zero dead-zone respiration monitoring with commodity RFIDs. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 79. [\[CrossRef\]](#)

25. Bakhtiari, S.; Elmer, T.W.; Cox, N.M.; Gopalsami, N.; Raptis, A.C.; Liao, S.; Mikhelson, I.; Sahakian, A.V. Compact millimeter-wave sensor for remote monitoring of vital signs. *IEEE Trans. Instrum. Meas.* **2011**, *61*, 830–841. [[CrossRef](#)]
26. Chuang, H.R.; Kuo, H.C.; Lin, F.L.; Huang, T.H.; Kuo, C.S.; Ou, Y.W. 60-GHz millimeter-wave life detection system (MLDS) for noncontact human vital-signal monitoring. *IEEE Sens. J.* **2011**, *12*, 602–609. [[CrossRef](#)]
27. Kao, T.Y.J.; Lin, J. Vital sign detection using 60-GHz Doppler radar system. In Proceedings of the 2013 IEEE International Wireless Symposium (IWS), Beijing, China, 14–18 April 2013; pp. 1–4.
28. Wang, C.; Xie, L.; Wang, W.; Chen, Y.; Bu, Y.; Lu, S. Rf-ecg: Heart rate variability assessment based on cots rfid tag array. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 85. [[CrossRef](#)]
29. Nandakumar, R.; Gollakota, S.; Watson, N. Contactless sleep apnea detection on smartphones. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, Florence, Italy, 18–22 May 2015; pp. 45–57.
30. Wang, T.; Zhang, D.; Zheng, Y.; Gu, T.; Zhou, X.; Dorizzi, B. C-FMCW based contactless respiration detection using acoustic signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *1*, 170. [[CrossRef](#)]
31. Peng, C.; Shen, G.; Zhang, Y.; Li, Y.; Tan, K. Beepbeep: A high accuracy acoustic ranging system using cots mobile devices. In Proceedings of the 5th International Conference on Embedded Networked Sensor Systems, Sydney, Australia, 6–9 November 2007; pp. 1–14.
32. Wang, A.; Gollakota, S. Millisonic: Pushing the limits of acoustic motion tracking. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–11.
33. Bachorowski, J.A.; Smoski, M.J.; Owren, M.J. The acoustic features of human laughter. *J. Acoust. Soc. Am.* **2001**, *110*, 1581–1597. [[CrossRef](#)] [[PubMed](#)]
34. Chen, J.S.; Huertas, A.; Medioni, G. Fast convolution with Laplacian-of-Gaussian masks. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *PAMI-9*, 584–590. [[CrossRef](#)] [[PubMed](#)]
35. Chan, A.; Litva, J. MUSIC and maximum likelihood techniques on two-dimensional DOA estimation with uniform circular array. *IEEE Proc. Radar Sonar Navig.* **1995**, *142*, 105–114. [[CrossRef](#)]
36. Polar H10 Heart Rate Sensor. Available online: <https://www.polar.com/sg-en/sensors/h10-heart-rate-sensor/> (accessed on 4 January 2024).
37. Pulse Rate. Available online: <https://www.bhf.org.uk/information-support/heart-matters-magazine/medical/ask-the-experts/pulse-rate> (accessed on 4 January 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.