Exploiting Passive Beamforming of Smart Speakers to Monitor Human Heartbeat in Real Time

Zhi Wang[1,2], Fusang Zhang[1,2], Siheng Li[1,2], Beihong Jin[1,2,3*]
[1]*State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences*
[2]*University of Chinese Academy of Sciences*
[3]*Beijing Key Laboratory on Integration and Analysis of Large-scale Stream Data*
Beijing, China
{wangzhi20, zhangfusang10, lisiheng19, jbh}@otcaix.iscas.ac.cn

***Abstract*—Currently, cardiac diseases have become one of the biggest health concerns. Existing heartbeat monitoring methods either require dedicated intrusive devices (e.g., ECG devices) that suffer high costs or leverage video camera analyses that are light-sensitive. In this paper, leveraging the acoustic signals sent by a speaker and received by a microphone array, we develop a prototype system to achieve the contactless and low-cost heartbeat monitoring. In particular, while we exploit the passive beamforming to enhance the user's heartbeat signal, we design a filtering method in frequency domain to remove the line-of-sight (LoS) impact and retain the target-reflected signals, and propose a wideband time-delay method to estimate the direction of arrival of target-reflected signal. Thus, our prototype is able to robustly estimate the human heartbeat and push the limit of acoustic sensing range. The experimental results show that our prototype achieves a heart rate monitoring at 1.7 m with the estimation error of 0.5 bpm, which is comparable to ECG or other contact-based solutions.**

***Index Terms*—Acoustic sensing, Vital sign monitoring, Contactless sensing, Health monitoring**

I. Introduction

Daily cardiac monitoring can help diagnose abnormal increase and decrease of heart rate, cardiac arrest and etc., thereby preventing many heart related diseases from happening. Traditional cardiac monitoring is usually carried out by medical professionals and employs special devices such as electrocardiographs [12]. Although these devices can achieve high-precision monitoring, they are usually expensive and need to be operated by professionals. For long-term monitoring of our health status, the convenient and low-cost solution is particularly desirable.

Existing heartbeat monitoring solutions that are relatively convenient and suitable for home use can be roughly divided into two categories: contact-based solutions and contactless solutions. Contact-based solutions employ wearable devices or sensors (such as finger clip pulse meters, smart watches, and wristbands) [10] [5] [6] to record the electrical signals of the heartbeat and generate the well-known ECGs. However, these devices need to directly contact with the user and let user always wear, which is not suitable for long-term continuous monitoring. Existing contactless solutions mainly depend on radio frequency (RF) signals (e.g., by using mmWave radars) [15] [9] [20] or cameras [11] [17] [18]. However, mmWave radars are relatively expensive and cameras have problems of occlusion and privacy leakage. Therefore, recent work explores acoustic signals to provide non-invasive and low-cost health monitoring solutions. In principle, the frequency modulated continuous wave (FMCW) signal sent by a speaker can be reflected by the human chest and the reflected signal can be received by the microphone(s). By analyzing the reflected signal, vital signs such as respiration and heartbeat can be obtained. For example, ApneaApp [13] utilizes a smart phone which contains a built-in speaker and microphone to send and receive FMCW signals and achieves human respiration monitoring and sleep apnea detection. Acousticcardiogram [14] utilizes the dual-microphone design of the smartphone to eliminate direct power leakage, and obtains the heart rate in the frequency domain. However, these smart phone based solutions are unsatisfactory because of their unstable performance and limited monitoring range, where the maximum monitoring distance is only 30 cm.

Recently, smart speakers which contains one or multiple speakers and microphones (such as Amazon Echo and Google Home) become popular in homes [8]. Considering that smart speakers have more powerful acoustic signal processing capabilities than mobile phones [3] [4] [2], we believe smart speaker based solutions can provide upgraded user experience in terms of sensing accuracy, distance and orientation.

Technically, smart speaker based heatbeat monitoring face three challenges. Firstly, in the smart speaker scenario where the speaker is close to the microphone array, traditional beamforming will be affected by the direct path signal from the speaker and cannot effectively enhance the signal reflected by target. However, eliminating the strong direct path signal from the speaker to the microphone from the design of hardware will destroy the normal playback of the speaker and disable the concurrent sensing ability. To deal with this challenge, we design a filtering method in frequency domain, which successfully eliminates the impact of the direct path signal in the frequency domain. It thus effectively combines multiple microphones to jointly improve the performance.

Secondly, for the FMCW acoustic signal, its bandwidth occupies more than 20% of the center frequency of the transmitted signal, that is, it belongs to the wideband signal. Thus, the phase delay at each microphone is no longer linearly correlated, it is impossible to estimate the direction of the

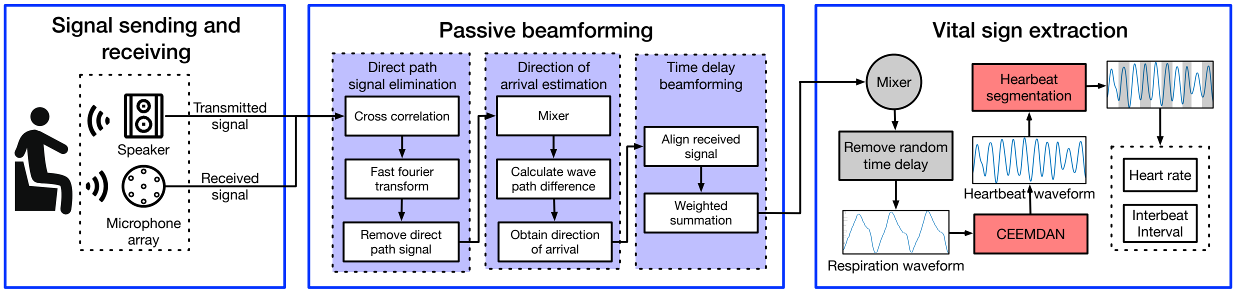

Fig. 1. Overview of our heartbeat monitoring system.

signal by taking a snapshot of the signal received by the microphone array at the same time. To deal with this challenge, we design a time-delay direction of arrival (DOA) estimation method for wideband acoustic signals. By mixing the signal received by the reference microphone with the signals received by other microphones, the frequency difference between the microphones can be obtained to infer the time delay of different microphones. We thus are able to obtain the DOA of reflected signals, align the signals of each microphone in the time domain and conduct weighted summation.

Lastly, the motions of human chest are mainly caused by respiration, heartbeat, and involuntary small body movements. It is a challenge to extract the weak heartbeat information that is submerged in mixed signals. In order to separate the superimposed heartbeat, respiration and other motions, we adopt a complete ensemble empirical mode decomposition method, which decomposes mixed signals into multiple independent intrinsic modes, from which the heartbeat waveform of the target is extracted.

In short, in the paper, we utilize a commodity speaker and a microphone array to develop a heartbeat monitoring prototype, obtaining the high-precision and real-time heartbeat in a low-cost and non-invasive way. The main contributions of this paper are summarized as follows.

- We propose a geometric model of a uniform circular microphone array for contactless sensing, which sets up the connection between DOA and wave path difference for each microphone and facilitates subsequent DOA estimation.
- We design the passive beamforming for the smart speaker scenario, presenting a filtering method in frequency domain to eliminate the direct path signal from the received signal, designing a wideband time-delay method to estimate the DOA of reflected signal and align received signals to enhance the sensing performance.
- We evaluate the effectiveness of the prototype under different experimental settings. Experimental results show that our prototype can achieve a heart rate monitoring error of 0.5 bpm at a maximum distance of 1.7 m with complete coverage of 360 degrees, which is comparable to contact-based solutions.

II. Our System

Our system is to monitor heartbeat by acoustic signals. As shown in Fig. 1, the system mainly consists of three modules. **Signal sending and receiving:** A speaker is employed to send the FMCW signal and a microphone array to receive signals. **Passive beamforming:** The received signals of multiple microphones are fused through passive beamforming to enhance the signal-to-noise ratio and finally improve the sensing distance.
Vital sign extraction: Since the heartbeat is affected by respiration and the involuntary motions of the body, we have to separate different signals to extract the heartbeat waveform. Further, based on the heartbeat form, we segment the heartbeat signal and estimate interbeat intervals (IBIs).

A. *Geometric Modeling*

For the uniform circular microphone array which is most commonly used in smart speakers, we give a geometric model, as shown in Fig. 2. Let the center of the microphone array be the origin of the coordinates. By numbering counterclockwise from the positive half of the x-axis, microphones are identified as $(x_i, y_i), i = 1, 2, ..., M$, where M is the number of microphones. We denote the angle between the 1st microphone and the x-axis as α, the incident angle of the target-reflected signal as (θ, ϕ) where θ and ϕ represent the azimuth angle and the elevation angle, respectively. Here, the azimuth angle is the included angle between the projection of target-reflected signal on the xOy plane and the x-axis, and the elevation angle is the included angle between the target-reflected signal and the xOy plane. Then the unit direction vector of the target-reflected signal can be expressed as:

$$\boldsymbol{I} = \begin{bmatrix} cos\theta cos\phi & sin\theta cos\phi & sin\phi \end{bmatrix} \tag{1}$$

Let the radius of microphone array be R and the 1st microphone be the reference point, its

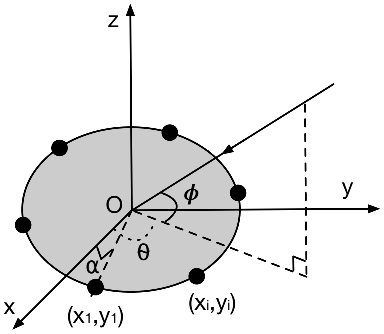

Fig. 2. Geometric model of the uniform circular microphone array.

position vector $\boldsymbol{m}_1$ is $\begin{bmatrix} Rcos\alpha & Rsin\alpha & 0 \end{bmatrix}$ and the position vector of the i-th microphone $\boldsymbol{m}_i$ is $\begin{bmatrix} Rcos(\alpha + \frac{2\pi}{M}(i-1)) & Rsin(\alpha + \frac{2\pi}{M}(i-1)) & 0 \end{bmatrix}$. Let $\boldsymbol{l}_i$ be the vector connecting the i-th microphone and the reference point, it can be expressed as $\boldsymbol{l}_i = \boldsymbol{m}_i - \boldsymbol{m}_1$. Then the wave path difference of the target-reflected signals received by the i-th microphone and the reference point can be expressed as:

$$\begin{aligned} \Delta_{w_i} &= < \boldsymbol{l}_i, \boldsymbol{I} > = < \boldsymbol{m}_i - \boldsymbol{m}_1, \boldsymbol{I} > \\ &= Rcos\phi(cos(\alpha + \frac{2\pi}{M}(i-1) - \theta) - cos(\alpha - \theta)) \end{aligned} \tag{2}$$

B. Signal Sending and Receiving

In our system, the speaker and the microphone array are required to be placed in the same location, hereafter they are collectively called ”devices”. The transmitted signal $x_{tx}(t)$ is an FMCW signal (chirp signal) whose frequency linearly increases over time, it can be expressed as:

$$x_{tx}(t) = Acos(2\pi(f_0 t + \frac{kt^2}{2})) \tag{3}$$

where $k = B/T$ is the slope of frequency change, B is the bandwidth, T is the sweep time, f_0 is the starting frequency, A is the amplitude of transmitted signal.

Let S be the distance between the target and the device, and c be the signal propagation speed in the air, the signal arrives at the target and is reflected to the i-th microphone after time $\tau_i = 2S/c$. Therefore, the received signal of i-th microphone $x_{ri}(t)$ can be treated as the transmitted signal with time delay τ_i, which is represented as:

$$x_{ri}(t) = D_i cos(2\pi(f_0(t - \tau_i) + \frac{k(t-\tau_i)^2}{2})) \tag{4}$$

where D_i is the amplitude of reflected signal received by i-th microphone.

C. Passive Beamforming

Beamforming is a technique that improves the signal-to-noise ratio of the received signals, eliminates interference sources and focuses the received signals from a specific direction through a sensing array. For example, the active beamforming is for strengthening the original acoustic signals. For vital sign monitoring via acoustic signals, since the signal change caused by heartbeat is very weak and the received signal of a single microphone might be interfered, the sensing distance is short and the performance is unstable. Thus, we design a passive beamforming method to combine received signals of multiple microphones, enhancing the received signals and improving the sensing performance.

a) Direct path signal elimination: In our system, since the speaker and the microphone array are placed in the same location, the microphone array will simultaneously receive the intense direct path signal from the speaker and the signal reflected by the human body. As shown in Figure 3, in the spectrogram of the received signals, the first chirp is the direct path signal and the second relatively strong chirp is the target-reflected signal. If directly using the spatial spectrum to estimate the DOA of the target-reflected signal, then the estimated direction will be the DOA of the direct path signal because the direct path signal plays the dominant role.

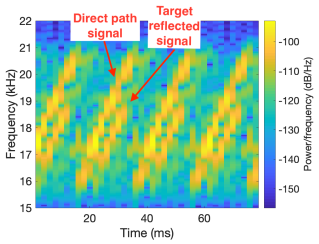

Fig. 3. The spectrogram of the received signals.

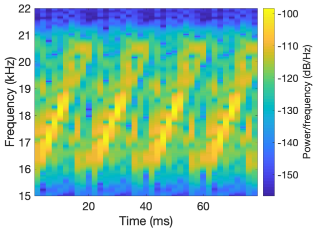

Fig. 4. The spectrogram after eliminating the direct path signal.

In order to accurately estimate the DOA of the target-reflected signal, we propose a method to eliminate the direct path signal, as shown in Algorithm 1. First, we calculate the cross-correlation coefficient between the transmitted signal x_{tx} and the received signal of the i-th microphone x_{ri} in the time domain, thereby obtaining the time delay τ_{di} of the direct path signal relative to the transmitted signal. Then we perform fast Fourier transform (FFT) on the received signal x_{ri} in a time window of length T_w, thus the frequency range of the direct path signal at time t will be $\begin{bmatrix} f_0 + (t + \tau_{di})k & f_0 + (t + \tau_{di} + T_w)k \end{bmatrix}$. Finally, we set the frequency of the direct path signal in each window to 0, and then perform inverse fast Fourier transform (IFFT) to obtain the received signal without the direct path signal. Fig. 4 shows the spectrogram after eliminating the direct path signal.

b) Direction of arrival estimation: After removing the direct path signal, the DOA of the target-reflected signal can be estimated. However, the bandwidth of the FMCW signal is relatively large in our system, and the frequency of the received signal is not a constant. Therefore, the phase delay at each microphone is no longer linearly correlated, and it is unable to estimate DOA of the target-reflected signal by taking a snapshot of the signals received by the microphone array at the same time.

Algorithm 1: Direct path signal elimination method.

```
Input: The transmitted signal $x_{tx}$, the received signal
       $x_{ri}$, $i = 1, ..., M$, sampling rate $f_s$.
for $i = 1; i \leq M; i = i + 1$ do
    calculate cross-correlation between $x_{tx}$ and $x_{ri}$ by
      $| < IFFT(FFT(x_{tx}), FFT(x_{ri}) > |$, obtain the
      sample delay $s_{di}$ by selecting peak;
    obtain time delay $\tau_{di} = s_{di}/f_s$;
    perform FFT on $x_{ri}$ in a time window of length
      $T_w$;
    set the result of FFT in frequency range
      $[f_0 + (t + \tau_{di})k \quad f_0 + (t + \tau_{di} + T_w)k]$ to 0;
    perform IFFT to obtain the received signal without
      the direct path signal.
end
```

Algorithm 2: Wideband DOA estimation method.

```
Input: The received signal $x_{ri}$, $i = 1, ..., M$.
Output: The incident angle of the target-reflected
        signal $\theta, \phi$.
for $i = 2; i \leq M; i = i + 1$ do
    multiply $x_{ri}$ and $x_{r1}$ and perform FFT to obtain $f_i$;
    $\Delta_{w_i} = f_i c/k$;
    put $\Delta_{w_i}$ into (2);
end
solve the equation set by the trust-region dogleg
  algorithm, obtain $\theta$ and $\phi$;
return $\theta, \phi$
```

We design a passive wideband DOA estimation method for the FMCW signal. As shown in Algorithm 2, we first calculate the time delay for each received signal, let the 1st microphone be the reference point, we perform the signal mixing operation by multiplying x_{ri} and x_{r1}, and then perform the FFT operation to obtain the frequency difference f_i between x_{ri} and x_{r1}. Through the relationship between frequency difference and wave path difference, we obtain the wave path difference Δ_{w_i} between the i-th microphone and the reference point. Finally, we put Δ_{w_i} into (2). By solving the equation set using the trust-region dogleg algorithm, we can obtain the DOA of the target-reflected signal.

c) Time delay beamforming: After obtaining the DOA of the target-reflected signal, we can obtain the time delay at each microphone which can be expressed as:

$$\Delta_{t_i} = \frac{Rcos\phi(cos(\alpha + \frac{2\pi}{M}(i-1) - \theta) - cos(\alpha - \theta))}{c} \tag{5}$$

Then we align the received signal of each microphone and compute the weighted summation of the received signals at all M microphones, obtaining the beamformed signal received at direction (θ, ϕ) as follows.

$$x_{rx}(t) = \sum_{i=1}^{M} x_{ri}(t + \Delta_{t_i}) \tag{6}$$

D. Vital Sign Extraction

After the beamforming process, we perform a mixing operation on the transmitted signal and the enhanced received signal to obtain the mixed signal. The mixed signal is composed of a high-frequency component and a low-frequency component, where the low-frequency component indicates the frequency difference between $x_{tx}(t)$ and $x_{rx}(t)$. After the mixed signal is passed through a low-pass filter to remove the high-frequency component, the left only contains the low-frequency component, which can be represented as:

$$x_m(t) = \frac{AD}{2} cos\Big(2\pi(f_0\tau - \frac{k(\tau^2 - 2t\tau)}{2})\Big) \tag{7}$$

where D is the amplitude of the enhanced received signal and τ is the delay between the transmitted signal and the enhanced received signal. In general, we can obtain τ by performing the FFT on the mixed signal. However, in the real system, the acoustic signal is firstly put into a buffer before it is sent out, so the system have a random delay and the frequency of mixed signal is determined by the signal propagation time in the air and the system random delay. We calculate the system random delay through analyzing few received signals, and eliminate its affect by a corresponding delay compensation on the transmitted signal.

After eliminating the system random delay, we get the frequency of mixed signal $f_b = k\tau$, so the distance S between the target and the device can be calculated as $S = \frac{c \cdot f_b}{2k}$. Then we separate target's vital sign information (amplitude and phase) at a distance of S from the mixed signal. However, due to limitation of signal bandwidth, the amplitudes of respiration and heartbeat waveforms are not great enough to change the frequency of the mixed signal. Therefore, we have to obtain the respiration waveform by the phase change of mixed signal.

Let Δs be the subtle movement distance of human chest, the mixed signal can also be represented as:

$$x_m(t) = \frac{AD}{2} cos(2\pi f_b t + \frac{4\pi f_0 \Delta s}{c}) \tag{8}$$

where the term $\frac{4\pi f_0 \Delta s}{c}$ is the phase change of mixed signal. Suppose that $\Delta s = 5mm$, $f_0 = 16kHz$, $c = 343m/s$, the corresponding phase change is 167.4°. That is, chest displacement of 1-5 mm will cause the signal phase change of 33.5° to 167.4°. Thus, by extracting phase change of mixed signal, we can obtain respiration waveform.

Further, we detect the heartbeat. Since the chest displacement caused by heartbeat is very small, it is submerged in the respiration waveform, what is worse, the respiration waveform also contains other tiny movements of the body. Therefore, we apply complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [16] to decompose the superimposed signal into the superposition of multiple intrinsic mode functions (IMFs) including respiration, heartbeat, tiny body movements, and other noise components.

Let $y(t)$ be the phase change we get from the mixed signal, $n_j(t)(j = 1, ..., N)$ be the white noise with zero mean and unit variance, $\widetilde{d_h}$ be h-th mode separated by CEEMDAN, $E_h(\cdot)$ be the operator which produces the h-th mode obtained by EMD, β_h be signal-to-noise ratio coefficient. Thus, in the light of CEEMDAN, the phase change of heartbeat signal can be obtained as follows.

Step 1: For every $j = 1, ..., N$, decompose each $y_j(t) = y(t) + \beta_0 n_j(t)$ by EMD until obtaining its first mode $E_1(y_j(t))$. The first mode of CEEMDAN can be represented as $\widetilde{d_1} = \frac{1}{N}\sum\limits_{j=1}^{N} E_1(y_j(t))$.

Step 2: At the first stage ($h = 1$), calculate the first residue: $r_1 = y(t) - \widetilde{d_1}$.

Step 3: Get $(h+1)$-th mode by recurrence formula $\widetilde{d_{h+1}} = \frac{1}{N}\sum\limits_{j=1}^{N} E_1(r_h + \beta_h E_h(n_j(t)))$ and $r_{h+1} = r_h - \widetilde{d_{h+1}}$, until the residue cannot be further decomposed by EMD or the max number of IMF is reached.

Step 4: Considering that the frequency of respiration waveform is [0.1-0.5Hz] and the frequency of heartbeat waveform is [0.8- 2Hz], we can obtain heartbeat waveform by differentiating frequency ranges.

For the resultant heartbeat waveform, we adopt EM algorithm to segment it. We deem that each heartbeat is continuous in time domain and has the same morphology i.e., template. Thereby, the segmentation of the heartbeat waveform can be converted into an optimization problem for each heartbeat. Through iterative optimization of the heartbeat template and the heartbeat segmentations, we can obtain the optimal heartbeat segmentations. On the basis of heartbeat segmentations, we can easily obtain IBIs of heartbeat.

III. Evaluation

A. Experimental Setup

We have implemented our prototype using an off-the-shelf speaker (JBL Jembe, 6 Watt, 80 dB) and a microphone array (UMA-8 USB MIC ARRAY-V2.0) [7]. The speaker and microphone array are connected to a laptop (MacBook Pro 2GHz with an Intel Core i5, 16 GB RAM) via the 3.5mm audio interface (AUX) and the USB interface respectively, as shown in Fig. 5. We make the speaker to transmit FMCW acoustic signal with $f_0 = 16kHz, B = 5kHz, T = 0.02s$, let the microphone array receive the signals using a $48kHz$ sampling rate, and employ the laptop to process the signals in real time. The demo can be viewed online at https://youtu.be/ButGm9Wf-CQ. Meanwhile, we employ a 3-lead ECG monitor i.e., Heal Force PC-80B as shown in Fig. 5 to record the ground-truth data, where we calculate heartbeat rate and IBIs from the ECGs. During the experiments, participants sit in front of the devices and breathe normally. We adopt the heartbeat rate error and the IBI error to evaluate the sensing performance. Here, the heartbeat rate error is the absolute value of the difference between the estimated heartbeat rate and the ground-truth rate. The unit is beats per minute (bpm). The IBI error is the absolute time difference between the estimated heartbeat interval and the ground truth. The unit is millisecond.

B. Sensing Performance Comparison

We first evaluate the sensing performance of our system in terms of heartbeat rate error and IBI error. In order to compare with our system, we build another system [19] which adopts same devices as our system except for using a single microphone. We also choose a commercial contact-based solution, i.e., Kiwi App [1] as a competitor. Kiwi app requires users to press their fingers on the phone camera, uses a flashlight to light up the fingers and measures the light absorbed by oxidized hemoglobin. In this experiment, we place the devices at a distance of 0.6 m from the participant and monitor each participant for a period of 5 mins.

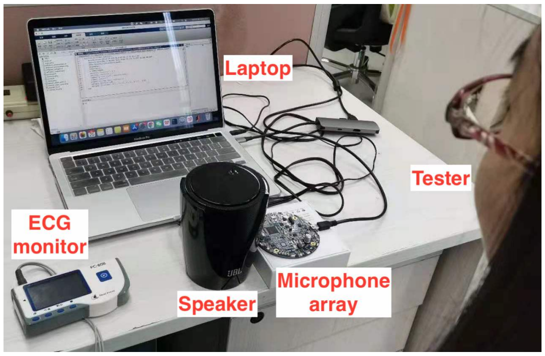

Fig. 5. Experimental setup.

Fig. 6 plots the cumulative distribution function (CDF) of the heartbeat rate errors of the three approaches. The achieved median errors of our system, single microphone system and Kiwi are 0.5 bpm, 0.75 bpm, and 3.07 bpm, respectively. Our system has the smallest error, and in the more than 70% cases the heartbeat rate error of our system is less than 1bpm. Compared with the single microphone system, the accuracy of our system is higher. This is because the microphone array can receive signals from multiple channels, so we can obtain more stable and accurate heartbeat information by beamforming. Compared with the Kiwi App, the accuracy of our system is much higher. The reason might be that Kiwi requires users to press their fingers on the camera. The stability of the contact between the fingers and the camera might affect the accuracy of heartbeat monitoring.

Fig. 7 plots the CDF of IBI errors. The achieved median IBI errors of our system, single microphone system, and Kiwi are 5.67ms, 11.46ms, and 17.53ms, respectively. The IBI error of our system is the lowest. IBI plays an important role in detecting most heart diseases and can be used for heart rate variability analysis.

C. Impact of User-device Distance

We evaluate the performance of our system while users sit at various distances away from the devices. Fig. 8 shows the impact of the distance, where the distance ranges from 0.5 m to 1.7 m with a step length of 0.2 m. The achieved median error of our system is still less than 2 bpm even at 1.5 m. And we can observe that as the user-device distance increases, the heart rate monitoring error slowly increases. This is because the ultrasound quickly attenuates when the distance increases, resulting that heartbeat information contained in the received signals weakens. Compared with the single microphone system, the maximum sensing distance of our system is 1.7 m. And even at 1.7 m, the achieved accuracy is still high enough to meet the requirement of daily heartbeat monitoring.

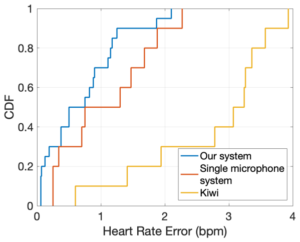

Fig. 6. Estimation of heart rate.

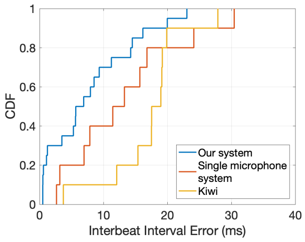

Fig. 7. Estimation of interbeat interval.

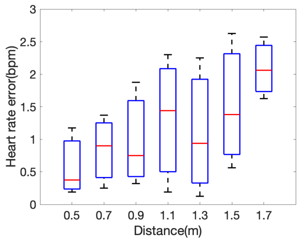

Fig. 8. Impact of user-device distance.

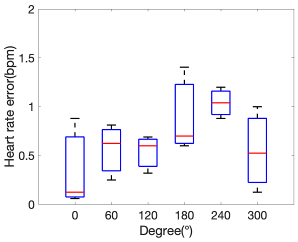

Fig. 9. Impact of user orientation.

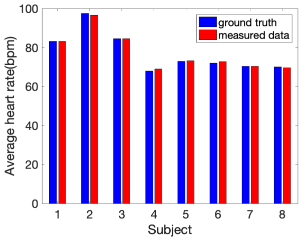

Fig. 10. Impact of different subjects.

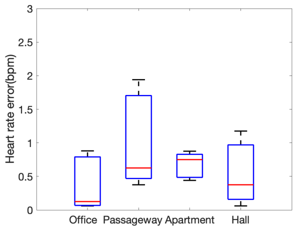

Fig. 11. Impact of different environments.

D. Impact of User Orientation

To evaluate the impact of the user orientation, we keep the speaker facing the human body all the time, simulating the rotary speaker in the smart speaker. Then we ask users to sit surrounding the microphone array, sitting in the range of 0-360 degrees with a step of 60 degrees. Fig. 9 shows the impact of the user orientation. The achieved median errors are all less than 2 bpm when users sit in different orientations. This is because our passive beamforming method enables our system to monitor the heartbeat omnidirectionally.

E. Impact of Different Subjects

In order to evaluate the impact of different subjects, we recruit 8 participants including 6 males and 2 females and monitor each participant for 10 mins. The participants have different ages (20∼30) and body conditions. Fig. 10 shows the estimated heartbeat rates and ground-truth data of 8 participants. Our system achieves high accuracy on diverse subjects. It indicates that our system can accurately monitor heartbeat for different subjects.

F. Impact of Different Environments

We conduct experiments in different indoor environments, including office, corridor, student dormitory, and hall. In each environment, we record heartbeat for 10 mins. As shown in Fig. 11, the achieved median errors of our system in the office, corridor, student dormitory, and hall are 0.125 bpm, 0.625 bpm, 0.75 bpm, and 0.375 bpm, respectively. There is no obvious difference in the four environments. It indicates that our system is not sensitive to the above four indoor environments.

IV. Conclusion

In this work, we propose a contactless and cost effective heartbeat monitoring system using a commodity speaker and a microphone array. We explore a passive beamforming method integrated with the above devices to demonstrate the robustness of contactless heartbeat monitoring using the acoustic signal. As the next step, we will pack the devices to be a smart speaker, provide our system to patients in hospitals and further improve the system in real applications.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants No. 62072450 and No. 61802373, and Open Research Fund Program of Beijing Key Laboratory on Integration and Analysis of Large-scale Stream Data.

References

[1] Kiwi application. http://patrickhealthappreview.blogspot.com/2014/06/app-review-kiwi-instant-heart-rate.html, 2014.

[2] Amazon echo dot 2nd generation. https://www.amazon.com/All-New-Amazon-Echo-Dot-Add-Alexa-To-Any-Room/dp/B01DFKC2SO, 2019.

[3] How ultrasound sensing makes nest displays more accessible. https://blog.google/products/google-nest/ultrasound-sensing/, 2019.

[4] Turn on ultrasound sensing. https://support.google.com/googlenest/answer/9509981?hl=en, 2019.

[5] Apple watch. https://www.apple.com/watch/, 2021.

[6] Fitbit wrist band. https://www.fitbit.com/, 2021.

[7] Uma-8 usb mic array-v2.0. https://www.minidsp.com/products/usb-audio-interface/uma-8-microphone-array, 2021.

[8] Worldwide smart speaker market evenue. https://www.statista.com/statistics/1022823/worldwide-smart-speaker-market-revenue/, 2021.

[9] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and C. Robert Miller. Smart homes that monitor breathing and heart rate. *CHI*, pages 837–846, 2015.

[10] Nicholas D Giardino, Paul M Lehrer, and Robert Edelberg. Comparison of finger plethysmograph to ecg in the measurement of heart rate variability, 2002.

[11] Jingjing Hu, Yunze He, Jie Liu, Min He, and Wenjin Wang. Illumination robust heart-rate extraction from single-wavelength infrared camera using spatial-channel expansion. *EMBC*, pages 3896–3899, 2019.

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

[13] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. *GetMobile*, pages 45–57, 2015.

[14] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. *IEEE INFOCOM*, pages 1574–1582, 2018.

[15] Salman Muhammad Raheel, James Coyte, Em Faisel Tubbal, Raad Raad, Philip Ogunbona, Christopher Patterson, and Dana Perlman. Breathing and heartrate monitoring system using ir-uwb radar. *ICSPCS*, pages 1–5, 2019.

[16] Eugenia María Torres, A. Marcelo Colominas, Gastón Schlotthauer, and Patrick Flandrin. A complete ensemble empirical mode decomposition with adaptive noise. *ICASSP*, pages 4144–4147, 2011.

[17] Wenjin Wang, C den Albertus Brinker, Sander Stuijk, and de Gerard Haan. Robust heart rate from fitness videos. *PHYSIOLOGICAL MEASUREMENT*, pages 1023–1044, 2017.

[18] P Bryan Yan, H S William Lai, K Y Christy Chan, C K Alex Au, Ben Freedman, C Yukkee Poh, and Ming-Zher Poh. High-throughput, contact-free detection of atrial fibrillation from video with deep learning. *JAMA CARDIOLOGY*, pages 105–107, 2020.

[19] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. Your smart speaker can ”hear” your heartbeat! *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2020.

[20] Mingmin Zhao, Fadel Adib, and Dina Katabi. Emotion recognition using wireless signals. *MobiCom*, pages 95–108, 2016.