

Noncontact Respiration Detection Leveraging Music and Broadcast Signals

Wentao Xie, Runxin Tian, Jin Zhang, *Member, IEEE*, and Qian Zhang¹, *Fellow, IEEE*

Abstract—Recent works have shown that acoustic signals can be leveraged to perform respiration monitoring with high accuracy and low energy consumption. Since smartphones, smart speakers, and many other IoT devices are already equipped with microphones and speakers, it is convenient to implement the acoustic sensing solutions on those devices. However, the existing technologies require the speaker to transmit certain ultrasonic signals to detect respiration. Although these signals are inaudible to adults, they are audible to children and pets and they may even have negative impacts on plants. In this article, instead of using ultrasonic signals, we are trying to leverage audible signals in daily lives, e.g., music or broadcasting audios, to detect human respiration. We design a respiration detection system which derives the respiration rate by continuously estimates the channel impulse response (CIR) using music and broadcast signals. We study the intersymbol interference (ISI) brought by the randomness of music and broadcast signal and give our strategy to minimize the interference. We also propose several techniques to resolve some practical issues, such as the multipath effect and sampling frequency offset between the speaker and the microphone. Extensive experiments are conducted to demonstrate the feasibility of our system. The result shows that our system can achieve high respiration detection accuracy with the mean error of less than 0.5 BPM when different audio signals are used.

Index Terms—Channel impulse response (CIR), music and broadcast signals, respiration detection.

Manuscript received May 15, 2020; revised June 29, 2020 and August 5, 2020; accepted August 28, 2020. Date of publication September 4, 2020; date of current version February 4, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61701216; in part by the Shenzhen Science, Technology and Innovation Commission Basic Research Project under Grant JCYJ20180507181527806; in part by the Guangdong Provincial Key Laboratory under Grant 2020B121201001; in part by the Guangdong Innovative and Entrepreneurial Research Team Program under Grant 2016ZT06G587; in part by the Shenzhen Sci-Tech Fund under Grant KYTDPT20181011104007; in part by the Research Grants Council under Contract CERG 16204418 and Contract R8015; and in part by the Guangdong Natural Science Foundation under Grant 2017A030312008. (Corresponding authors: Jin Zhang; Qian Zhang.)

Wentao Xie is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, and also with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: wxieaj@cse.ust.hk).

Runxin Tian is with the School of Computing, National University of Singapore, Singapore (e-mail: tianrunxin@u.nus.edu).

Jin Zhang is with the Guangdong Provincial Key Laboratory of Brain-Inspired Intelligent Computation, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: zhang.j4@sustech.edu.cn).

Qian Zhang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: qianzh@cse.ust.hk).

Digital Object Identifier 10.1109/JIOT.2020.3021915

I. INTRODUCTION

RECENT years have witnessed the surge of ubiquitous healthcare technologies [4]. Many of these technologies aim at monitoring the user's respiration rate. The traditional respiration monitoring systems, such as thoracic impedance pneumography [5] and capnography [6], require the subjects wearing sensors which is time consuming and it is not a pleasant experience for the patients. Since respiration rate measurement is so important to healthcare [7], a more convenient way of respiration monitoring is highly desired.

Recent works have shown that respiration can be monitored in a cost efficient and noncontact way. Benetazzo *et al.* [8] and Nam *et al.* [9] proposed that respiration can be monitored using cameras. However, these systems usually require a good lighting condition [8] and not energy efficient. Besides, image-based systems face the risk of privacy leakage. Researchers also propose that RF radars can be used to detect human respiration [4], [10], [11]. Although these systems accurately detect human respiration in a noncontact way with no risk of privacy leakage, the cost of radars is prohibitive. Wang *et al.* [12] and Liu *et al.* [13] proposed that WiFi signals can be leveraged to detect human respiration. However, since WiFi transceivers are often fixed, those systems can work only when the user is in the designated areas.

In recent years, researchers have turned their attention into acoustic sensing technologies mainly because the audio transceivers (i.e., speakers and microphones) are cost efficient and the demand for smart speakers, such as Amazon Echo [14], has grown rapidly. Generally, an acoustic sensing system transforms the speaker and microphone into a sonar system. By carefully designing the transmitted signal and analyzing the received signal, disturbances around the acoustic system, such as chest movement caused by respiration, can be detected [15]–[17].

The transmitted signals in the above-mentioned works are mostly ultrasonic (e.g., chirps and OFDM symbols) working at the frequency band around 20 kHz. We have one major concern about these ultrasonic signals: although these signals are inaudible to adults, they are still audible to children [1], [18] and pets [2], [19], and the ultrasounds may also have negative impacts on plants [3]. Specifically, research in [18] shows that children at 5–14 years of age can even detect the 25-kHz stimulus at 57-dB intensity while adults at 16–20 years old cannot. Warfield [19] gave the approximate maximum hearing range of cats and dogs to be 45 and 67 kHz, respectively, which is way over the frequency band of the recently proposed acoustic systems. Consequently, it may not be safe to use

ultrasonic-based sensing devices in a home scenario, especially for those families who have children and pets.

In this work, instead of using ultrasounds, we are trying to use the audible sounds in daily lives, such as music and broadcast signals to detect human respiration. In this way, one can get his/her respiration monitored while listening to music or radio. However, it is challenging to achieve accurate respiration detection using nonpredefined signals, such as music, the following explains why. The existing works achieve object detection using acoustic signals through 1) phase extraction [20], [21]. Because the signal's reflection path is changing due to the moving object, the phase of the signal is changing accordingly. If the transmitted signal has a stable phase (e.g., continuous wave (CW) or frequency modulated CW (FMCW)), the object's moving trajectory can be obtained by extracting the phase change of the received signal; 2) Doppler effect [22]. Since a moving target gives rise to a frequency shift with an amplitude proportional to the moving speed, the speed information of the target can be derived by examining the frequency of the received signal if the frequency components of the sent signal are stable (e.g., CW); 3) channel estimation [23], [24]. Channel impulse response (CIR) indicates the geographical information of the surrounding reflectors of the transceivers including the target object. Thus, by estimating the CIR, the target's location can be derived. Above are the most commonly used methods for acoustic signals. However, applying these techniques to music or broadcast signals is challenging. Since music and broadcast signals are random in nature, they do not have a stable phase. Thus, extracting respiration information from the phase of the received signal is difficult. Music and broadcast signals have varying frequencies. Hence, it is almost impossible to detect the slight frequency shift caused by human respiration. Therefore, we turn our attention to detecting human respiration through channel estimation.

In this work, we use a CIR estimation technique to extract the human respiration. The CIR provides us with the distance information of the surrounding reflectors. Since the human chest is one of those reflectors, we can design an algorithm to extract the chest's motion from CIR. However, the most commonly used channel estimation methods require the transmitted signals to have a good autocorrelation property [25]. It should also successfully handle intersymbol interference by adding cyclic prefixes or guard intervals. In the existing works [23], [24], [26], [27], the transmitted signals are well designed to have the above properties. Nevertheless, our system estimates the CIR using music and broadcast signals which do not have such properties. Thus, the key challenge of our system rises: how to accurately estimate the CIR using music and broadcast signals? We also face some practical challenges. First, there are multiple reflection paths between the speaker and the microphone. We need to overcome the multipath effect to achieve accurate breathing pattern extraction. Second, the frequency offset between the speaker and the microphone exists because they may use separate oscillators. The frequency shift influences the estimation of CIR and the impact accumulates over time.

In this article, we propose a respiration monitoring system that overcomes the above challenges using music and

broadcast signals. We observe that although music or broadcast signal has poor correlation property in the time domain, we can still estimate the channel frequency response (CFR) in the frequency domain. Then the CFR can be converted to the CIR using inverse Fourier transformation. The resulting CIR is accurate enough to extract human respiration. We also observe that since human respiration is periodic, we can easily separate the dynamic paths (caused by breathing) from the static paths (caused by static objects). In this way, we can extract the chest motion from the CIR in a multipath environment. Additionally, the line-of-sight (LOS) path is a static path and the lag of the LOS path should be a constant. Thus, we can use this property to solve the frequency offset problem by aligning all the estimated CIRs by their LOS path. In our system, there are two main components. In the first component, we leverage an algorithm to estimate the CIR using music or broadcast signals. In the second component, we design a respiration pattern extraction algorithm to extract the respiration signal from the CIR. In the channel estimation algorithm, we use a certain strategy to minimize the intersymbol interference brought by the randomness of the transmitted signal. In the respiration extraction algorithm, we handle the multipath effect problem and we design a synchronization scheme to compensate the frequency offset.

We highlight the main contributions of our work as follows.

- 1) We propose a human respiration monitoring system using daily life audible signals, such as music and broadcast signals. To the best of our knowledge, we are the first to achieve acoustic-based respiration monitoring without using predefined signals.
- 2) We use random signals, such as music and broadcast to estimate the CIR and we propose a strategy to minimize the interframe interference caused by the randomness of the signal. We propose a respiration pattern extraction algorithm to extract the respiration pattern from the estimated CIR. The respiration pattern extraction algorithm resolves some practical issues, such as multipath effect and frequency offset. We also use a search-refine strategy to reduce the power consumption and latency of our system.
- 3) We implement our system on a home audio system to demonstrate the feasibility of our system. Extensive experiments are conducted and the experimental result shows that our system can achieve high respiration detection accuracy.

The remainder of this article is organized as follows. Section II gives an overview of our design. Sections III and IV present the design details of the system. Section V evaluates the performance of our system. Section VI discusses the related work of this article. Section VII discusses the limitations and future directions of this work. Section VIII concludes this article.

II. SYSTEM OVERVIEW

In this section, we give an overview of our system. As has been stated in the previous section, we derive the human respiration pattern by estimating the CIR. The principle of this design is as follows. The CIR indicates the distance

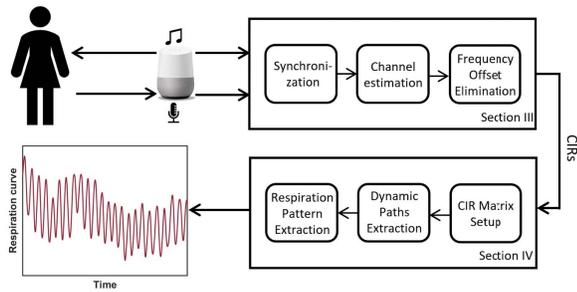


Fig. 1. System diagram.

and reflective property of the surrounding reflectors. In our system, the user's chest is one of the reflectors. Thus, the CIR of our system contains the distance information of the user's chest. Therefore, our system performs CIR in the first place. However, simply estimating the channel is not enough. Since there is more than one reflector in the environment, the distance information of the user's chest, along with other reflectors, are mixed in the CIR. Thus, our system needs to extract the user's chest motion from the CIR. Therefore, we design two modules for our system. The first module estimates the CIR using music or broadcast signals. The second module extracts the user's respiration pattern from the CIR. The workflow of our system is shown in Fig. 1. The following two sections describe the working process of these two modules.

A. CIR Estimation

This module estimates the CIR using music and broadcast signals. Our system first synchronizes the received signal with the originally transmitted signal by performing cross-correlation on the two signals. Next, the system cuts the original signal and the received signal into segments. Then, the system computes CIRs using those audio segments. However, as the speaker and the microphone use separate oscillators, a frequency offset may exist which will result in CIR drifting, i.e., the CIR drifts over time. We compensate for the frequency offset by an observation that the LOS path of a CIR is a static path and the LOS path always has large power. At the end of this module, the system stores a series of CIRs as input to the next module. The details of this module are described in Section III.

B. Respiration Detection

This module detects the respiration based on the prestored CIRs. The system first stacks a number of CIRs to form a matrix called a CIR matrix. The CIR matrix indicates how the CIR is changing over time. Second, our system separates the dynamic paths, caused by the moving chest, from the static paths, caused by the surrounding stationary objects, by performing autocorrelation on the CIR matrix. The autocorrelation function measures the periodicity of a sequence. Third, the system tracks all peaks' moving trajectories of the dynamic paths. Finally, the system selects the peak trajectory with the best periodicity. The chosen peak trajectory directly represents the breathing pattern. Section IV presents the design details of this module.

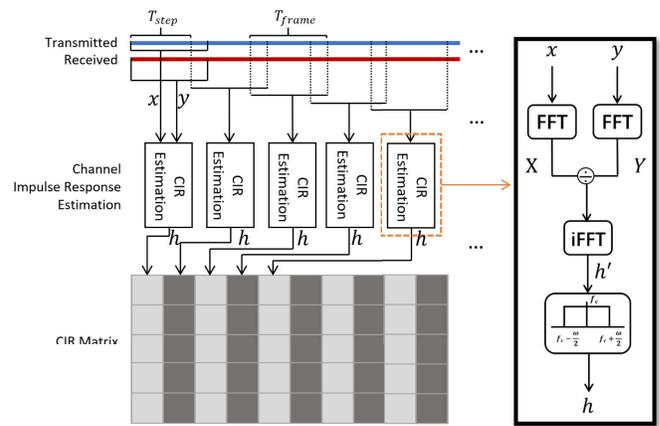


Fig. 2. Process of CIR estimation.

III. CHANNEL IMPULSE RESPONSE ESTIMATION VIA MUSIC AND BROADCAST SIGNALS

In this section, we discuss how our system estimates the CIR using music and broadcast signals. The key challenge in this module is how to estimate the CIR using audio signals with poor autocorrelation property and how to assure the quality of the estimated CIR. In addition, since it is often the case that the speaker and the microphone use separate oscillators, there is a sampling frequency offset between the two devices that degrades the CIR estimation accuracy. Unlike the commonly used carriers, such as CW or FMCW where frequency offset can be easily eliminated by tracking the constant frequency change over time [28] or by tracking the drifting of the peak frequency [29], it is difficult to directly compute the frequency offset from music or broadcast signals because they are random in nature. Hence, the second challenge in this module is how to eliminate the frequency offset based on music and broadcast signals. The overall process of this module is shown in Fig. 2. We first synchronize the received signal with the transmitted one. Second, we compute the CIR in the frequency domain. Third, our system eliminates the impact of frequency offset caused by the clock difference between the speaker and the microphone. At the end of this section, we analyze the interference brought by our CIR estimation method and give our strategy to reduce the impact of the interference. The following sections elaborate on the details of this module.

A. Synchronization

First, our system needs to know when the microphone starts to receive the transmitted signal. Thus, our system performs synchronization in the first place. The system computes the delay of the received signal by performing a cross-correlation between the original signal and the received signal as follow: $R(t) = x_{\text{ori}}(t) * x_{\text{rec}}^*(-t)$, where $x_{\text{ori}}(t)$ and $x_{\text{rec}}(t)$ represent the transmitted original signal and the received signal, respectively. The lag of the maxima of $R(t)$ indicates the delay of the received signal and is the offset that the received signal needs to be shifted.

B. Channel Impulse Response Estimation

In this section, we elaborate on our approach to estimating the CIR. First, our system performs segmentation on the synchronized audio signals, then the system computes the CIR using the audio segments.

1) *Audio Signal Segmentation*: Since the system is designed to monitor the human respiration continuously, the system is required to estimate the CIR frequently. In a wireless communication system, channel information is often estimated frame by frame, e.g., in WiFi systems. However, since music and broadcast signals are continuous over time, they have no frame structures where chunks of data are separated by blank intervals. Thus, in our system, we manually cut the audio signals into segments. In detail, in every T_{step} seconds, our system backtracks T_{frame} seconds of the original signal to form a signal segment, and the same goes to the received signal. After that, every received signal segment is paired up with an original signal segment. The received signal segment can be viewed as the summation of multiple delayed copies of the original segment. Note that here, T_{frame} plays an essential role in deciding the channel estimation accuracy. Section III-D later in this section discusses how to choose a proper T_{frame} . The system stores the original/received signal pairs for CIR estimation.

2) *Channel Impulse Response Estimation*: Let $x[n], y[n]$ denote an original signal segment and its corresponding received signal segment. Let $X[k]$ and $Y[k]$ denote the discrete Fourier transformation of $x[n]$ and $y[n]$. Then, the following equation holds: $y[n] = x[n] * h[n]$, where $h[n]$ is the CIR and n is from 0 to $N-1$ where N is the segment length. According to the convolution property of discrete Fourier transform [30], the corresponding frequency response of CIR is the quotient of the Fourier transformation of $y[n]$ and $x[n]$, i.e.,

$$\hat{H}[k] = \frac{Y[k]}{X[k]}, \quad k = 0 \text{ to } N-1.$$

Then our system applies inverse Fourier transform (iFFT) to $\hat{H}[k]$ to obtain the CIR $\hat{h}[n]$, i.e.,

$$\hat{h}[n] = \text{iFFT}(\hat{H}[k]).$$

Fig. 3(a) shows a CIR generated by our system. We further smooth the computed CIR by a low-pass filter. The filtered results are shown in Fig. 3(c).

Note that some signal segments may contain blank intervals (e.g., pauses between two words in a piece of a broadcast signal), and few audio signals are transmitted in these cases where the system may fail to generate a valid CIR. Fig. 3(b) shows an invalid CIR. The system detects those invalid frames by measuring the standard deviation of the CIR. If a frame's standard deviation exceeds a threshold Th_{σ} , the system discards the frame and copies the last frame to the current frame. The revised version of CIR in Fig. 3(b) is shown in Fig. 3(d).

C. Frequency Offset Elimination

In practical use, the frequency offset between the speaker and the microphone may exist. This is because the speaker and the microphone may use separate crystal oscillators. Due to imperfect hardware, the sampling rate of the microphone and

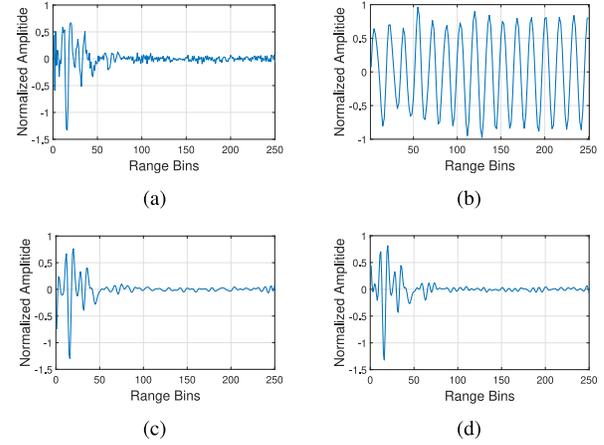


Fig. 3. (a) and (b) are the CIRs generated by the system. (c) is the filtered signal of (a). (d) is the altered signal of (b).

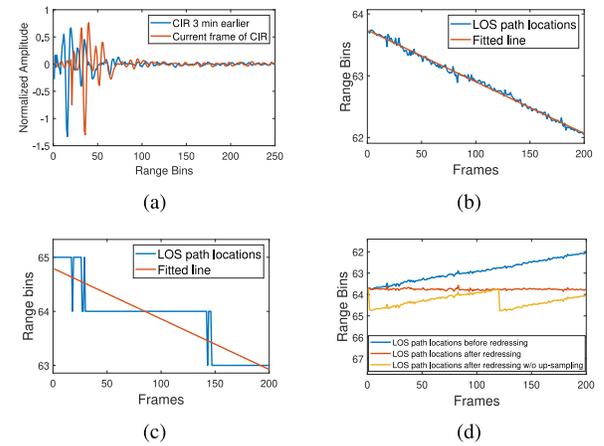


Fig. 4. (a) CIR drifting effect. (b) Distance sequence of the LOS path and the fitted line (with upsampling). (c) Distance sequence of the LOS and the fitted line (without upsampling). (d) Distance sequence of the LOS path after CIR redressing.

the speaker may not be exactly the same. This will lead to the speaker and the microphone experiencing different times when sending or receiving the same length of data. Since our system continuously cuts the original signal and the received signal into segments of the same length, the time offset between each pair of signals will accumulate and the resulting CIR will drift over time. For example, Fig. 4(a) shows a CIR as well as the CIR estimated by our system a short time earlier. We can see that the CIR is delayed over time.

We eliminate this effect by an observation that the LOS path is a static path and should always have a large power. This is because in a smart speaker, the microphone and the speaker are placed close to each other. Our system first measures the CIR drifting speed by tracking the drift of the LOS path, then the system compensates the offset frame by frame. The following presents the design details.

1) *Measuring the CIR Drifting Speed*: For the first step, the system upsamples each CIR 100 times by performing cubic spline interpolation. The spline interpolation improves the distance resolution of CIR thus providing us with more detailed information about the peak locations.

Algorithm 1: Peak Tracking Algorithm

Data: The list of all CIRs C , the acceptance window size w , the minimum peak height p

Result: The list of all distance sequences DS

```

1  $DS \leftarrow emptyList()$ ;
2  $i \leftarrow 1$ ;
3 while  $i \leq length(C)$  do
4    $locs \leftarrow$  locations of peaks in  $C[i]$  with height  $\geq p$ ;
5   if  $isEmpty(DS)$  then
6      $j \leftarrow 1$ ;
7     while  $j \leq length(locs)$  do
8        $ds \leftarrow emptyList()$ ;
9        $ds \leftarrow append(ds, locs[j])$ ;
10       $DS \leftarrow append(DS, ds)$ ;
11       $j \leftarrow j + 1$ ;
12   end
13 else
14    $j \leftarrow 1$ ;
15    $itr \leftarrow 1$ ;
16   while  $j \leq length(DS)$  do
17      $ds \leftarrow DS[j]$ ;
18      $DS[j] \leftarrow append(ds, ds[i - 1])$ ;
19     while  $itr \leq k \leq length(locs)$  do
20       if  $||ds[i] - locs[k]|| \leq w$  then
21          $ds[i] \leftarrow locs[k]$ ;
22          $DS[j] \leftarrow ds$ ;
23          $itr \leftarrow itr + 1$ ;
24       break
25     end
26      $k \leftarrow k + 1$ ;
27   end
28    $j \leftarrow j + 1$ ;
29 end
30 end
31  $i \leftarrow i + 1$ ;
32 end

```

The LOS path has large power, and the system uses this property to track the drifting trajectory of the LOS path. The system first tracks all high peaks' locations using a peak tracking algorithm described in Algorithm 1. Then the system selects the peak whose moving trajectory has the best linearity as the LOS path. The linearity is measured by the Pearson correlation coefficient. The basic idea of the peak tracking algorithm is: the algorithm locates every high peak frame by frame. If two peaks in the adjacent frames are close to each other, we consider the two peaks belong to the same peak clan. The algorithm records all peak clans' location sequences. We call a peak clan's location sequence a distance sequence.

Next, we fit the LOS path's distance sequence with a line. The line shows the trend of the CIR drifting and its slope, denoted by k , is the drifting speed we want. An example of the distance sequence of the LOS path and its fitted line is shown in Fig. 4(b).

2) *Compensating the Frequency Offset:* The final step is to redress the CIRs to compensate for the drift. Let Δn_i denotes

the number of samples that the i th CIR needs to be shifted forward or backward. We calculate Δn_i as

$$\Delta n_i = i \cdot k.$$

Next, we downsample the CIRs 100 times to restore the original sampling rate. An example of the redressed CIR is shown as the red line in Fig. 4(d). The blue line shows the original distance sequence of the LOS path. Thus, the impact of the frequency offset is successfully eliminated.

3) *Upsampling Scheme:* Note that we use an *upsampling scheme* to solve the clock difference problem. We do so for the following two reasons.

- 1) As is stated before, upsampling the CIRs using cubic spline interpolation improves the distance resolution of the CIR. Thus, the distance sequence is smoother after upsampling. The obtained distance sequence of the LOS path without upsampling and its fitted line is shown in Fig. 4(c). Readers can compare it with Fig. 4(b).
- 2) The drifting speed k is usually not an integer. Without upsampling, we can only shift the CIR when the drifting speed accumulates to an integral value, i.e., $i \cdot k \geq 1$. For example, if $k = 0.15$, we can only shift the CIR one sample forward or backward after seven frames of CIRs. With 100 times of upsampling, we can adjust the CIR every frame if the drift offset is greater than 0.01 which is an easy criterion to meet in practice. The result of CIR redressing without upsampling is shown as the yellow line in Fig. 4(d).

D. Combating the Interference Brought by the Randomness of Music and Broadcast Signals

In a wireless communication system, to estimate the channel model usually requires the transmitted signal to be well designed for the sake of accuracy [31]. However, music and broadcast signals, which can be viewed as random signals, are not designed for channel estimation. Besides, since we manually cut the audio signal into segments, there is no blank interval between segments. Hence, interference between successive frames of CIRs may occur. In this section, we first briefly discuss the mathematical model of our system. Then, we analyze how the interference is brought to our system and we propose our strategy to reduce the interference.

1) *Mathematical Model of Our System:* Suppose the wireless channel around the device has L propagation paths, and the i th path is associated with a transmission delay τ_i as well as an attenuation coefficient α_i . The impulse response of the wireless channel is

$$h[n] = \sum_{i=1}^L \alpha_i \delta[t - \tau_i] \quad (1)$$

and the corresponding CFR is

$$H[k] = \sum_{i=1}^L \alpha_i e^{-j2\pi \tau_i k / N}. \quad (2)$$

Assume that the system arrives at a time point, and it should calculate the CIR now. Let x denotes the latest segment of the

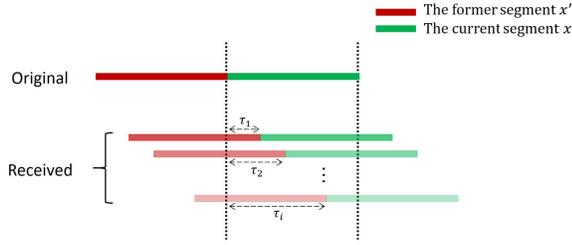


Fig. 5. Details of audio signal segmentation.

original signal and x' denotes the former segment (as shown in Fig. 5). Then, the received segment is of the following formulation:

$$y[n] = \sum_{i=1}^L \alpha_i (x[n - \tau_i] + x'[n + N - \tau_i]), \quad n = 0 \text{ to } N - 1. \quad (3)$$

Let $\beta'[n]$ denotes the difference between $x'[n]$ and $x[n]$, i.e.,

$$\beta'[n] = \frac{x'[n]}{x[n]}, \quad n = 0 \text{ to } N - 1 \quad (4)$$

and for the sake of simplicity, we let

$$\beta_i[n] = \begin{cases} 1, & \text{for } n = 0 \text{ to } N - \tau_i - 1 \\ \beta'[n], & \text{for } n = N - \tau_i \text{ to } N - 1. \end{cases} \quad (5)$$

The system first performs FFT on the received $y[n]$ to calculate its Fourier transform $Y[k]$

$$Y[k] = \sum_{i=1}^L \alpha_i e^{-j2\pi \tau_i k/N} \sum_{n=0}^{N-1} \beta_i[n] x[n] e^{-j2\pi kn/N}. \quad (6)$$

In order to get the CFR, the system divides $Y[k]$ by $X[k]$, i.e.,

$$\hat{H}[k] = \frac{Y[k]}{X[k]} = \sum_{i=1}^L \alpha_i e^{-j2\pi \tau_i k/N} \frac{\sum_{n=0}^{N-1} \beta_i[n] x[n] e^{-j2\pi kn/N}}{\sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}}. \quad (7)$$

Then, for simplicity, we let

$$\Gamma_i[k] = \frac{\sum_{n=0}^{N-1} \beta_i[n] x[n] e^{-j2\pi kn/N}}{\sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}}. \quad (8)$$

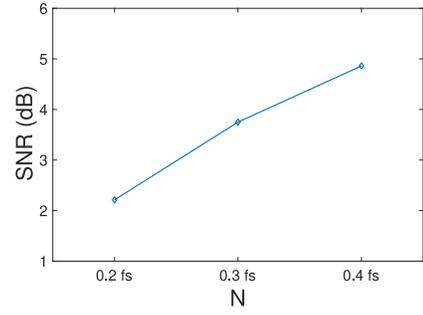
Now $\hat{H}[k]$ becomes the summation of the multiplication of two signals, i.e.,

$$\hat{H}[k] = \sum_{i=1}^L \alpha_i e^{-j2\pi \tau_i k/N} \cdot \Gamma_i[k]. \quad (9)$$

After that, the system applies the inverse Fourier transformation to $\hat{H}[k]$ to compute the CIR, i.e., $\hat{h}[n]$. According to the time-shifting property and the multiplication property of discrete-time Fourier transform [30]

$$\hat{h}[n] = \sum_{i=1}^L \alpha_i \delta[n - \tau_i] * \gamma_i[n] \quad (10)$$

where $\gamma_i[n]$ is the time-domain signal transformed from $\Gamma_i[k]$.

Fig. 6. SNR of $\gamma_i[n]$ when different N is selected.

2) *Interference and Its Reduction Strategy*: First, we analyze how interference is introduced to our system. Comparing (1) with (10) and (2) with (9), the interference is introduced because of the existence of $\gamma_i[n]$ or $\Gamma_i[k]$. The estimated CIR is valid only if the impact of $\gamma_i[n]$ or $\Gamma_i[k]$ is limited which means $\Gamma_i[k]$ should be close to a vector of ones and $\gamma_i[n]$ should be close to a pulse signal according to (9) and (10).

Next, we present our strategy to reduce interference. From (8), $\beta_i[n]$ plays a major role in deciding the value of $\Gamma_i[k]$. In order to obtain a good CIR, $\beta_i[n]$ should be as close to one as possible. It is easy to infer from (5) that we can increase the frame length N , to get a better $\beta_i[n]$. An intuitive observation is that a large value of N means that the frame is long enough so that the effect of the delay profile does not significantly affect the CIR. The following explains why simply increasing the frame length can work.

Since the power decreases sharply along with the distance an acoustic signal travels, and in a room setting, walls together with furniture will eventually reflect or absorb all the signals, there is usually a maximum transmission distance of an acoustic signal. Assume that this maximum transmission distance is D and suppose the sampling rate of our system is f_s , given that the speed of sound is v_s , the delay τ_i is limited by $\tau_i \leq (2Df_s/v_s)$. Let η be the ratio of τ_i and N . The η indicates the percentage of none-one values in $\beta_i[n]$, then $\eta = (\tau_i/N) \leq (2Df_s/v_s N)$. It is obvious that when N becomes larger, η becomes smaller. In this way, more elements in $\beta_i[n]$ are one. Thus, $\Gamma_i[k]$ has less impact on $\hat{H}[k]$ and $\gamma_i[n]$ has less impact on $\hat{h}[n]$. Consequently, the interference is reduced. For example, let $D = 4$ m and given that the speed of sound is 340 m/s and suppose N/f_s has the value of 0.4 s (which is a relatively large value) in our system. Then $\eta \leq 0.059$. That means over 94% of the elements in $\beta_i[n]$ are 1. In this way, $\Gamma_i[k]$ is close to an ones vector and $\gamma_i[n]$ is close to a unit pulse. Fig. 6 shows the average SNR of $\gamma_i[n]$ s of a 180-s music signal with different segment length N . The SNR is calculated as: $\text{SNR} = \tau_i[0]^2 / \sum_{k=1}^{N-1} \tau_i[k]^2$.

Note that although increasing N can reduce the impact of interference, a large N may decrease the time resolution of our system. This is because we use N samples of an audio signal to compute a frame of CIR. If N is too large, the chest may move a long distance during the time of a frame. In this way, the CIR may fail to indicate the current chest position. Thus, in our system, we empirically set N to be $0.4 f_s$.

TABLE I
NOTATIONS

Symbol	Meaning	Default
T_{frame}	The duration of the original/received signal segments.	0.4s
T_{step}	The step interval between two adjacent original/received signal segment.	0.1s
K	The amount of CIRs in a CIR matrix.	200
T_{update}	The updating period of CIR matrix.	2s

IV. RESPIRATION DETECTION USING CHANNEL IMPULSE RESPONSE

After obtaining the CIRs, we need to derive the respiration pattern from them. Since generally, the system is in a multipath environment, the signals reflected by the user's chest, together with signals reflected by other surrounding objects, are collected by the microphone. From the CIR's perspective, the channel taps (or paths) that caused by the reflection of the user's chest are mixed with channel taps caused by other objects. Thus, the key challenge of this module is to mitigate the multipath effect so that chest motion can be correctly derived from the CIR. The module works in the following steps. We first set up a CIR matrix using a series of CIRs to derive the locations of the channel taps that are caused by breathing based on the fact that the chest moves periodically when breathing. Then, the system separates the channel taps caused by breathing from the channel taps caused by other objects. Next, the system derives the breathing pattern based on the variation of separated channel taps. Additionally, We use a *search-refine strategy* to efficiently extract the breathing pattern. The following sections elaborate on the details of this module. Table I shows the basic definitions used in this section.

A. CIR Matrix Setup

When our system is running, the system stacks K CIRs to form a CIR matrix and the CIR matrix is updated every T_{update} , i.e.,

$$\mathbf{M}_{CIR}(\mathbf{t}) = \left[\mathbf{h}'_{\lfloor t/T_{update} \rfloor}, \mathbf{h}'_{\lfloor t/T_{update} \rfloor + 1}, \dots, \mathbf{h}'_{\lfloor t/T_{update} \rfloor + K - 1} \right]$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The CIR matrix shows the dynamic change of the CIR with time. Fig. 7(a) is an example of a CIR matrix. Each row of data belongs to a range bin where a range bin presents the distance that sound propagates from the speaker to the microphone. Each column of data is a frame of CIR. It is obvious that generally, all K frames of CIR are similar. This is because although the person is breathing, the overall environment is almost static except the chest movement. However, if we take a closer look at the CIR matrix from the 350th range bin to the 450th range bin, we could easily observe a vibration pattern. This pattern is caused by breathing and it is what we want to derive from the CIR matrix.

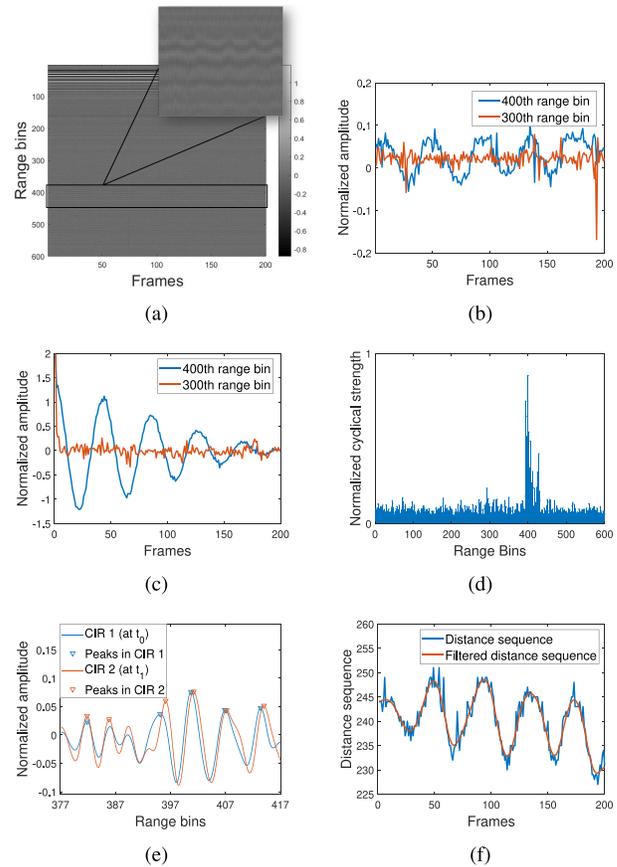


Fig. 7. (a) CIR matrix. (b) Two range sequences. (c) Autocorrelation profile of the two range sequences in (b). (d) Cyclic strength of all the range bins. (e) CIRs at different times. (f) Selected distance sequence and the filtered result.

B. Dynamic Paths Extraction

To extract the vibration pattern from the CIR matrix, first, we need to locate the paths that correspond to the chest vibration. In a CIR, there are multiple reflection paths that consist of the static paths, caused by the surrounding stationary objects, as well as the dynamic paths, caused by the moving chest. In Fig. 7(a), the paths between the 350th range bin to the 450th range bin are dynamic paths and the rest of the paths are static paths. We know that human respiration is almost periodic in nature. Thus, we extract the dynamic paths by leveraging the periodic property of human breathing.

To extract the dynamic paths, we need to select the range bins where the dynamic paths locate. As is analyzed above, the vibration pattern in the CIR matrix is caused by the back and forth movement of the chest. Hence, our objective is to choose the range bins where sequences show strong periodicity. Here, the sequence at a range bin is called a range sequence (i.e., a row in a CIR matrix). For example, Fig. 7(b) shows the range sequence of the 400th and 300th range bin of the CIR matrix in Fig. 7(a). Evidently, the range sequence at the 400th range bin shows strong periodicity while the range sequence at the 300th range bin does not. We measure the periodicity of each range sequence by its autocorrelation as follows. We first standardize the range sequence by calculating the z-scores as $Z_i = [(X_i - \mu)/\sigma]$, where μ and σ are the mean and

the standard deviation of the sequence, respectively. Then, we perform autocorrelation to the standardized range sequence. The autocorrelation function is defined as follows:

$$R(k) = \frac{E[(X_i - \mu)(X_{i+k} - \mu)]}{\sigma^2}$$

where E is the expectation, μ and σ are the mean and the standard deviation of the sequence. The cyclical strength of a range sequence is defined as the summation of the peak-valley differences of the sequence's autocorrelation profile. The higher the cyclical strength is, the more periodic the sequence is. Fig. 7(c) shows the autocorrelation profiles of the corresponding range sequences in Fig. 7(b). After defining the cyclical strength of the range sequence, we identify the range bins where the dynamic paths locate by selecting the range bins with the highest cyclical strength. We compute the cyclical strength of all range bins in Fig. 7(a), and the result is shown in Fig. 7(d). The range bins around the 400th bin show strong periodicity, and these range bins are where the dynamic paths locate.

C. Respiration Pattern Extraction

Since we have selected the dynamic paths from the CIR matrix, one more thing we need to do is to track the dynamic paths' moving trajectory which directly represents the breathing pattern.

We use the peak tracking algorithm described in Section III to obtain all the distance sequences of the dynamic paths. Note that not all the distance sequence are valid. For example, Fig. 7(e) shows the CIR at t_0 and t_1 and the figure is zoomed into the dynamic paths. The three peaks from the 397th range bin to the 417th range bin move away at the same pace while the peak at around the 395th range bin moves a lot and the peaks before the 390th range bin are almost stationary. In this case, the rightmost three peaks are the desired peaks. Again, because human respiration is almost periodic in nature, the system measures the periodicity of each distance sequence by autocorrelation. The system selects the distance sequence with the highest cyclical strength and then passes the sequence through a low pass filter. The resulting curve is the desired breathing pattern. The selected distance sequence, as well as the filtered curve, are shown in Fig. 7(f).

D. Search-Refine Strategy

Note that autocorrelation requires multiplication operation which consumes large computing resources. In the dynamic paths extraction process, the system performs autocorrelation on every range bin. When selecting the best distance sequence, the system performs autocorrelation again on all the distance sequences. Thus, without further optimization, the system will have high energy consumption and have high latency.

Our system uses a search-refine strategy when extracting the respiration pattern to reduce energy consumption and latency. This strategy contains two stages given as follows.

- 1) *Searching Stage*: In the searching stage, the system first extracts the dynamic paths and then extracts the breathing pattern as described above. When the CIR matrix is updated, the system moves into the refining stage.

- 2) *Refining Stage*: In the refining stage, the system continues to track the chosen peak, i.e., appends the selected distance sequence. Meanwhile, the system monitors the cyclical strength of the recent K items of the distance sequence. If the cyclical strength drops below a threshold, the system returns to the searching stage. With the search-refine strategy, the system performs a large amount of calculation only in the searching stage. Thus, the energy consumption is reduced and the latency is reduced.

V. EVALUATION

In this section, we conduct comprehensive experiments to evaluate our system.

A. Experimental Setup

We use an HTC U Ultra smartphone, an Edifier-R1000TC Loudspeaker, and a macOS laptop with 1.4-GHz Quad-Core Intel Core i5 processor to build our system. The loudspeaker is connected to the smartphone. The loudspeaker plays audio pieces while the microphone on the smartphone records the audio signal and transmits the recorded audio to a MATLAB server on the laptop through Wi-Fi. The MATLAB server processes the recorded signal and generates the respiration waveform. The smartphone and the loudspeaker are working with a sampling rate of 48 kHz. We use the respiration module of the SCHUHFRIED Biofeedback system [32] to record the actual respiration rate as the ground truth. The device is a contact sensor system that directly measures the subject's breathing amplitude. The sampling rate of this device is 20 Hz. We recruited 11 volunteers (five males and six females) to participate in the experiments. The volunteers are asked to sit in front of the smartphone and the loudspeaker and the distance between the smartphone/loudspeaker and the subject is set to be 25 cm by default. The volunteers are asked to wear a shirt when doing the experiments, except where noted. We select eight pieces of music and news broadcast to play on the loudspeaker, including two rock songs, two pop songs, two ballad songs, and two news broadcasting pieces. Each piece of audio is 120 s long. In each experiment, the volume of the loudspeaker is set to be 50% of the maximum volume.

We use the following three metrics to evaluate the performance of our system: 1) normalized root-mean-square error (NRMSE), a metric to measure the variation of the error of our system; 2) the 90th percentile error (P90) (in BPM), a metric to measure the overall accuracy of our system; and 3) maximum error (in BPM), a metric to measure the worst possible performance of our system.

B. System Performance Evaluation

We conduct comprehensive experiments to evaluate our system.

- 1) *Overall Evaluation*: An example of the respiration rate estimated by our system during a 6-min song¹ is shown in Fig. 8. Fig. 8(a) shows the comparison of the respiration

¹Yanni. "If I Could Tell You." *If I Could Tell You*. Virgin, 2000.

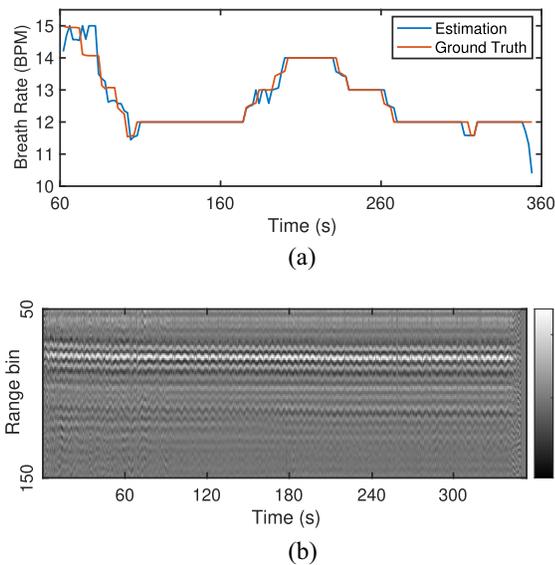


Fig. 8. System's performance under an entire song. (a) Breathing rate estimated by our system versus the ground truth. (b) CIR matrix of the entire song.

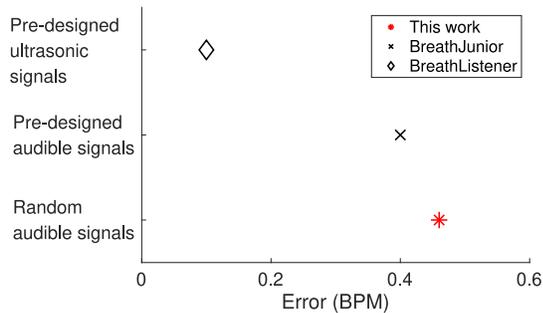


Fig. 9. Mean error and the signal design of our system, BreathListener [17] and BreathJunior [33].

rate estimated by our system with the ground truth. Fig. 8(b) presents all the CIRs estimated by our system during the entire song. A comparison of our system with two similar systems, BreathListener [17] and BreathJunior [33], in terms of signal design and mean error is shown in Fig. 9. Note that these three systems are designed to detect human respiration in very different scenarios. We present this comparison just to show that although we use random audible signals, such as music and broadcasting signals, instead of pre-designed and ultrasonic signals, we still can achieve acceptable respiration detection accuracy. BreathListener targets at monitoring a driver's breathing condition when driving. In their evaluation, the system is implemented on a smartphone and the smartphone is put on the instrument panel, cab door, cup holder, and the driver's pocket. BreathJunior achieves respiration monitoring for infants. They build their system on a smartspeaker. In their experiments, the system is placed 40–60 cm away from the infant. To validate this comparison, the mean error of our work shown in Fig. 9 is the result when the distance between the user and the system is 50 cm. In the implementation of our system, the loudspeaker has an external power source and the smartphone is only responsible for recording and transmitting data to the server. The computational burden is mainly on the

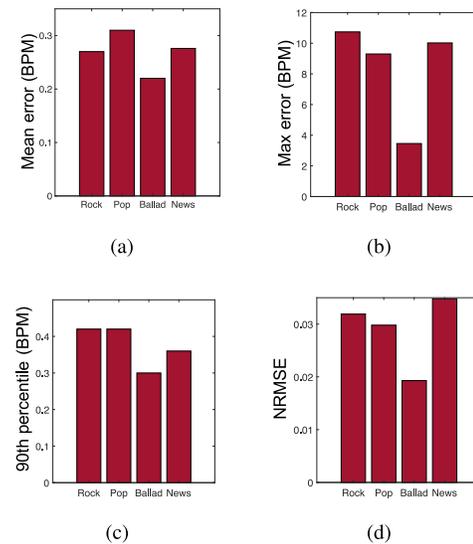


Fig. 10. Evaluation with different kinds of audio signals. (a) Mean error. (b) Maximum error. (c) 90th percentile error. (d) NRMSE.

server. We record the server's CPU usage during our experiments and the result shows that when our system is running, the CPU usage rises from 1% (the idle state) to 15.5% on average.

2) *Evaluation for Different Audio Genres*: We first examine the performance of our system when different kinds of audio signals are used. The system should achieve high respiration detection accuracy when different audio signals are used. In this set of experiments, all the eleven volunteers are involved. Fig. 10 presents the result. Note that the average error is low but the maximum error can be as high as 10 BPM. This is because of the dynamic nature of music and broadcast signals where long blank intervals or intervals with extremely low power exist. When nearly no signal is transmitted by the speaker, the system is unable to correctly extract the breathing pattern. For example, the song used in Fig. 8 has a very low volume at the beginning and end. Fig. 8(b) shows that there are blurs at the beginning and there is nearly no valid CIR at the end. This results in larger estimation errors at the beginning and at the end. In the following experiments, we use a 120 s music piece with nearly no blank intervals from the rock music set to prevent the influence of blank intervals to our experiments.

3) *Impact of the Subject's Orientation*: We conduct experiments with different orientations to figure out how the user's orientation to the speaker influences the system performance. In this set of experiments, three volunteers (two males and one female) are involved. We test the orientation from 0° to 90° as shown in Fig. 11. From the results, we can conclude that, with the increase of the orientation degree, the performance of our system decreases. The result shows that our system can still work when the user is 30° to the loudspeaker.

4) *Impact of the Distance Between the Subject and the System*: We conduct several experiments to examine our system under different distances and three volunteers (two males and one female) participate in this set of experiments. The distance we set between the subject and the devices ranges

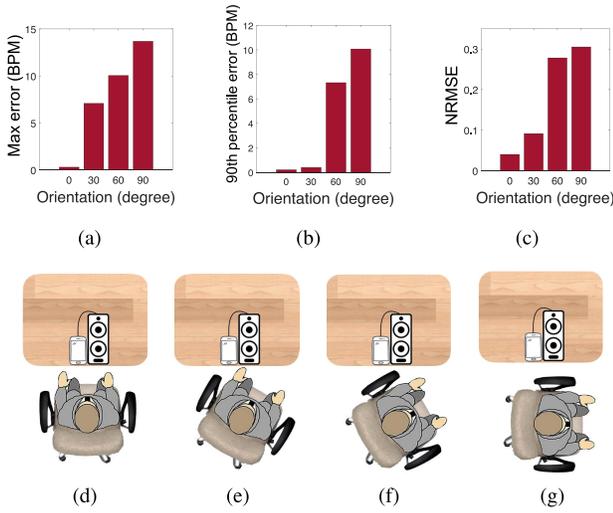


Fig. 11. Evaluation with different orientations. (a) Maximum error. (b) 90th percentile error. (c) NRMSE. (d)–(g) The illustration of rotating the user’s orientation by 0°, 30°, 60°, and 90°.

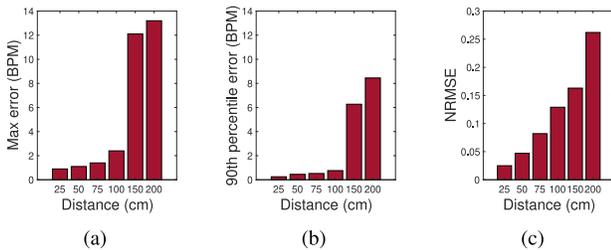


Fig. 12. Evaluation with different distances. (a) Maximum error. (b) 90th percentile error. (c) NRMSE.

from 25 to 200 cm. We also collect the mean acoustic intensities using an Android App² at these distances for the reader’s reference. The acoustic intensities are 78, 74, 69, 65, 58, and 54 dB at 25, 50, 75, 100, 150, and 200 cm, respectively. As is shown in Fig. 12, the experimental result indicates that our system can operate with high accuracy when the distance between the subject and the device is within 100 cm. However, when the distance increases to 150 cm, the accuracy of our system decreases notably. This is reasonable because the energy of the acoustic signal drops sharply when propagating in the air. After traveling 3 m back and forth, the received signal is very weak. In addition, the motion of a chest when breathing is very tiny which makes the respiration hard to detect.

5) *Evaluation Under Multiperson Scenarios:* We evaluate the performance of our system when there are people in the background. We consider the following two scenarios: when a subject is using our system, there is another person (the disturber) either 1) sitting right next to the subject or 2) walking around the subject. Three volunteers (two males and one female) participate in this set of experiments and in each experiment, one of the volunteers is the system user and one of the other two is the disturber. As is shown in Fig. 13, when someone sitting next to the user, the accuracy degrades slightly. This is mainly because even if the speaker and the microphone are facing at the subject, some signals may also be reflected by the

²Sound Meter. ABC Apps, 2020.

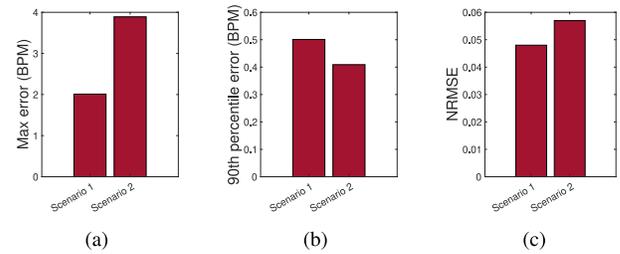


Fig. 13. Evaluation with multiple people scenarios. (a) Maximum error. (b) 90th percentile error. (c) NRMSE.

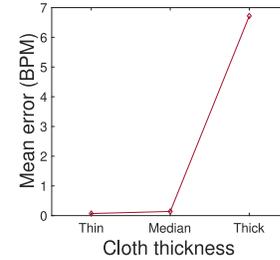


Fig. 14. Influence of clothing.

disturber because of the scattering effect of the sound wave. But still, our system can achieve decent accuracy under this situation. In the second case when someone walks around the user, the accuracy is barely declined because the disturber is out of the reflection path between the device and the subject. One factor that may influence the system performance might be the noise that the disturber makes when walking.

6) *Impact of Clothing:* We also evaluate the influence of clothing. In this experiment, we ask the two subjects (two males) to wear clothes of three levels of thickness: 1) thin (a shirt); 2) median thickness (a shirt and a thin coat); and 3) thick (a shirt and a thick coat). The result is shown in Fig. 14. The result indicates that our system can work with high accuracy when the user is wearing medium-thickness clothes or thin clothes. However, when the clothes are too thick, the system fails to detect respiration.

7) *Impact of Environmental Noise:* We conduct three experiments with two volunteers (two males) to evaluate the performance of our system under noisy environments.

- 1) In the first experiment, another person is sitting behind the subject while using the keyboard. The noise level in this experiment is around 45 dB.
- 2) In the second experiment, another person is sitting behind the subject and reading a newspaper aloud. The noise level is around 54 dB.
- 3) In the third experiment, the subject sits in a room with a high power air conditioner operating. The noise level in this experiment is around 50 dB. The overall result is shown in Fig. 15. The result shows that our system works well in common noisy environments.

VI. RELATED WORK

In this section, we briefly review the existing work on non-contact sensing which can be grouped into two categories: 1) noncontact respiration monitoring methods and 2) acoustic sensing methods.

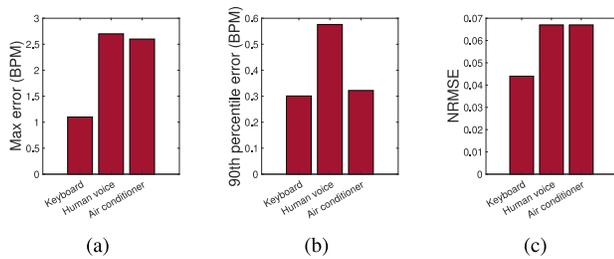


Fig. 15. Evaluation with background noise. (a) Maximum error. (b) 90th percentile error. (c) NRMSE.

A. Noncontact Respiration Monitoring Systems

WiBreathe [11] is a respiration monitoring system based on radio signals, which can achieve high fidelity and noninvasive breathing monitoring. UbiBreathe [34] is a WiFi-based breath monitoring system. It measures the received signal strength (RSS) to extract the user's respiration rate. WiSleep [13] and [12] exploit the channel state information (CSI) in WiFi signals to monitor the respiration. VitalRadio [4] is a smart home device that also uses the phase change of RF signals to monitor people's heartbeat as well as respiration. EZ-Sleep [35] uses an FMCW radar to monitor the users' sleeping status and insomnia. The above systems either require additional hardware like RF radar or need the user to stay at the designated areas. [8] and [9] are camera-based methods that use cameras to record user's chest motion when breathing. Camera-based methods require good lighting conditions and may face the risk of privacy leakage.

B. Acoustic Sensing Systems

In recent years, researchers have turned their attention to acoustic sensing technologies. LLAP [20] uses the phase shift of a sinusoidal signal to estimate the distance change between the target object and the smartphone to perform tracking. FingerIO [23] and CovertBand [26] let the speaker sends OFDM symbols and uses the cross-correlation profile to extract the object motion. Strata [27] uses the 26-b GSM signal to estimate the CIR and derives the target motion by analyzing the phase change of each channel tap. EchoTrack [24] uses the FMCW signal as the transmitted signal. EchoTrack passes the received signal through a matched filter to localize the target object. Wang *et al.* [16] used the FMCW signal as well. The authors developed a technique called C-FMCW to detect the user's respiration when sleeping. Unlike traditional FMCW-based approaches, C-FMCW works by estimating the round-trip propagation time of acoustic signals. AudioGest [22] is a gesture recognition system that makes use of the Doppler effect of the reflected acoustic signal to estimate the in-air hand movement. ApneaApp [15] lets the smartphone send FMCW signals and detects the frequency shift at the received signal to extract the chest motion when breathing at sleep. Millisonic [21] is a headset tracking system that achieves mm-level accuracy. They use the phase shift of FMCW signals to perform tracking. BreathListener [17] is a respiration monitoring system in the driving context, which achieves fine-grained breath pattern monitoring using the generative adversarial network (GAN). Similar to our work, BreathJunior [33] uses audible signals to detect the respiration

of infants. They let the speaker send pseudorandomly generated Gaussian white noise and transform the received signal reflections into FMCW chirps which preserve the multipath reflection information. The respiration pattern is derived from transformed chirps. One major difference between our work and the above works is that they use predesigned signals to perform sensing while our system does not need the transmitted signal to be predesigned.

VII. DISCUSSION

We consider our work as the first step to explore the feasibility of using random signal, such as music and broadcast to perform sensing. We design extensive experiments to validate our system. The experimental result shows that our proposed method can accurately estimate the respiration rate in some real-world scenarios using music and broadcast signals. However, the current implementation still has the following three limitations.

- 1) As is shown in Fig. 14, wearing thick clothes severely degrades the system performance. This is due to the fact that sound wave is a mechanical wave, thick clothes may reflect or absorb the acoustic signal before the signal reaches the user's chest. Thus, the displacement of the subject's chest is too small to be detected.
- 2) Although our system can successfully detect the user's breathing rate when the speaker plays an audio piece, there might be relatively long blank intervals in music and broadcast signals (the beginning and end of a song as shown in Fig. 8). Thus, to achieve a reliable respiration monitoring, our system needs to be integrated with other sensing methods to have the ability of incessantly monitoring.
- 3) Our system extracts human respiration from the CIR leveraging the periodicity of human breathing. Thus, a relatively stable breathing rate is the prerequisite of our system and our system may fail to detect abnormal breathing patterns, such as apnea.

To resolve these three challenges, more designs are expected. Since this work is a first step toward the above blueprint, we leave the limitations and challenges to our future work.

VIII. CONCLUSION

In this article, we present a respiration monitoring system using music and broadcast signals. Different from the existing works, our system uses audible music and broadcast signal as the transmitted signal rather than the inaudible, predesigned signals which have negative effects on children, pets, and plants. The system detects the breathing pattern by estimating and analyzing the CIR. We evaluate the system on commodity speaker and microphone using several commonly heard audio pieces. Our experimental result shows that our system can work with high accuracy under several real-world circumstances.

REFERENCES

- [1] M. Ueda, K. Ashihara, and H. Takahashi, "How high-frequency do children hear?" in *Proc. Internoise 45th Int. Congr. Exposit. Noise Control Eng.*, 2015, pp. 21–24.
- [2] R. R. Fay and L. A. Wilber, *Hearing in Vertebrates: A Psychophysics Databook*, 1989.

- [3] J. A. T. da Silva and J. Dobránszki, "Sonication and ultrasound: Impact on plant growth and development," *Plant Cell Tissue Organ Cult.*, vol. 117, no. 2, pp. 131–143, 2014.
- [4] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 837–846.
- [5] J. N. Wilkinson and V. U. Thanawala, "Thoracic impedance monitoring of respiratory rate during sedation—Is it safe?" *Anaesthesia*, vol. 64, no. 4, pp. 455–456, 2009.
- [6] M. B. Jaffe, "Infrared measurement of carbon dioxide in the human breath: 'breathe-through' devices from Tyndall to the present day," *Anesth. Analg.*, vol. 107, no. 3, pp. 890–904, Sep. 2008.
- [7] M. Elliott and A. Coventry, "Critical care: The eight vital signs of patient monitoring," *Brit. J. Nurs.*, vol. 21, no. 10, pp. 621–625, 2012.
- [8] F. Benetazzo, A. Freddi, A. Monteriù, and S. Longhi, "Respiratory rate detection algorithm based on RGB-D camera: Theoretical background and experimental results," *Healthcare Technol. Lett.*, vol. 1, no. 3, pp. 81–86, Sep. 2014.
- [9] Y. Nam, Y. Kong, B. Reyes, N. Reljin, and K. H. Chon, "Monitoring of heart and breathing rates using dual cameras on a smartphone," *PLoS ONE*, vol. 11, pp. 1–15, Mar. 2016.
- [10] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 95–108.
- [11] R. Ravichandran, E. Saba, K. Chen, M. Goel, S. Gupta, and S. N. Patel, "WiBreathe: Estimating respiration rate using wireless signals in natural settings in the home," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, St. Louis, MO, USA, Mar. 2015, pp. 131–139.
- [12] H. Wang *et al.*, "Human respiration detection with commodity wifi devices: Do user location and body orientation matter?" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 25–36.
- [13] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-Sleep: Contactless sleep monitoring via WiFi signals," in *Proc. IEEE Real Time Syst. Symp.*, Rome, Italy, Dec. 2014, pp. 346–355.
- [14] *Amazon. Echo (3rd Gen)—Smart Speaker with Alexa*. Accessed: Aug. 5, 2020. [Online]. Available: <https://www.amazon.com/all-new-Echo/dp/B07R1CXKN7>
- [15] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. 13th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2015, pp. 45–57.
- [16] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-FMCW based contactless respiration detection using acoustic signal," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–20, Jan. 2018.
- [17] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li, "BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals," in *Proc. 17th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2019, pp. 54–66.
- [18] S. E. Trehub, B. A. Schneider, B. A. Morrongiello, and L. A. Thorpe, "Developmental changes in high-frequency sensitivity: Original papers," *Audiology*, vol. 28, no. 5, pp. 241–249, 1989.
- [19] D. Warfield, "The study of hearing in animals," in *Methods of Animal Experimentation, IV*. London, U.K.: Academic, 2012, pp. 43–143.
- [20] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 82–94.
- [21] A. Wang and S. Gollakota, "Millisonic: Pushing the limits of acoustic motion tracking," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, p. 18.
- [22] W. Ruan, Q. Z. Sheng, P. Xu, L. Yang, T. Gu, and L. Shangguan, "Making sense of doppler effect for multi-modal hand motion detection," *IEEE Trans. Mobile Comput.*, vol. 17, no. 9, pp. 2087–2100, Sep. 2018.
- [23] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "FingerIO: Using active sonar for fine-grained finger tracking," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2016, pp. 1515–1525.
- [24] H. Chen, F. Li, and Y. Wang, "EchoTrack: Acoustic device-free hand tracking on smart phones," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Atlanta, GA, USA, May 2017, pp. 1–9.
- [25] O. O. Oyerinde and S. H. Mneney, "Review of channel estimation for wireless communication systems," *IETE Techn. Rev.*, vol. 29, no. 4, pp. 282–298, 2012.
- [26] R. Nandakumar, A. Takakuwa, T. Kohno, and S. Gollakota, "CovertBand: Activity information leakage using music," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–24, Sep. 2017.
- [27] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, "Strata: Fine-grained acoustic-based device-free tracking," in *Proc. 15th Annu. Int. Conf. Mobile Syst. Appl. Serv.*, 2017, pp. 15–28.
- [28] H. Rahul, S. S. Kumar, and D. Katabi, "MegaMIMO: Scaling wireless capacity with user demand," in *Proc. ACM SIGCOMM*, vol. 4, 2012, p. 1.
- [29] W. Mao, J. He, and L. Qiu, "Cat: High-precision acoustic motion tracking," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, 2016, pp. 69–81.
- [30] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2009.
- [31] M. Pukkila, *Channel Estimation Modeling*, Nokia Research Center, 2000.
- [32] (2018). *Schuhfried—Biofeedback*. Accessed: Apr. 1, 2018. [Online]. Available: <https://www.schuhfried.com/biofeedback/>
- [33] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–16.
- [34] H. Abdelnasser, K. A. Harras, and M. Youssef, "Ubibreathe: A ubiquitous non-invasive WiFi-based breathing estimator," in *Proc. 16th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2015, pp. 277–286.
- [35] C.-Y. Hsu, A. Ahuja, S. Yue, R. Hristov, Z. Kabelac, and D. Katabi, "Zero-effort in-home sleep and insomnia monitoring using radio signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–18, Sep. 2017.



Wentao Xie received the B.E. degree from the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, in 2019. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, and the Department of Computer Science and Engineering, Southern University of Science and Technology.

His research interests include Internet of Things, wireless sensing, and mobile healthcare.



Runxin Tian received the B.E. degree from the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, in 2020. He is currently pursuing the master's degree with the School of Computing, National University of Singapore, Singapore.

His main research interests include distributed streaming computing, systems, and networking.



Jin Zhang (Member, IEEE) received the B.E. and M.E. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2004 and 2006, respectively, and the Ph.D. degree from the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong.

She is currently an Assistant Professor with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. Her research interests are mainly

in next-generation wireless networks, network economics, mobile computing in healthcare, cooperative communication, and networks.



Qian Zhang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from Wuhan University, Wuhan, China, in 1994, 1996, and 1999, respectively.

In September 2005, she joined Hong Kong University of Science and Technology (HKUST), Hong Kong, where she is currently a Tencent Professor of engineering and the Chair Professor of the Department of Computer Science and Engineering. She is also serving as the Co-Director of Huawei-HKUST innovation lab and the Director of Digital Life Research Center, HKUST. Before that, she was with Microsoft Research Asia, Beijing, China, from July 1999, where she was the Research Manager of the Wireless and Networking Group. She has published more than 400 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is the inventor of more than 50 granted and 20 pending International patents. Her current research interests include Internet of Things, smart health, mobile computing and sensing, wireless networking, as well as cyber security.

Dr. Zhang is currently serving as the Editor-in-Chief for the IEEE TRANSACTIONS ON MOBILE COMPUTING. She was a Members-at-Large of the IEEE Communications Society from 2016 to 2018.