Enabling Contact-free Acoustic Sensing under Device Motion

JIALIN LIU*, School of Software, Dalian University of Technology, Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China and University of Massachusetts Amherst, USA
DONG LI*, University of Massachusetts Amherst, USA
LEI WANG†, School of Software, Dalian University of Technology, Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, China
FUSANG ZHANG, Institute of Software, Chinese Academy of Sciences, China
JIE XIONG†, University of Massachusetts Amherst, USA

Recent years have witnessed increasing attention from both academia and industry on contact-free acoustic sensing. Due to the pervasiveness of audio devices and fine granularity of acoustic sensing, it has been applied in numerous fields, including human-computer interaction and contact-free health sensing. Though promising, the limited working range hinders the wide adoption of acoustic sensing in real life. To break the sensing range limit, we propose to deploy the acoustic device on a moving platform (i.e., a robot) to support applications that require larger coverage and continuous sensing. In this paper, we propose SonicBot, a system that enables contact-free acoustic sensing under device motion. We propose a sequence of signal processing schemes to eliminate the impact of device motion and then obtain clean target movement information that is previously overwhelmed by device movement. We implement SonicBot using commercial audio devices and conduct extensive experiments to evaluate the performance of the proposed system. Experiment results show that our system can achieve a median error of 1.11 *cm* and 1.31 *mm* for coarse-grained and fine-grained tracking, respectively. To showcase the applicability of our proposed system in real-world settings, we perform two field studies, including coarse-grained gesture sensing and fine-grained respiration monitoring when the acoustic device moves along with a robot.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: device motion, acoustic signals, contact-free sensing, large-scale

ACM Reference Format:
Jialin Liu, Dong Li, Lei Wang, Fusang Zhang, and Jie Xiong. 2022. Enabling Contact-free Acoustic Sensing under Device Motion. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 128 (September 2022), 27 pages. https://doi.org/10.1145/3550329

*Both authors contributed equally to the paper.
†Lei Wang (email: lei.wang@ieee.org) and Jie Xiong (email: jxiong@cs.umass.edu) are the corresponding authors.

Authors' addresses: Jialin Liu, School of Software, Dalian University of Technology, Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian, China and University of Massachusetts Amherst, Massachusetts, USA, jialinliu@umass.edu; Dong Li, University of Massachusetts Amherst, Massachusetts, USA, dli@cs.umass.edu; Lei Wang, School of Software, Dalian University of Technology, Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian, China, lei.wang@ieee.org; Fusang Zhang, Institute of Software, Chinese Academy of Sciences, Beijing, China, fusang@iscas.ac.cn; Jie Xiong, University of Massachusetts Amherst, Massachusetts, USA, jxiong@cs.umass.edu.

2474-9567/2022/9-ART128 $15.00
https://doi.org/10.1145/3550329

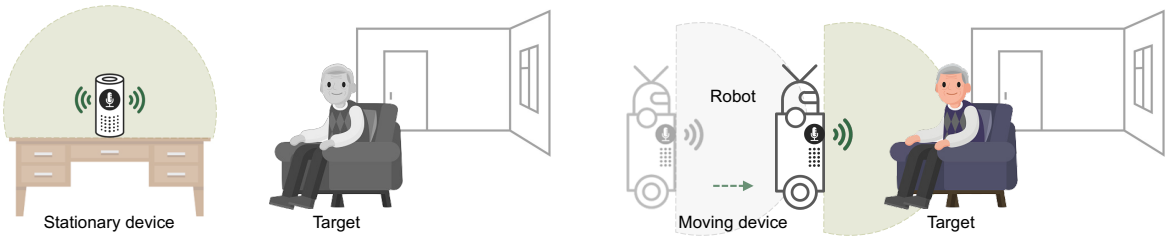
1 INTRODUCTION

Acoustic sensing that extends the primary use of speakers and microphones from audio playing and voice-based interaction to contact-free motion tracking has been gaining a tremendous amount of attention from both academia and industry recently. Acoustic sensing exhibits numerous advantages including pervasiveness [26, 29, 48] and fine-grained sensing granularity [57, 60]. On one hand, speakers and microphones are essential components on a variety of commodity devices in our daily life, including smartphones, smart speakers, smart TVs, etc. On the other hand, owing to the inherent low propagation speed in the air (i.e., 340 m/s), acoustic sensing can achieve a finer sensing granularity compared to other sensing modalities such as Wi-Fi sensing [30, 61, 65] and LoRa sensing [62, 63]. Recent efforts have successfully pushed the granularity of acoustic sensing from centimeter-level to millimeter-level [14, 54, 64]. As a result, acoustic sensing has enabled diverse applications, ranging from coarse-grained motion sensing such as human tracking [16] and gesture recognition [14, 25] to fine-grained activity sensing including finger tracking [28, 57, 60], lip reading [47, 66, 67], respiration monitoring [43, 55, 56], and eye blink detection [21].

Although promising, the limited range of acoustic sensing hinders its wide adoption in real life. For example, the state-of-the-art acoustic sensing systems can only achieve a sensing range of 2 m for respiration monitoring [55, 56] and 4.5 m for gesture recognition [25]. The main reason is that signals experience large attenuation when reflected from a target and become much weaker than the direct path signals. Researchers have endeavored to address this issue by employing microphone arrays [14, 55] and machine learning techniques [24, 25]. However, the sensing range is still limited to a few meters.

To break the sensing range limitation, we seek to deploy the acoustic sensing device on a mobile platform (e.g., a robot) to support applications that require larger coverage and continuous sensing, as shown in Fig. 1. For example, healthcare robots nowadays play an increasingly important role in providing well-being to both patients and caregivers [1, 3, 5, 23, 34, 36, 37, 41, 45, 50, 51]. The caregivers can remotely monitor the vital signs of patients in a contact-free manner, which reduces their risk of exposure to infectious diseases such as COVID-19. Healthcare robots can follow the elderly or patients to provide 24-hour continuous care and supervision. As essential components of voice-based human-computer interaction [3, 5, 34, 36, 37], the built-in speakers and microphones on healthcare robots can therefore be leveraged to provide fine-grained human sensing in a contact-free manner. Although most healthcare robots on the market [3, 36, 37] are equipped with other types of sensors such as camera and LiDAR, these sensing modalities have their limitations on human activity sensing. LiDAR sensors on robots are usually employed for detecting obstacles and identifying objects. While it can capture the shape of the target, the range accuracy of LiDAR sensors is not high enough for fine-grained sensing. For example, the range error of LiDAR sensors is usually 1–2% of the distance [44]. It indicates that, if the target is 2 m from the robot, the range error would be 2 to 4 cm, which is too large for respiration sensing with a subtle chest displacement of 0.5 cm. Cameras are another common components of robots for detecting surrounding objects. However, due to privacy concerns, people are usually unwilling to be continuously monitored by cameras. Compared with these sensing modalities, acoustic signals can achieve fine-grained sensing on the scale of sub-millimeter and do not raise privacy concerns. Therefore, we believe acoustic sensing well complements existing sensing modalities. For example, a robot can first identify the human target with the help of LiDAR, and then initiate respiration monitoring using acoustic sensing when it approaches the target.

The fundamental principle of acoustic sensing is that the target movement affects acoustic signals reflected from the target, so we can obtain target movement information by analyzing signal variations. Yet, current acoustic sensing is based on a critical assumption that the sensing device remains stationary [9, 19, 38, 43, 55, 56]. On the contrary, acoustic sensing under device motion (e.g., the acoustic device moves with a robot) is entirely different from sensing with a stationary device. The reason is that when the acoustic device is stationary, the signal variations are purely caused by target movements. However, the signal variations contain both target and



(a) The stationary scenario.

(b) The device motion scenario.

Fig. 1. Prior studies utilize stationary devices for contact-free acoustic sensing, resulting in a limited sensing range (green area). Our system aims at breaking the limits of acoustic sensing range by employing a moving robot to achieve a larger coverage and continuous sensing.

device movement information when the device is moving. Hence, we face critical challenges to make acoustic sensing work under device motion.

The first challenge is how to extract clean target movement information when the signal variations are caused by both target and device movements. Device movements can easily overwhelm target movements. One straightforward way to address device motion is to introduce motion sensors such as inertial measurement units (IMU) to capture the information of device motion. However, IMU-based solutions suffer from severe accumulated errors and bring additional data fusion burden [7]. Therefore, we aim to seek a solution without requiring any additional sensors.

To address this challenge, we leverage a static object in the environment as a reference to cancel out the effect of device motion. The key intuition is that the signals reflected from a reference[1] only contain the device movement information. In contrast, the signals reflected from the target contain both device and target movement information. We can then eliminate the effect of device motion by subtracting the signal variation in the reference reflection (containing only device movement) from that in the target reflection (containing both device and target movements) to extract clean target movement information.

In this paper, we employ the chirp signal design to separate signal reflections from the target and reference. The chirp signal design has a fine resolution (e.g., 4.25 cm for a signal bandwidth of 4 kHz) in resolving reflectors located at different distances. Reflectors separated by more than 4.25 cm will be placed into different range bins. In this case, we can obtain the signal reflection from the target and from the reference separately. However, we encounter the second challenge during the process of tracking the target and reference, i.e., the reflections become unstable when the device is moving. The desired target and reference reflections sometimes become weak, while the undesired multipath from other uninterested objects become strong. This phenomenon will cause the range information of the target and reference wrongly estimated.

To address the second challenge, we leverage the spatial continuity of movement to mitigate the effect of unstable reflections. Even though the noise and undesired multipath sometimes can be strong to interfere with the target and reference tracking, we observe that they appear randomly and can be eliminated as outliers. In contrast, the target and reference reflection can form two continuous trajectories owing to spatial continuity.

The third challenge we encounter is the insufficient phase sampling rate[2] under device motion. Specifically, we can obtain one phase sample from each single chirp. Since the phase measurement is wrapped within the interval $[-\pi, \pi]$, we need to unwrap the phase measurement to recover the true distance change between the

[1]We term the static object as the "reference" hereafter.

[2]The phase sampling rate mentioned here is different from the 48 kHz audio signal sampling rate in our system. The phase sampling rate is the number of transmitted chirps per second.

device and target. When the device moves slowly, the phase measurement can be correctly unwrapped. However, if the device moves fast, the values of two consecutive phase measurements will be more than π, making phase measurements wrongly unwrapped. Intuitively, we can increase the phase sampling rate by shortening the chirp duration to avoid this issue. However, a short chirp will severely decrease the sensing range, resulting in poor sensing performance.

To address the third challenge, we propose a new chirp transmitting scheme to deliberately make adjacent chirps overlap to increase the phase sampling rate. Unlike traditional chirp sending schemes in which one chirp will only be sent after the previous chirp is completely sent out, we start transmitting the second chirp before the transmission of the first chirp is completed. With this novel design, the phase sampling rate can be increased, breaking the device speed constraint on sensing under device motion.

In this paper, we propose a system named SonicBot to explore the capability of acoustic sensing under device motion. We implement SonicBot on a development platform for extensive experiments. We demonstrate the feasibility and reliability of the proposed system using two representative sensing applications, coarse-grained gesture recognition and fine-grained respiration monitoring. Our contributions are summarized as follows:

- To the best of our knowledge, this is the first time that device motion is considered in acoustic sensing. We believe sensing under device motion is a critical step towards real-world adoption of acoustic sensing.
- We propose a device motion cancellation scheme to extract the target movement in the presence of device motion. We propose a chirp transmitting scheme that can increase the phase sampling rate to address the phase unwrapping issue caused by device motion.
- We implement our system on a pair of commodity speaker and microphone, and conduct comprehensive experiments to evaluate its performance. We showcase our system with two representative applications. Experiment results demonstrate the effectiveness of the proposed solution in real-world settings.

2 RELATED WORK

As an integration of acoustic sensing with healthcare robots, SonicBot is closely related to contact-free acoustic sensing and sensing on healthcare robots. In this section, we discuss the literature most related to our study.

2.1 Acoustic Sensing

In the last decade, acoustic sensing has enabled plentiful sensing applications, which can be divided into two categories in terms of sensing granularity: coarse-grained and fine-grained activity sensing.

2.1.1 Coarse-grained Activity Sensing. Coarse-grained activity sensing aims at sensing large movements such as meter-level human walking [16, 17] and decimeter-level hand gesture [19, 25, 38, 58]. Acoustic sensing systems can capture target movement information by continuously tracking the relative distance between target and sensing device. For example, EchoSpot [16] achieves human tracking using a single pair of speaker and microphone in an indoor environment. Due to the large reflection area of human body, the reflected signals are strong enough to be captured within a room. However, room-scale hand tracking is a non-trivial task since hand is much smaller than human body, resulting in a much weaker reflection [14, 58]. Therefore, RTrack [25] combines signal processing techniques with recurrent neural network (RNN) to push hand tracking range to 4.5 *m*.

2.1.2 Fine-grained Activity Sensing. Different from sensing large movements, fine-grained activity sensing focuses on sensing subtle displacement such as respiration [43, 55, 56] and movement of small body parts such as finger tapping [28, 46, 57, 60] and eye blink [21]. Even though subtle movement cannot cause apparent distance change as that caused by large movement, it can still induce signal variations that delicate signal processing schemes can resolve. For example, a 5 *mm* chest displacement during breathing induces an obvious signal phase variation, which is significant enough to be detected [43, 55]. However, existing acoustic sensing systems can only achieve

fine-grained activity sensing within a limited range. For example, C-FMCW [56] achieves a maximum respiration sensing range of 2 m. Compared with coarse-grained activity, the sensing range of fine-grained activity is much shorter since the signal variation induced by subtle displacement could be easily buried in noise.

To push the range boundary of acoustic sensing, we propose to deploy the sensing device on a moving robot to enable applications that require large coverage and continuous sensing. One typical application is the healthcare robot, which continuously senses the target's physical status.

2.2 Sensing on Healthcare Robots

As the aging problem becomes more severe worldwide, there is an increasing demand for healthcare robots to help relieve the heavy burden of elderly care. Besides, the current COVID-19 pandemic has also stimulated a great need for contact-free nursing. We envision contact-free sensing on healthcare robots to become a general trend. Here, we introduce two functions of healthcare robots, i.e., physical rehabilitation and respiration monitoring, which are closely related to the coarse-grained and fine-grained activity sensing mentioned above, respectively.

2.2.1 Physical Rehabilitation. Many healthcare robots provide engaging physical rehabilitation of upper-limb function for patients with Parkinson's disease, Alzheimer's disease, or stroke [1, 40, 41, 49]. They aim at improving the patients' upper-limb performance by asking them to perform gestures to interact with the healthcare robot [4]. For example, iRebot [41] is a gesture-controlled robot that can provide physical rehabilitation with entertainment. However, it requires patients to wear customized sensors on their wrists, causing an extra burden on patients. To avoid adding wearable sensors on patients, Rijanto *et al.* [35] employ computer vision for rehabilitation on a robot, but camera-based methods cannot work well in poor lighting conditions and also raise privacy concerns.

2.2.2 Respiration Monitoring. Continuous vital sign sensing such as respiration monitoring is a critical function on healthcare robots for patients with respiratory diseases and the elderly who need 24/7 surveillance, especially under the COVID-19 pandemic [12, 27]. Most healthcare robots still employ traditional contact-based sensors such as oximeter to measure the respiration rate [12, 13], which is cumbersome and will result in a high risk of cross-infection. Dedicated sensors such as remote photoplethysmography (rPPG) [42, 52] and infrared thermometer [27, 32] are employed for contact-free respiration monitoring. Yet, these methods are sensitive to motion and cannot achieve continuous sensing under device motion.

3 PRELIMINARY

3.1 Chirp Signal Primer

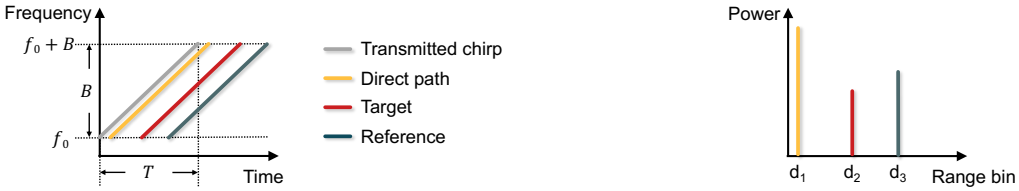
Chirp signal is a continuous wave whose frequency increases linearly with time as $f_0 + \frac{B}{T}t$, where f_0, B, and T are the initial frequency, bandwidth, and chirp duration, respectively. The transmitted chirp can be represented as

$$x(t) = \cos\left(2\pi\left(f_0 t + \frac{B}{2T}t^2\right)\right). \tag{1}$$

Considering a simple scenario where there only exists one reflector in the environment. The transmitted chirp is reflected from the reflector, and then received at the microphone. As shown in Fig. 2a, the received chirp is a delayed version of the transmitted chirp, which can be represented as

$$y(t) = \alpha\cos\left(2\pi\left(f_0(t-\tau) + \frac{B}{2T}(t-\tau)^2\right)\right), \tag{2}$$

where α is the signal attenuation, and τ is the time-of-flight (ToF) of the signal in the air. τ can be used to calculate the distance of the reflector with respect to the speaker-microphone pair as $d = \frac{\tau V_s}{2}$, where V_s is the speed of the sound. To obtain the distance information of the reflector, we obtain the mixed signal by multiplying the received chirp with the transmitted chirp. After applying the product-to-sum formula $\cos A \cos B =$



(a) The transmitted chirp and the received chirps.

(b) The FFT result of the mixed signal.

Fig. 2. The transmitted chirp (a) travels through multiple paths with different delays and (b) falls into different range bins.

$(\cos(A-B)+\cos(A+B))/2$ and filtering out the high frequency component $\cos(A+B)$, the mixed signal can be represented as

$$m(t) = \frac{1}{2}\alpha \cos\left(2\pi\left(\frac{B}{T}\tau t + f_0\tau - \frac{B}{2T}\tau^2\right)\right). \tag{3}$$

We further simplify the mixed signal as

$$m(t) = \frac{1}{2}\alpha \cos(2\pi f t + \varphi), \tag{4}$$

where $f = \frac{B}{T}\tau$ is the beat frequency, and $\varphi = 2\pi\left(f_0\tau - \frac{B}{2T}\tau^2\right) \approx 2\pi f_0\tau$ is the initial phase. In a typical setting, $f_0 >> \frac{B}{2T}\tau$, so we can omit the quadratic term. Taking multipath into consideration, the mixed signal can be written as a superposition of reflections from N paths

$$m(t) = \sum_{i=1}^{N} \frac{1}{2}\alpha_i \cos(2\pi f_i t + \varphi_i), \tag{5}$$

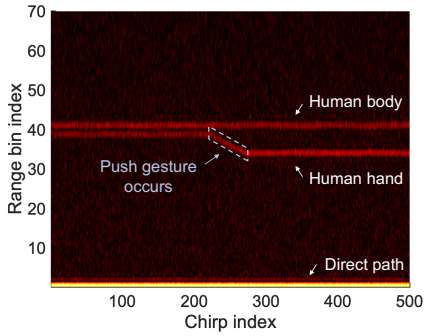
where α_i, f_i, and φ_i denote the signal attenuation, beat frequency, and initial phase for the i-th path, respectively. The beat frequency can present us a coarse-grained estimate of the absolute distance (i.e., the ranging information) of the reflector with respect to the device. Specifically, the signals reflected from reflectors at different distances have different ToFs, leading to different beat frequencies. Therefore, we can calculate the target's coarse-grained distance d based on the beat frequency as

$$d = \frac{V_s f T}{2B}. \tag{6}$$

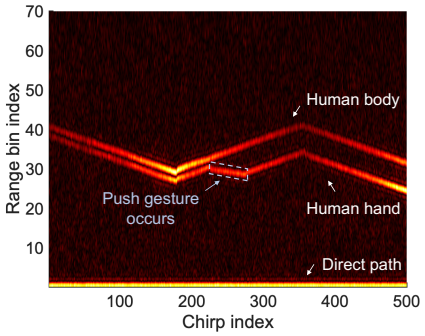
The beat frequency can be obtained by performing Fast Fourier Transform (FFT) on the mixed signals. Then, multiple signals reflected from reflectors at different distances can be separated into different range bins as shown in Fig. 2b. The size of the range bin is relevant to the range resolution, which is determined by the frequency bandwidth B of the chirp and can be calculated as $\frac{V_s}{2B}$. That means if two reflectors are located $\frac{V_s}{2B}$ apart with respect to the device, the two reflections will fall into two different range bins. Without loss of generality, we consider a bandwidth of 4 kHz, which yields a range resolution of 4.25 cm. If the moving distance of the reflector exceeds the size of the range bin, the reflector moves from one range bin to another and thus we can calculate the distance change. In contrast, if the moving distance of the reflector is less than one range bin, we need to refer to the phase change to track its movement.

The initial phase contains fine-grained information to track the displacement Δd of the reflector with respect to the device. The displacement Δd from distance d_0 to d_1 can be represented as

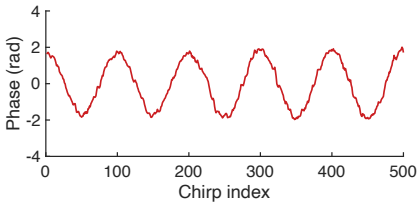
$$\Delta d = d_1 - d_0 = \frac{\tau_1 V_s}{2} - \frac{\tau_0 V_s}{2} = \frac{V_s}{2}(\tau_1 - \tau_0). \tag{7}$$



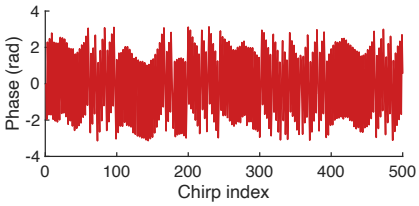
(a) The range profile under the stationary scenario.



(b) The range profile under the device motion scenario.



(c) The phase change under the stationary scenario.



(d) The phase change under the device motion scenario.

Fig. 3. In the stationary scenario, (a) hand gesture can be recognized based on the range information of the hand, and (c) respiration pattern can be obtained from the phase information extracted from the the range bin of the human body. While in the device motion scenario, (b) the range bin of the human hand keeps changing, and (d) the phase variation induced by respiration is overwhelmed by device motion.

Since the initial phase mentioned above is represented as $\varphi \approx 2\pi f_0 \tau$, we can obtain the relationship between the displacement Δd and the phase variation $\Delta\varphi$ as

$$\Delta d = \frac{V_s}{2}\left(\frac{\varphi_1}{2\pi f_0} - \frac{\varphi_0}{2\pi f_0}\right) = \frac{V_s \Delta\varphi}{4\pi f_0}. \tag{8}$$

To summarize, the range information can be adopted for coarse-grained motion sensing. In contrast, the phase information can imply the subtle displacement that happens within a range bin, which can be employed for fine-grained activity sensing.

3.2 Sensing with Stationary Device

In this subsection, we introduce how contact-free acoustic sensing systems work with stationary devices.

3.2.1 Coarse-grained Activity. Since the coarse-grained activity such as hand gesture usually involves a movement scale larger than the size of a range bin, it can be sensed using the absolute distance information. Specifically, the target reflection can be observed on the range profile with high energy as shown in Fig. 3a. Suppose that a human target sitting in front of the device performs a "push" gesture. We can observe that the hand gets closer from the range profile shown in Fig. 3a. Therefore, we can achieve gesture recognition based on the range information.

3.2.2 Fine-grained Activity. Despite the range information can imply the target movement, it is not enough for fine-grained activity sensing since the displacement of fine-grained activity is smaller than one range bin. In this case, we need to employ the phase information for subtle movement sensing. We first detect the range

bin in which the target is located and then zoom into it to extract the phase variation. We employ respiration sensing as an example to illustrate fine-grained activity sensing. When sensing with a stationary device, since the distance between the device and the target is relatively fixed and the 5 *mm* chest movement is within a range bin of 4.25 *cm*, the range bin of the human body will not change during the sensing process, which can be observed in Fig. 3a. Therefore, we can take a one-time effort to identify the target reflection and then zoom into the target range bin to obtain the phase change caused by chest movement. As Eq. 8 depicts, the fine-grained movement causes the initial phase of the mixed signals to vary. The 5 *mm* chest movement causes the reflection path length change and induces around a phase variation of 3.69 *rad*. By measuring the phase variation, we can obtain the respiration pattern of the target as shown in Fig. 3c.

3.3 Sensing under Device Motion

Acoustic sensing under device motion is dramatically different from sensing with a stationary device. We introduce the impact of device motion on coarse-grained and fine-grained activity sensing, respectively.

3.3.1 Coarse-grained Activity. For coarse-grained gesture recognition, we can see from Fig. 3b that the push gesture is distorted by device motion. The reason is that besides the target movement, device motion is also contained in target reflection. Thus, the push gesture cannot be accurately recognized under device motion.

3.3.2 Fine-grained Activity. In the device motion scenario, the respiration pattern can no longer be extracted by the approach used in the stationary scenario. Specifically, the device motion has the following two main effects on respiration sensing:

(1) The range bin of the target changes. Under device motion, the distance between the target and the device varies significantly, which is much larger than the size of one range bin (i.e., 4.25 *cm*), thus the range bin where the target is located continuously changes from one to another as shown in Fig. 3b. Consequently, the chest movement occurs in different range bins as the device moves, which requires us to continuously detect the target's range bin.
(2) The target movement is overwhelmed by device motion. Even though we can accurately identify the target's range bin, the phase variation extracted from the target reflection contains the chest movement and the device motion information. Both chest movement and device motion contribute to the reflection path length change, causing phase variations to be superimposed. We show the phase variation extracted from the target's range bin in Fig. 3d. We can see that the phase variation no longer indicates the respiration pattern under device motion.

In summary, device motion degrades sensing performance and even causes failure of existing sensing systems. We therefore propose SonicBot to address the issue of device motion in acoustic sensing.

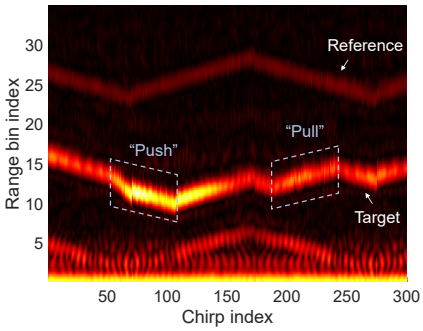
4 DEVICE MOTION CANCELLATION

Our basic idea is to employ a static object in the environment as a reference to extract the signal variation caused by the device motion, which can be removed to obtain clean target movement information. This section introduces the principle of the proposed solution and conducts experiments to evaluate its feasibility and effectiveness.

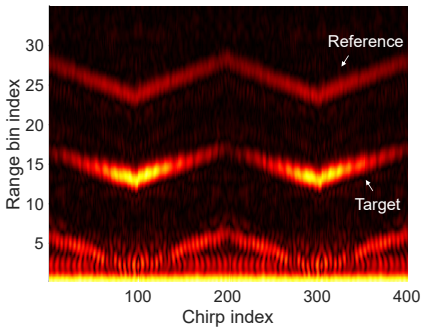
4.1 Principle

Suppose there exist two reflectors in the environment, which are the target and the reference. We denote the target reflection and reference reflection as m_{Tar} and m_{Ref}, respectively

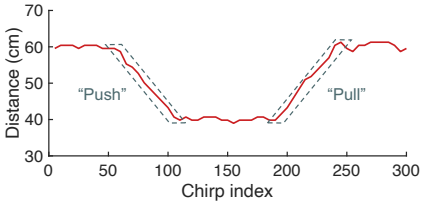
$$m_{Tar}(t) = \frac{1}{2}\alpha_{Tar}\cos(2\pi f_{Tar}t + \varphi_{Tar}), \tag{9}$$



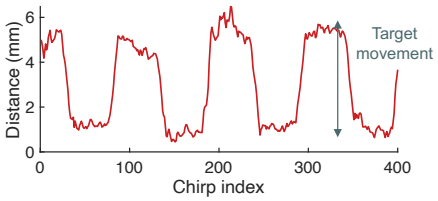
(a) The range profile of sensing coarse-grained movement.



(b) The range profile of sensing fine-grained movement.



(c) The recovered target movement using range information.



(d) The recovered target movement using phase information.

Fig. 4. An illustration of applying the proposed motion cancellation scheme for acoustic sensing under device motion.

$$m_{Ref}(t) = \frac{1}{2}\alpha_{Ref}\cos(2\pi f_{Ref}t + \varphi_{Ref}), \tag{10}$$

where f_{Tar}, f_{Ref} contain the range information, and φ_{Tar}, φ_{Ref} are the phase information of the target and the reference, respectively.

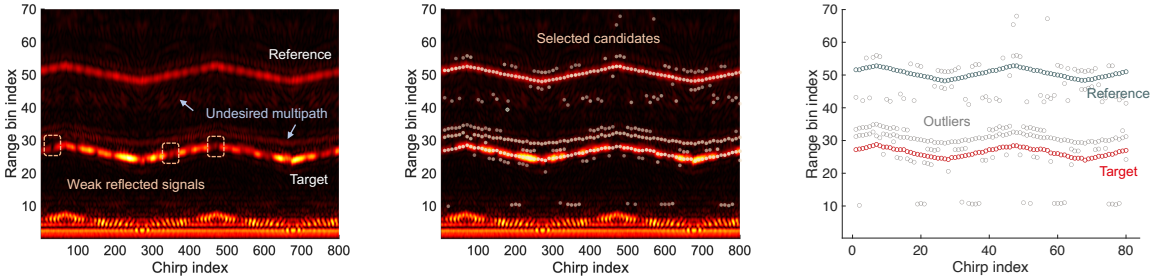
Since the reference is static, the variation of range and phase information obtained from the reference reflection is purely induced by device motion. Therefore, we can infer the device motion information from reference reflection. On the other hand, the target reflection contains both target movement information and device motion information. So we can recover clean target-induced signal variation by subtracting the signal variation in the reference reflection from that in the target reflection. Based on Eq. 6, the coarse-grained target moving distance Δd_{Tar}^{Coarse} can be recovered from the range information as

$$\Delta d_{Tar}^{Coarse} = \frac{V_s T}{B}(\Delta f_{Tar} - \Delta f_{Ref}). \tag{11}$$

If the target movement is within a range bin, we need to further involve the phase information for fine-grained displacement measurement Δd_{Tar}^{Fine}, which can be calculated based on Eq. 8 as

$$\Delta d_{Tar}^{Fine} = \frac{4\pi f_0}{V_s}(\Delta\varphi_{Tar} - \Delta\varphi_{Ref}). \tag{12}$$

Through the above subtraction operation, the impact of device motion can be eliminated and clean target movement information can be recovered.



(a) Undesired multipath and weak reflected signals will result in wrongly-estimated range information.

(b) We select five peaks with high reflected energy as candidates in each time window.

(c) The target and the reference can be successfully detected by our proposed solution.

Fig. 5. Illustration for (a) wrongly-estimated range information, (b) selected candidate peaks, and (c) the clustering result of target and reference detection.

4.2 Feasibility Evaluation

We conduct experiments to evaluate the effectiveness of the device motion cancellation scheme. The device is placed on a linear guide slide and moves back and forth for 20 cm at a constant speed of 5 cm/s. We use a cardboard with a size of 8×15 cm^2 as target and attach it to a cozmo robot to control its movement, and the target is placed 0.6 m away from the device. We adopt a 19×19 cm^2 cardboard box as the reference and put it 1 m away from device. To evaluate coarse-grained activity sensing, we let the cozmo robot move forward and backward for 20 cm to simulate the "push" and "pull" gesture, respectively. To evaluate the fine-grained activity sensing, we let the cozmo robot move back and forth for 5 mm to simulate the chest movement.

Fig. 4 shows the effect of applying the device motion cancellation scheme on coarse-grained and fine-grained movement sensing. From Fig. 4a, we can see that the coarse-grained target movement is severely distorted by the device motion. However, it can be accurately recovered by subtracting the range change of the reference reflection from that of the target reflection as shown in Fig. 4c. For fine-grained activity sensing as shown in Fig. 4b, the fine-grained target movement cannot be recovered from the range information. We can recover the subtle movement pattern of the target by subtracting the phase change of the reference reflection from that of the target reflection as demonstrated in Fig. 4d.

In brief, utilizing a static object in the environment as a reference to cancel out device motion is an effective scheme for recovering both coarse-grained and fine-grained target movement information.

5 ESTIMATING RANGE INFORMATION OF THE TARGET AND THE REFERENCE

To apply the device motion cancellation scheme proposed in Sec. 4, accurate range information of the target and the reference is critical. This section introduces the problem encountered during the process of estimating the range information and addresses the problem by applying spatial continuity to constrain target movements.

5.1 Problem Description

We employ range profiles to localize and track reflectors in the environment. The range profile shows a high-energy trajectory which implies the location of the reflector at different timestamps as shown in Fig. 5a. However, due to rich multipath in indoor environment, we can observe several other peaks on the range profile, which means that there exist other undesired multipath reflections besides the target and reference we are interested. To eliminate the undesired multipath interference including the direct path and reflections from other static

objects, prior studies [14, 25] adopt background subtraction to make the interested reflection stand out on the range profile. Although background subtraction works effectively under the stationary scenario, it does not work under device motion due to the following reasons.

First, in a multipath-rich environment, the multipath conditions at two locations might be dramatically different. During the process of device motion, the multipath condition keeps varying and some undesired multipath would appear randomly on the range profile. Fig. 5a shows a range profile under device motion. The device moves forward and backward for 20 cm. The human target sits in front of the device at 1 m and there is a wall behind the human target at 2 m with respect to the device. From the figure, we can observe that the undesired multipaths pinpointed by arrows vary over chirps and are strong enough to interfere with target range estimation. In stationary scenario, the multipath reflected from static objects are relatively stable over chirps, and thus, can be removed by a subtraction operation between range profiles of two adjacent chirps. However, this background subtraction method does not work well in the device motion scenario due to varying multipath. Second, device motion causes the signal reflected by the target or the reference at certain locations to be very weak. As we can see from Fig. 5a, several dark gaps marked in the dashed boxes make the target's trajectory discontinuous on the range profile. Both of these two reasons cause wrongly estimated target/reference range which further impacts device motion cancellation.

5.2 Spatial Continuity of Movement

We propose a method to eliminate the wrongly-estimated target/reference range information caused by device motion. The proposed method is based on the fact that the movement of an object is continuous in space. Thus, we can rule out the undesired multipath which randomly appears on the range profile and remedy those gaps where the reflections from the target or the reference are weak through the following steps.

First, we assume that the range bin of the target and the reference do not change within a short time window (e.g., 0.2 s). We can remove the random noise and partial undesired multipath by summing up the range profiles from multiple chirp measurements within this time window.

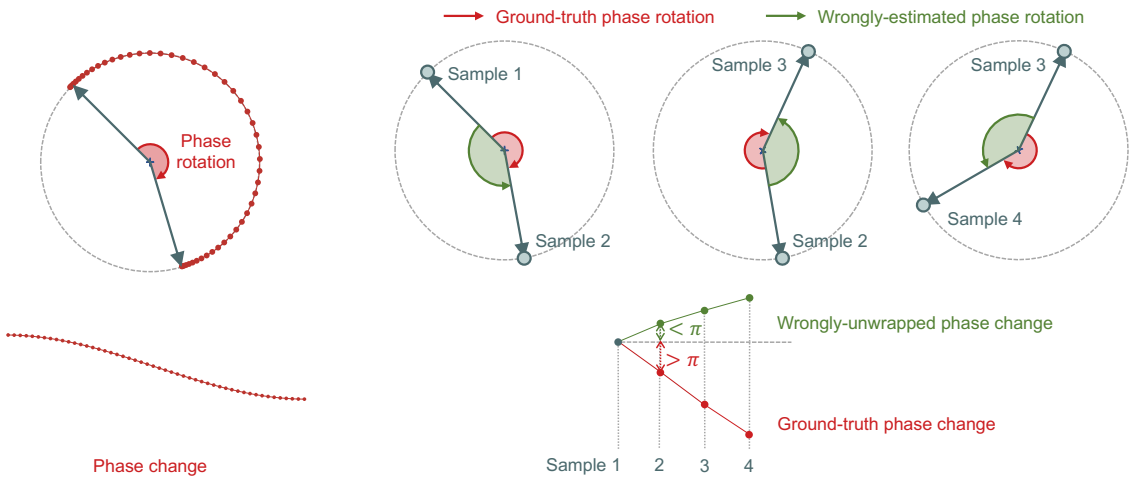
Second, we pick a certain number of peaks with high reflection signal strength as candidates from the summed range profile of each time window as shown in Fig. 5b since they indicate the possible locations of the target and the reference. We empirically set the number of candidate peaks as 5 [11]. There are two factors we take into consideration for selecting candidates. (i) The direct path has much stronger power than the reflection paths, so the candidate searching starts from the distance of 0.3 m to eliminate influence of the direct path. (ii) One reflector may cause two to three adjacent range bins to have high energy on the range profile, so we regard them as one candidate.

Third, there are still some undesired multipath reflections that have intense reflection energy. Therefore, we apply the DBSCAN algorithm to the selected candidates to rule out the residual undesired multipath. According to spatial continuity, the range points of the target and the reference are recognized as two classes while the random points caused by undesired multipath are regarded as outliers. Thus, we can obtain two clear trajectories corresponding to the target and the reference, respectively, as shown in Fig. 5c. Last but not least, if there are still some range information missing in the trajectories due to the weak reflections, we then remedy those missing range information by predicting them according to the anterior and posterior range information.

To sum up, the spatial continuity helps us eliminate interference from the undesired multipath reflections to achieve robust identification of the target and the reference.

6 IMPROVING THE PHASE SAMPLE RATE

So far, accurately estimating the range information of target and reference is enough for coarse-grained activity sensing such as gesture recognition, whose displacement is in the range of 10 to 30 cm. Yet, the range information



(a) The phase change under the stationary scenario.

(b) The large phase change between two consecutive samples causes wrong phase unwrapping under device motion.

Fig. 6. (a) The phase change induced by respiration between two samples is within π under the stationary scenario since the chest movement is only 5 *mm* and its speed is slow. (b) However, the phase change under the device motion scenario is induced by respiration plus device motion which will cause a significant phase change between two samples, resulting in the phase being wrongly-unwrapped.

is not enough to sense finer-grained movements such as respiration because a 5 *mm* displacement of the chest is much smaller than the size of a range bin, i.e., 4.25 *cm*. As a result, we need to further extract the phase information that contains fine-grained movement information. However, there exists an insufficient phase sampling rate problem in chirp signal design which severely limits the capability of fine-grained activity sensing under device motion. This section first introduces the insufficient phase sampling rate problem. Then, we propose an overlapped chirp transmitting scheme to mitigate the effect caused by this problem.

6.1 Problem Description

Phase information is widely adopted for respiration monitoring since it can provide fine-grained displacement measurements [43, 55]. We can obtain one phase sample from every single chirp. For example, if the chirp duration is 0.04 *s*, the acoustic sensing system can transmit 25 chirps per second, which means that we can obtain 25 phase samples within one second for sensing. When the device is stationary, the target reflection only contains the chest movement. According to Eq. 8, a 5 *mm* chest displacement causes a phase change of 3.69 *rad*. In this case, the phase changes on the I-Q plane as illustrated in Fig. 6a. The phase rotates clockwise when the chest moves towards the device and counter-clockwise when the chest moves away. By measuring the phase change, we can recover the respiration pattern.

In contrast, the phase change under device motion is induced by both target and device. The device motion is much larger than the chest movement and it causes the phase wrapping problem because the phase change between two consecutive chirps can be more than π, resulting in the displacement being wrongly estimated as shown in Fig. 6b.

Specifically, in order to avoid the phase wrapping problem, the phase change between two consecutive samples cannot be more than π, which corresponds to a displacement of 4.25 *mm*. Hence, to make respiration sensing

(a) The traditional chirp transmitting scheme.

(b) The overlapped chirp transmitting scheme.

Fig. 7. Different from the traditional chirp transmitting scheme (a), our overlapped chirp transmitting scheme (b) can increase the chirp rate without sacrificing sensing range.

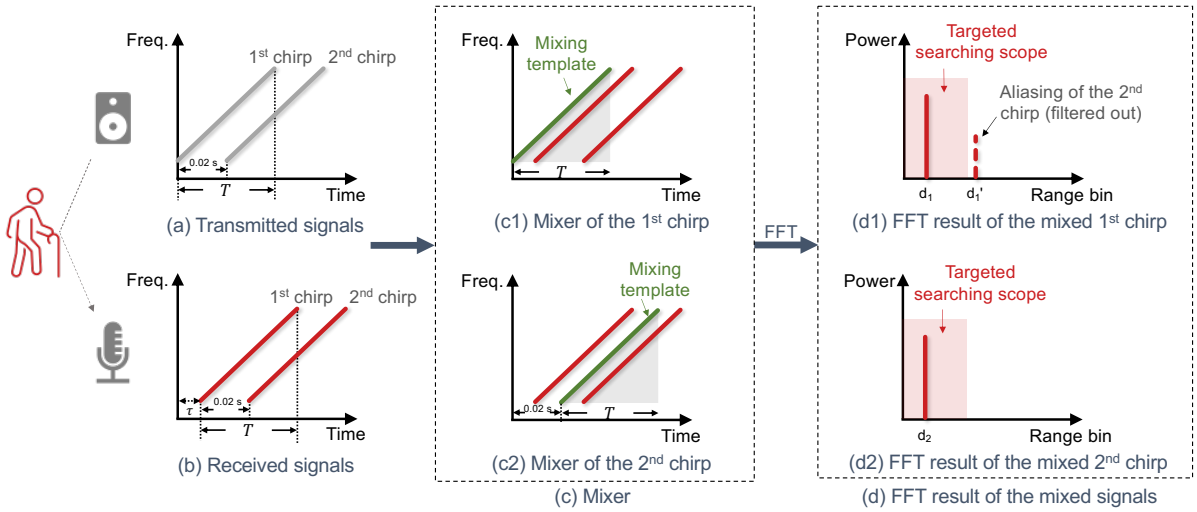


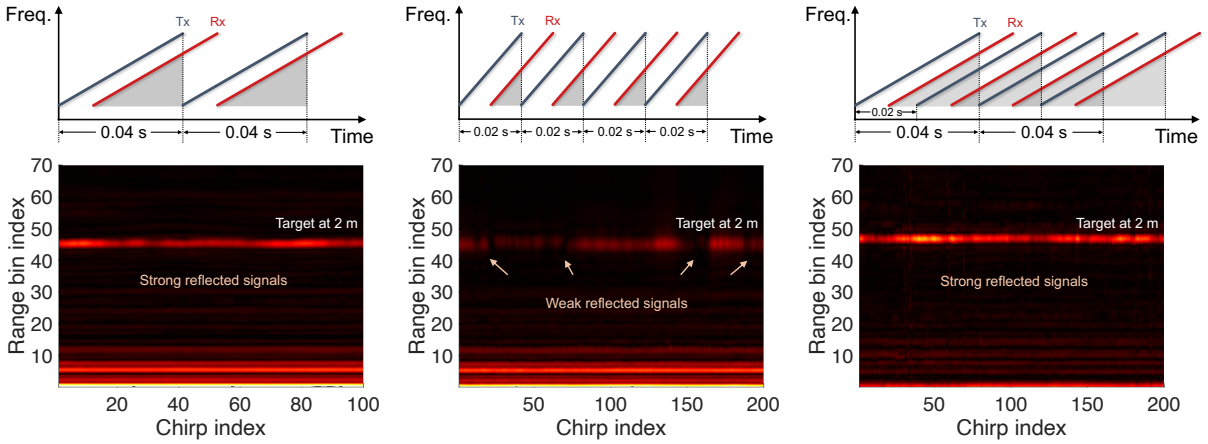
Fig. 8. The workflow of the overlapped chirp transmitting scheme.

work under a device moving speed of 0.2 m/s[3], the phase sampling rate should be at least $\frac{0.2 \times 1000\ mm/s}{4.25\ mm/s} = 47\ Hz$. Therefore, the chirp duration should not exceed 0.0213 s to meet the required phase sampling rate. The typical chirp duration setting for activity sensing is 0.04 s [14, 25], which cannot meet the phase sampling rate required under device motion. If we decrease the chirp duration to 0.0213 s, the sensing range would also be decreased. As a result, we cannot simply shorten the chirp duration to improve the phase sampling rate.

6.2 Overlapped Chirp Transmitting Scheme

We propose an overlapped chirp transmitting scheme that deliberately makes the adjacent chirps overlap, which can increase the phase sampling rate without decreasing the sensing range. Unlike traditional chirp sending scheme shown in Fig. 7a, in which one chirp will only be sent after the previous chirp is completely sent out, we start transmitting the second chirp before the transmission of the first chirp is completed as shown in Fig. 7b. Since chirps overlap with each other, we call it overlapped chirp transmitting scheme. Through transmitting overlapped chirps, we increase the number of transmitted chirps per second, improving the phase sampling rate.

[3]The safe speed of the healthcare robot for human-robot interaction is no more than 0.2 m/s within a distance of 2 m [2].



(a) Traditional scheme with 0.04 s chirps.

(b) Traditional scheme with 0.02 s chirps.

(c) Overlapped scheme with 0.04 s chirps.

Fig. 9. Range profiles when different chirp transmitting schemes are adopted.

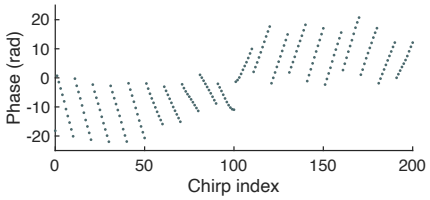
The workflow of overlapped chirp transmitting scheme is shown in Fig. 8. Specifically, the overlapped chirps are transmitted from the speaker, reflected by the target, and then received by the microphone. We mix the received overlapped chirps one by one with a single transmitted chirp (i.e., the mixing template as shown in Fig. 8c) to obtain the mixed signals. Note that the overlap would create an aliasing peak from the next chirp, which is 3.43 m further away with respect to the true peak for a gap of 0.02 s between overlapped chirps as shown in Fig. 8d. As the targeted sensing range in this work is 3 m, the aliasing peak can be filtered out since it is out of the sensing range. After getting the mixed signals of target reflection, we can further obtain the phase information for each chirp.

The phase sampling rate can be calculated by the time shift between the start timestamps of two consecutive chirps. For example, if we employ overlapped 0.04 s chirps with a time shift of 0.02 s, then the phase sampling rate is $\frac{1}{0.02\ s} = 50\ Hz$, which is doubled compared with $\frac{1}{0.04\ s} = 25\ Hz$ using traditional 0.04 s chirps. Although repeating a 0.02 s chirp twice can achieve the same phase sampling rate as overlapped 0.04 s chirps, the later achieves a better sensing performance. This is because the reflected signal strength is related to (i) the SNR of the received signals and (ii) the number of mixed samples [15, 25, 53], i.e., the size of the grey areas shown in Fig. 9. Another way to increase the SNR is to employ the cross correlation method [56]. However, compared with the proposed scheme, the sensing granularity of the cross correlation method is coarser, and its computational complexity is relatively higher.

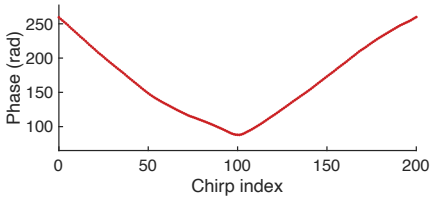
We conduct experiments[4] to show the performance difference for three schemes, i.e., (a) traditional scheme with 0.04 s chirps, (b) traditional scheme with 0.02 s chirps, and (c) overlapped 0.04 s chirps. We can see from Fig. 9 that the reflections of using 0.04 s chirps (i.e., Fig. 9a and Fig. 9c) are stronger than that of using 0.02 s (Fig. 9b). Scheme (c) achieves similar signal strength as scheme (a). On the other hand, scheme (c) obtains more phase samples within the same time window and therefore can track a target with a larger speed.

Although we can further increase the phase sampling rate by shortening the time shift, i.e., increasing the overlap between two sending chirps, there exists a boundary that limits the improvement of the phase sampling rate. The reason is two-fold. First, the upper bound of the phase sampling rate depends on the required

[4]We put a cozmo robot 2 m away from the sensing device. The robot moves back and forth for 5 mm and the sensing device is stationary.



(a) Before eliminating the initial phase values.



(b) After eliminating the initial phase values.

Fig. 10. (a) Different initial phase values from different range bins cause the phase change to be discontinuous. (b) According to the spatial continuity of movement, we can eliminate the initial phase difference and make the phase change continuous.

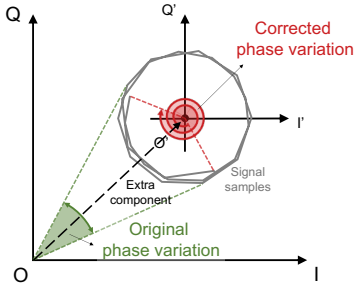


Fig. 11. Illustration of phase correction in multipath-rich environments.

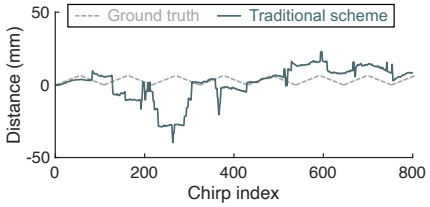
sensing range. If the targeted sensing range is 3 m, the time shift between adjacent chirps cannot be less than $\frac{3\ m \times 2}{340\ m/s} = 0.017\ s$. Otherwise, we cannot receive the reflected chirp before the next chirp is transmitted. Second, the total transmission power of the speaker is fixed. If we transmit too many chirps within a time window, the energy of each chirp will severely decrease, degrading the sensing range. As a result, there is a trade-off between phase sampling rate and sensing range. In our system, we employ a chirp duration of 0.04 s with a time shift of 0.02 s. Under this setting, we can achieve respiration sensing under a device moving speed of 0.2 m/s and still maintain a sensing range of 3 m.

6.3 Splice Phase from Different Range Bins

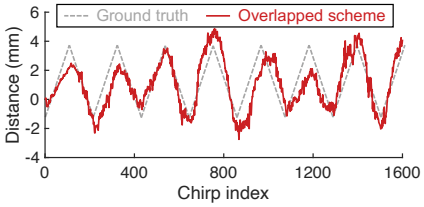
As mentioned in Sec. 4, we need the knowledge of the phase information from the range bins of the target and the reference for fine-grained movement sensing. Since the chest movement occurs in different range bins under device motion, the phase variation caused by chest movement is included in different rang bins. Therefore, we need to splice the phase measurements from multiple range bins.

However, the initial phase values are different for different range bins according to Eq. 4, thus the phase change between two different range bins is discontinuous as shown in Fig. 10a. We propose a simple yet effective method to eliminate the effect of different initial values by connecting their head and tail. The reason is that the target movement is continuous in space, so the phase variation induced by the target movement is continuous as well. Fig. 10b shows that the phase change extracted from different range bins are continuous after applying the proposed method.

In complex multipath environments, the composite signal contains not only the target reflection but also reflections from static objects in the environment [30, 53]. The extra component caused by the static objects will



(a) The traditional chirp transmitting scheme.



(b) Our overlapped chirp transmitting scheme.

Fig. 12. (a) Due to insufficient phase sampling rate, target movement cannot be accurately sensed using the traditional chirp transmitting scheme under device motion. (b) The overlapped chirp transmitted scheme is used to increase the phase sampling rate without sacrificing the sensing range.

move the composite signal samples away from the origin as shown in Fig. 11. In this case, if we calculate the phase variation with respect to the origin of the coordinate, we are not able to obtain the true phase variation induced by motions. To address this issue, we notice that within a short time interval (e.g., 0.1 s), the extra component caused by the static objects can be assumed as a constant. The phase variations are caused mainly by the device and target movements. We can therefore apply a circle fitting algorithm [33] to find the center of the circle which corresponds to the extra component induced by static objects. By taking the center of the circle as the new coordinate origin, the impact of reflections from static objects can be removed.

After increasing the phase sampling rate and splicing the phase change, we can recover the target movement by subtracting the phase change in the reference reflection from that in the target reflection under device motion. Fig. 12 shows the result of recovering a 5 mm target movement under the 0.2 m/s device moving speed. We can see that the phase wrapping problem occurs if we utilize the traditional chirp transmitting scheme, causing the target movement to be undetectable as shown in Fig. 12a. In contrast, our overlapped chirp transmitting scheme can provide a sufficient phase sampling rate to successfully recover the target movement as shown in Fig. 12b.

6.4 Address the Phase Inconsistency Issue

In practical scenarios, there may be a certain angle between the robot moving direction and the reference surface as illustrated in Fig. 13, which results in inconsistency between target reflection path length change and reference reflection path length change. Suppose the target and the reference are located at different directions with respect to the device moving direction. As shown in Fig. 13, we denote the direction difference between the device moving direction and device-target direction as α_t and that between the device moving direction and device-reference direction as α_r, respectively. Now if the device moves a distance of ΔL, the reflection paths from the target/reference will change from the dashed lines to the solid lines as marked in the figure. We denote the reflection path length change as ΔP_t and ΔP_r for the target and the reference, respectively, which can be approximated as $\Delta P_t = \Delta L \cos \alpha_t$ and $\Delta P_r = \Delta L \cos \alpha_r$. We can see that if $\alpha_t \neq \alpha_r$, the path length changes are not equal. Thus the phase variations caused by the device motion at the target and at the reference are different, which means the impact of device motion at the target cannot be fully eliminated by the device motion captured at the reference.

To address this problem, we propose an optimization-based phase compensation scheme to convert the phase variation at the reference to be the same as that at the target for cancellation. For a small time interval, we can assume the device moves at a constant speed. For each time interval, we denote the compensation weight ω as

$$\omega = \frac{\Delta\varphi_t}{\Delta\varphi_r} = \frac{\Delta P_t}{\Delta P_r} = \frac{\Delta L \cos \alpha_t}{\Delta L \cos \alpha_r} = \frac{\cos \alpha_t}{\cos \alpha_r}. \tag{13}$$

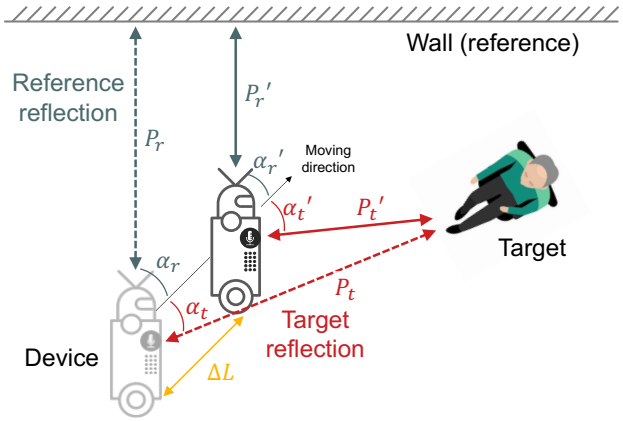


Fig. 13. The path length variations caused by target moving at the target and at the reference are different.

By applying the weight on the phase variation at the reference, we can address the phase inconsistency issue caused by different angles (directions), and the device motion can be eliminated.

However, without the knowledge of α_t and α_r, how to estimate the compensation weight is non-trivial. In this paper, we turn it into an optimization problem. Suppose there are N chirps within a time interval, we denote the phase measurements of the n-th chirp of the target reflection and the reference reflection as ϕ_{t_n} and ϕ_{r_n}, respectively. To find the best weight ω for each time interval, we optimize the following function

$$\arg\max_{\omega,C} \left| \sum_{n=1}^{N} e^{j\left((\omega\phi_{r_n}+C)-\phi_{t_n}\right)} \right|, \tag{14}$$

where C is a constant value. The function obtains the weight ω and the constant offset C that can minimize the difference between the phase of reference reflection and the phase of target reflection after compensation [59]. The size of the time interval determines the number of chirps N within a time interval. Note that a smaller interval brings more frequent weight estimation and therefore we try to set a larger interval. For applications with slower movements such as respiration, the time interval can be set to a larger value (e.g., 3 s). For fast movement such as hand tremor caused by Parkinson's disease, we can set a smaller time interval (e.g., 100 ms).

7 EVALUATION

In this section, we evaluate the performance of the proposed system through benchmark experiments and real-world applications. Benchmark experiments are employed to verify the effectiveness of the proposed schemes and study the impact of varying parameters and conditions. Furthermore, we showcase gesture recognition as coarse-grained sensing and respiration monitoring as fine-grained sensing with the proposed system to demonstrate sensing under device motion in real-world scenarios.

7.1 Implementation

The prototype of SonicBot is shown in Fig. 14. There are three components in the prototype, i.e., the sensing device, the moving platform, and the ground truth collector. We introduce the implementation details below.

7.1.1 Sensing Device. We adopt a general-purpose speaker (i.e., STAPEZ low distortion speaker) [18] for transmitting inaudible acoustic chirp signals, and a Seeed ReSpeaker 6-Mic Circular Array [39] as the receiver. Both

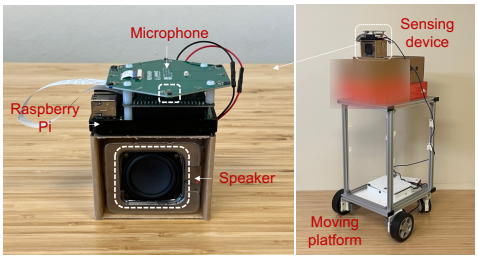


Fig. 14. The prototype of SonicBot system.

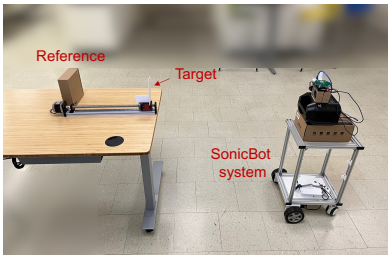
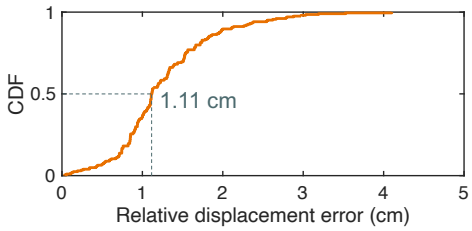
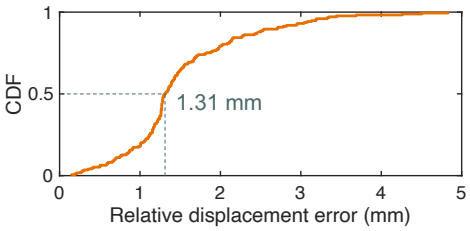


Fig. 15. Setup for benchmark experiment.



(a) The result of coarse-grained movement sensing.



(b) The result of fine-grained movement sensing.

Fig. 16. The CDF of the relative displacement error under device motion.

speaker and microphone array are connected on an Raspberry Pi 3 B+ platform and we control the Raspberry Pi to transmit and receive acoustic signals. Note that although the receiver has six microphones, only one microphone is used for sensing.

7.1.2 Moving Platform. We employ the ZiiROBOT smart active cart [68] to carry the sensing device. It is programmable to tune its moving distance, direction, and speed. To make sure the sensing device is at a suitable height for hand tracking and respiration sensing, we add two boxes to raise its height to 0.8 m.

7.1.3 Signal Parameter Settings. We adopt the frequency band from 18 kHz to 22 kHz (4 kHz bandwidth) which is inaudible for human ears for sensing. We set the chirp duration as 40 ms. The speaker and microphone audio sampling rate is 48 kHz. The signals are processed and analyzed by MATLAB on a MacBook Pro equipped with an Intel i7 processor and 16 GB memory.

7.1.4 Ground Truth Measurements. In benchmark experiment, we place the target on a 40 cm FUYU FSL40 linear guide slide [6], which is controlled by an Arduino UNO board. The accuracy of the linear guide slide is 0.03 mm and we use the moving distance of the slide as the ground truth. In field studies, we utilize the Hexoskin wearable smart clothing [10] to record the ground truth of the respiration rate of our participants, and manually record the push and pull gestures performed by our participants.

7.2 Benchmark Experiments

We verify the effectiveness of SonicBot by varying the parameters and conditions to study the impact of different factors on the sensing performance.

7.2.1 Experiment Setup. We show the benchmark experiment setup in Fig. 15. The details of the default setup are described as follows. We use an 8 $cm \times 15$ cm cardboard as the target and place it on the linear guide slide. We

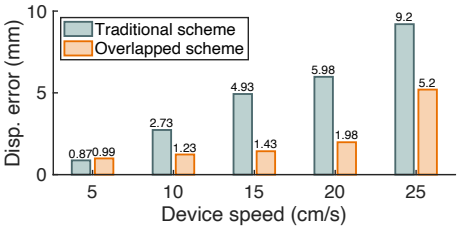


Fig. 17. Displacement error vs. speed.

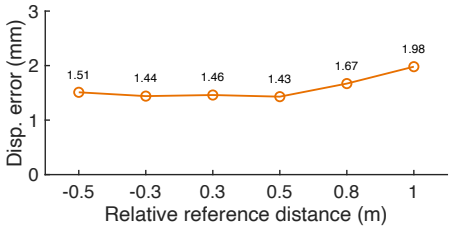


Fig. 18. Displacement error vs. reference-target distance.

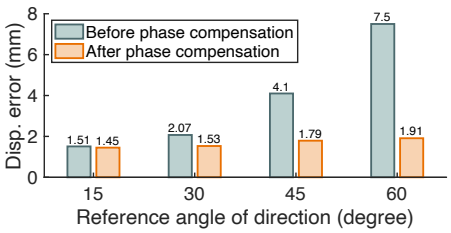


Fig. 19. Displacement error vs. reference direction.

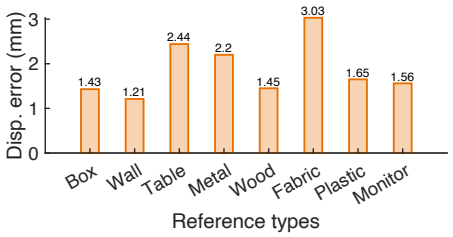


Fig. 20. Displacement error vs. reference type.

adopt a 19 $cm \times 19$ cm cardboard box as the reference. According to the survey, the acceptable minimum distance between human target and an assistive robot [2] is 0.92 m. Thus, we set the distance between the target and the device as 1.5 m and let the robot approach the target until the distance is 1 m. We place the reference 2 m away. We let the robot continuously approach and move away from the target with a moving distance of 0.5 m. The default moving speed of the robot is 0.15 m/s. The time shift of the overlapped transmitting scheme is set as 0.02 s. Experiments are conducted in a hall. Each data collection process lasts for 30 seconds, and we repeat the experiment 20 times for each setting. We let the target move forward and backward with a displacement of 20 cm and 5 mm to simulate coarse-grained and fine-grained movements, respectively.

7.2.2 Overall Performance. We use the relative displacement error as the metric to evaluate the performance of our system. As shown in Fig. 16, our system can achieve a median error of 1.11 cm and 1.31 mm for tracking coarse-grained and fine-grained movements, respectively.

7.2.3 Impact of Device Moving Speed. To explore the impact of device moving speed for fine-grained target movement, we conduct experiments under different speed settings, i.e., 5, 10, 15, 20 and 25 cm/s. We compare the displacement errors between using the traditional chirp transmitting scheme and the proposed overlapped scheme. As shown in Fig. 17, the displacement error increases with the speed goes up. The reason is that as the device moving speed increases, the number of the phase samples becomes insufficient. Fortunately, we can double the phase sampling rate by applying the overlapped transmitting scheme with a time shift of half chirp length, achieving a median error of 1.98 mm under the speed of 20 cm/s, significantly outperforming the traditional transmitting scheme. Note that for healthcare robot, the maximum safe speed is usually 20 cm/s [2]. We observe that the performance degrades when we further increase the speed to 25 cm/s. This is because the 0.02 s time shift between two consecutive chirps still cannot meet the required phase sampling rate at 25 cm/s. If large sensing distance is not required, more time shift can be applied to further increase the phase sampling rate.

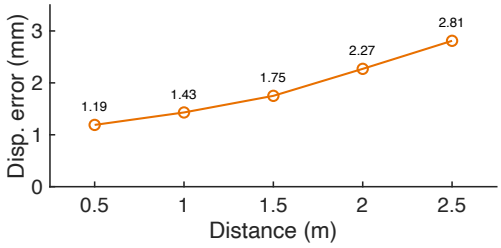


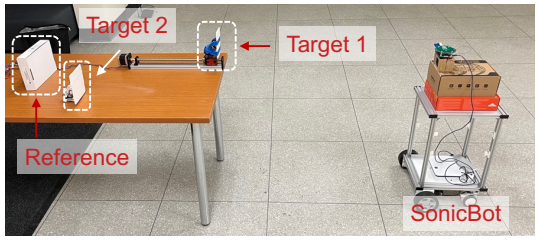
Fig. 21. Displacement error under different target-device distances.

7.2.4 Impact of Relative Distance between the Reference and the Target. To explore the impact of relative distance between the reference and the target, we conduct experiments under different distances of the reference with respect to the target, i.e., -0.5, -0.3, 0.3, 0.5, 0.8 and 1 *m*. The negative relative distance means the reference is in front of the target and the positive distance indicates the reference is behind the target. The results are shown in Fig. 18. We can observe from the results that the average median displacement error is 1.46 *mm* when the reference is within 0.5 *m* with respect to the target. However, the error slightly increases as the reference is further away. We believe this is because there is a delay between the target reflection and the reference reflection and this delay is larger as the relative distance increases. Therefore, the reflection from the target and the reflection from the reference might not represent the movements happening at the same time, causing an inconsistency between the target reflections and the reference reflection.

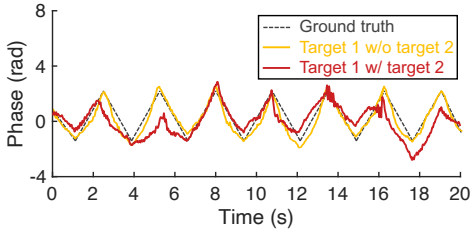
7.2.5 Impact of Reference Angle of Direction. To explore the impact of reference angle of direction with respect to the device, we conduct experiments under different angles from -60 degrees to 60 degrees at a step size of 15 degrees. We define the device moving direction as 0 degrees. We can observe from Fig. 19 that the median displacement error increases with the reference's angle of direction becomes larger. This is caused by the phase inconsistency as we analyzed in Sec. 6.4. After applying our phase compensation method, the error is reduced, proving the effectiveness of the proposed method. However, we still observe a slight error increase as the angle becomes larger. The reason is that the reference reflection becomes weaker due to the speaker's radiation pattern. We further conduct an experiment to demonstrate the feasibility of the phase compensation method in real-world settings. We program the robot to move following an S-shaped trajectory and keep the other setup as default. If the proposed compensation method is not applied, the median displacement error is 6.87 *mm*. After the proposed compensation method is applied, the error is reduced to 2.25 *mm*.

7.2.6 Impact of Distance. We conduct experiments to evaluate the effect of different distances between the sensing device and the target. The distance between the target and the device is increased from 0.5 m to 2.5 m at a step size of 0.5 m. The reference is 0.5 *m* further than the target with respect to the device. The results are shown in Fig. 21. We can observe that the error goes up as the distance increases because the SNR of the reflected signal becomes lower. However, we can see that the achieved distance error is still relatively small, i.e., 2.81 *mm* at a distance of 2.5 *m*.

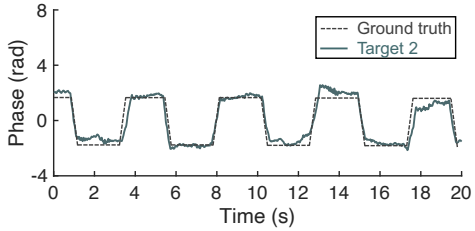
7.2.7 Impact of Reference Type. To explore the impact of reference type, we employ different objects as the reference, including the cardboard box, wall, table, metal cabinet, wooden cabinet, fabric couch, plastic trash bin, and computer monitor. The results are shown in Fig. 20. The wall achieves the best performance since it has a large reflection area and flat surface. When the fabric couch is chosen as the reference, the median displacement error is higher than other reference types. The reason is that different materials have different sound reflection



(a) Experiment setup.



(b) Obtained moving patterns of target 1 in the absence and presence of target 2.



(c) Obtained moving patterns of target 2.

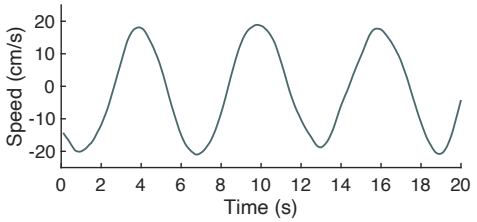
Fig. 22. Experiment results for multi-target sensing.

capabilities. Also, flat and smooth surfaces are able to reflect more signal. Therefore, the wall is a good reference and it is also the most common reflector in indoor environments.

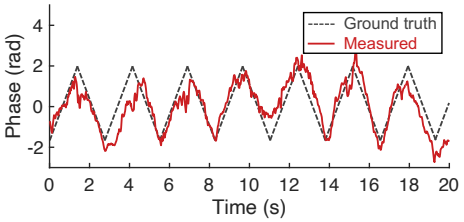
7.2.8 Impact of Ambient Noise. We conduct experiments under different ambient noises to evaluate their impact. We consider music and human speech as two sound sources and evaluate two different volume levels for each sound type. We use the Decibel X app [22] on iPhone 11 Pro to measure the sound pressure level of the environment and place it behind the sensing device. Note that the motor on the moving platform has an inherent noise level of 39.7 dB. We can observe from the results shown in Fig. 24 that the errors under different sound pressure levels are similar. The reason is that the frequency the ambient noise is below 14 kHz and therefore does not interfere with the signal in the range of 18-22 kHz we adopt for sensing.

7.2.9 Impact of Multiple Targets. We evaluate the performance of SonicBot on multi-target sensing. The experiment setup is shown in Fig. 22a. Target 1, target 2, and reference are placed at a distance of 1.2 m, 1.8 m, and 2 m, respectively with respect to SonicBot. Target 1 and target 2 move back and forth for a same displacement of 5 mm but in different patterns. Specifically, target 1 moves along a linear guide slide and its phase variation pattern is a triangle shape as shown in Fig. 22b, while target 2 is carried by the cozmo robot and the phase variation pattern is a square shape as shown in Fig. 22c. The ground-truth phase variation patterns are marked as dashed lines. As Fig. 22 shows, although the extracted target movements are slightly interfered when there are multiple targets, both movement patterns of target 1 and 2 can still be accurately recovered.

7.2.10 Impact of Varying Speeds. We conduct experiments to evaluate the performance of SonicBot under varying speeds. Specifically, we program the moving platform to move at varying speeds (lower than 0.2 m/s) and other experiment settings are kept the same as default. The result is shown in Fig. 23 and we can see that SonicBot can successfully recover the target movement information, demonstrating the feasibility of SonicBot under varying speeds.



(a) Speed variation of the device.



(b) Obtained moving patterns of the target.

Fig. 23. The recovered target movement under varying device speeds.

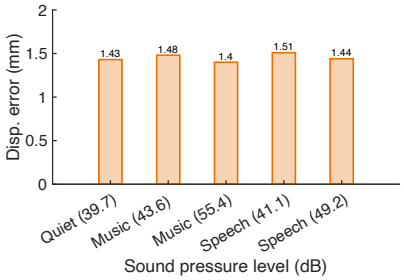


Fig. 24. The results under different noises.

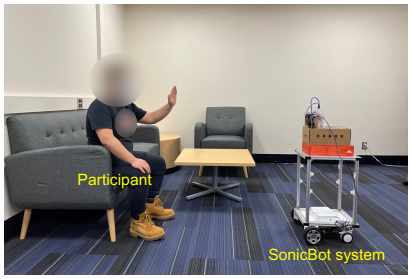
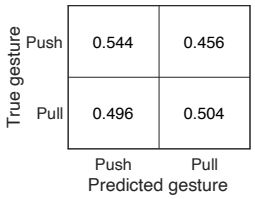
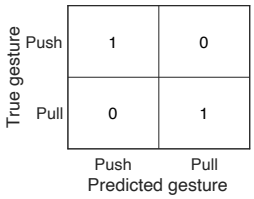


Fig. 25. A demonstration of the field studies in a living room.



(a) Before device motion cancellation.



(b) After device motion cancellation.

Fig. 26. The performance of push and pull gesture recognition.

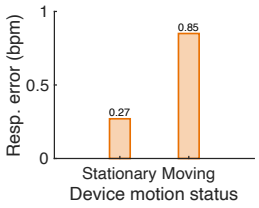


Fig. 27. The performance of respiration sensing when the device is stationary and moving.

7.3 Field Studies

We recruit five participants (including four males and one female, aged from 19 to 29) to evaluate the performance of our system on gesture recognition and respiration sensing in different indoor environments.

7.3.1 Experiment Setup. Experiments are conducted in three different environments. For gesture recognition, we present the confusion matrix of push and pull gesture recognition. We ask the participants to perform push and pull gestures five times during each 30-second data collection period. For respiration monitoring, we compare the measured respiration rate with the ground truth. We ask the participants to breathe naturally sitting in a chair or couch. Each data collection period lasts for 1 minute, and we repeat it for 10 times. A demonstration of the setup in the living room is shown in Fig. 25. We conduct experiments with both stationary device and moving device. The device moving parameters are the same as those described in the benchmark experiments.

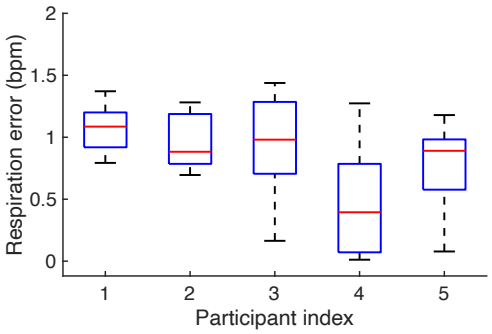


Fig. 28. The respiration rate error for different users.

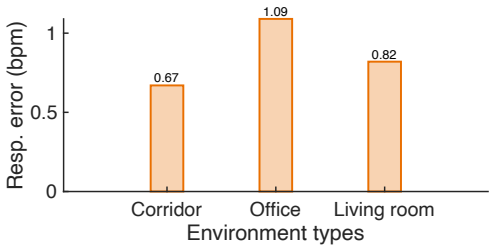


Fig. 29. The respiration rate error in different environments.

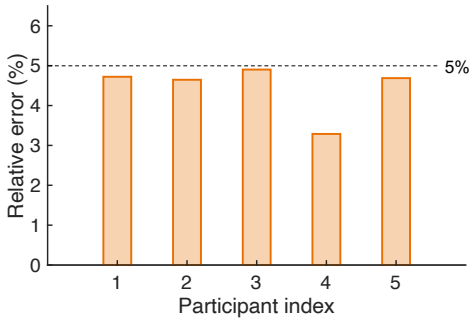


Fig. 30. Relative respiration error of different users.

7.3.2 Overall Performance. We show the confusion matrix of push and pull gesture recognition in Fig. 26. We compare the performance before applying the device motion cancellation scheme (Fig. 26a) and after the cancellation (Fig. 26b). The hand movements are overwhelmed by device motion without cancellation, resulting in a low accuracy of 52.4%. After applying the proposed scheme, we can effectively eliminate device motion and achieve an accuracy of 100%.

We evaluate the respiration sensing under device motion and compare the performance with that obtained when the device is stationary. The results are shown in Fig. 27. The respiration rate of most human beings falls in the range of 12 to 20 *bpm*. The absolute error of commodity respiration monitoring devices is required to be below 1 *bpm* [20, 55, 56]. SonicBot can achieve a respiration rate error of 0.27 *bpm* when the device is static and 0.85 *bpm* when the device is moving, satisfying the requirement.

7.3.3 Impact of User. Fig. 28 shows the respiration rate error of different participants. Our system can achieve an average error of 0.85 bpm among all participants. We can observe that participant 4 has a lower respiration error compared with other participants. The reason is that the chest movement of that participant's respiration is more significant than other participants. We also show the relative error of different users in Fig. 30. We can see that the relative errors are always lower than 5% [8, 31].

7.3.4 Impact of Environment. We conduct experiments in three different environments, including a 3 m wide corridor, a 5.5 $m \times 6$ m office, and a 3.4 $m \times 4$ m living room. Fig. 29 shows the respiration rate error in different

environments. In the corridor, SonicBot obtains a strong reflection from the wall as reference, resulting in a low respiration error. In the office, the environment is more complex compared with the other two environments since there are different types of furniture, resulting in a slightly higher error.

8 DISCUSSION

Large open space. The proposed system aims to serve as a healthcare robot for the elderly and patients in indoor environments such as homes and hospital wards. Due to the fast attenuation over distance for acoustic signals [14, 25], the proposed method may have difficulty working in a large open space. While this work focuses on dealing with device motion, we would like to mention that some recent studies [15, 25] proposed advanced signal processing and machine learning schemes to increase the sensing range.

Varying reference. Our system works under the assumption that we can obtain a stable reflection from the surrounding static object as the reference for eliminating the device motion. However, according to our observation from real-life experiments, when there are multiple static objects, the best reference may change. Hence, dynamically selecting the high-quality static object as the reference and seamlessly switching between references remains an interesting future work.

Limited device moving speed. Although our overlapped chirp transmitting scheme can mitigate the problem caused by device moving speed, due to the trade-off between sensing distance and phase sampling rate, the maximum device speed under which our system can work well is still limited given a particular sensing range requirement. We leave this problem as our important future work.

9 CONCLUSION

In this paper, we introduce a novel scheme that enables contact-free acoustic sensing on a moving robot. To achieve this goal, we study the reasons why traditional sensing methods become invalid under device motion, and propose a device motion cancellation scheme to support acoustic sensing under device motion. With an increasing demand for contact-free sensing on a moving robot for the elderly, COVID-19 patients, and caregivers, we believe the proposed system herein opens up a new pathway towards ubiquitous contact-free sensing.

REFERENCES

[1] Sivakumar Balasubramanian, Roberto Colombo, Irma Sterpi, Vittorio Sanguineti, and Etienne Burdet. 2012. Robotic assessment of upper limb motor function after stroke. *American journal of physical medicine & rehabilitation* 91, 11 (2012), S255–S269.

[2] Christopher Brandl, Alexander Mertens, and Christopher M Schlick. 2016. Human-robot interaction in assisted personal services: factors influencing distances that humans will accept between themselves and an approaching service robot. *Human Factors and Ergonomics in Manufacturing & Service Industries* 26, 6 (2016), 713–727.

[3] Buddy. 2018. Buddy the Emotional Robot. https://buddytherobot.com/en/buddy-the-emotional-robot/.

[4] Carlos Cifuentes, Ariel Braidot, Luis Rodríguez, Melisa Frisoli, Alfonso Santiago, and Anselmo Frizera. 2012. Development of a wearable ZigBee sensor system for upper limb rehabilitation robotics. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 1989–1994.

[5] CNN. 2021. Meet Grace, the ultra-lifelike nurse robot. https://www.cnn.com/2021/08/19/asia/grace-hanson-robotics-android-nurse-hnk-spc-intl/index.html.

[6] FUYU. 2021. High Precision Ball Screw Linear Motion Guide. https://www.fuyumotion.com/high-precision-ball-screw-linear-motion-guide-product/.

[7] Zhihui Gao, Ang Li, Dong Li, Jialin Liu, Jie Xiong, Yu Wang, Bing Li, and Yiran Chen. 2022. MOM: Microphone based 3D Orientation Measurement. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 132–144.

[8] Linfei Ge, Jin Zhang, and Jing Wei. 2018. Single-frequency ultrasound-based respiration rate estimation with smartphones. *Computational and mathematical methods in medicine* 2018 (2018).

[9] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. Soundwave: using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1911–1914.

[10] Hexoskin. 2021. Hexoskin Smart Garments. https://www.hexoskin.com/.

[11] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. 2015. Spotfi: Decimeter level localization using wifi. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. 269–282.

[12] Sébastien Laniel, Dominic Létourneau, Mathieu Labbé, François Grondin, Janice Polgar, and François Michaud. 2017. Adding navigation, artificial audition and vital sign monitoring capabilities to a telepresence mobile robot for remote home care applications. In *2017 International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 809–811.

[13] Min Ho Lee, Ho Seok Ahn, Kevin Wang, and Bruce A MacDonald. 2014. Design of a healthcare sensor managing system for vital sign measuring devices. In *International Conference on Simulation, Modeling, and Programming for Autonomous Robots*. Springer, 521–532.

[14] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 150–163.

[15] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. LASense: Pushing the Limits of Fine-grained Activity Sensing Using Acoustic Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[16] Jie Lian, Jiadong Lou, Li Chen, and Xu Yuan. 2021. EchoSpot: Spotting Your Locations via Acoustic Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–21.

[17] Jie Lian, Xu Yuan, Ming Li, and Nian-Feng Tzeng. 2021. Fall Detection via Inaudible Acoustic Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–21.

[18] BDNC (HOLDING) LIMITED. 2019. STAPEZ brand low distortion speaker. http://www.newbdnc.com/wp-content/uploads/datasheets/BFC-4448-24-4-006.pdf.

[19] Kang Ling, Haipeng Dai, Yuntang Liu, and Alex X Liu. 2018. Ultragesture: Fine-grained gesture sensing and recognition. In *2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 1–9.

[20] Haipeng Liu, John Allen, Dingchang Zheng, and Fei Chen. 2019. Recent development of respiratory rate measurement technologies. *Physiological measurement* 40, 7 (2019), 07TR01.

[21] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. 2021. BlinkListener: " Listen" to Your Eye Blink Using Your Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.

[22] SkyPaw Co. Ltd. 2021. *Decibel X: dB Sound Level Meter.* https://apps.apple.com/us/app/decibel-x-db-sound-level-meter/id448155923

[23] Saturnino Maldonado-Bascon, Cristian Iglesias-Iglesias, Pilar Martín-Martín, and Sergio Lafuente-Arroyo. 2019. Fallen people detection capabilities using assistive robot. *Electronics* 8, 9 (2019), 915.

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

[25] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-based room scale hand motion tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[26] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. 2017. Indoor follow me drone. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 345–358.

[27] Haimiao Mo, Shuai Ding, Shanlin Yang, Xi Zheng, and Athanasios V Vasilakos. 2020. The Role of Edge Robotics As-a-Service in Monitoring COVID-19 Infection. *arXiv preprint arXiv:2011.08482* (2020).

[28] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1515–1525.

[29] Rajalakshmi Nandakumar, Alex Takakuwa, Tadayoshi Kohno, and Shyamnath Gollakota. 2017. Covertband: Activity information leakage using music. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–24.

[30] Kai Niu, Fusang Zhang, Jie Xiong, Xiang Li, Enze Yi, and Daqing Zhang. 2018. Boosting fine-grained activity sensing by embracing wireless multipath effects. In *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*. 139–151.

[31] Shermeen Nizami, Amente Bekele, Mohamed Hozayen, Kimberley J Greenwood, JoAnn Harrold, and James R Green. 2018. Measuring uncertainty during respiratory rate estimation using pressure-sensitive mats. *IEEE Transactions on Instrumentation and Measurement* 67, 7 (2018), 1535–1542.

[32] Carina Barbosa Pereira, Xinchi Yu, Michael Czaplik, Rolf Rossaint, Vladimir Blazek, and Steffen Leonhardt. 2015. Remote monitoring of breathing dynamics using infrared thermography. *Biomedical optics express* 6, 11 (2015), 4378–4394.

[33] Vaughan Pratt. 1987. Direct least-squares fitting of algebraic surfaces. *ACM SIGGRAPH computer graphics* 21, 4 (1987), 145–152.

[34] Corinne Purtill. 2019. Stop Me if You've Heard This One: A Robot and a Team of Irish Scientists Walk Into a Senior Living Home. https://time.com/longform/senior-care-robot/.

[35] Estiko Rijanto, Erik Adiwiguna, Latif Rozaqi, Aryo Putro Sadono, and M Hafil Nugraha. 2019. Experimental Performance Evaluation of Computer Vision for an Upper Limbs Rehabilitation Robot. In *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*. IEEE, 59–63.

[36] SoftBank Robotics. 2008. Pepper the humanoid and programmable robot. https://www.softbankrobotics.com/emea/en/nao.
[37] SoftBank Robotics. 2014. NAO the humanoid and programmable robot. https://www.softbankrobotics.com/emea/en/pepper.
[38] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 474–485.
[39] Seeed. 2020. ReSpeaker 6-Mic Circular Array kit for Raspberry Pi - Seeed Wiki. https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/.
[40] Ava D Segal, Mark C Lesak, Anne K Silverman, and Andrew J Petruska. 2020. A Gesture-Controlled Rehabilitation Robot to Improve Engagement and Quantify Movement Performance. *Sensors* 20, 15 (2020), 4269.
[41] Ava D Segal, Mark C Lesak, Neil E Suttora, Anne K Silverman, and Andrew J Petruska. 2020. iRebot: An interactive rehabilitation robot with gesture control. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 5158–5161.
[42] Ruchika Sinhal, Kavita Singh, and MM Raghuwanshi. 2020. An overview of remote photoplethysmography methods for vital sign monitoring. *Computer Vision and Machine Intelligence in Medical Image Analysis* (2020), 21–31.
[43] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. 2020. SpiroSonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[44] IEEE Spectrum. 2016. Sweep Is a $250 LIDAR With Range of 40 Meters That Works Outdoors: Finally an affordable LIDAR for robots and drones. https://spectrum.ieee.org/sweep-lidar-for-robots-and-drones.
[45] Takuma Sumiya, Yutaka Matsubara, Miyuki Nakano, and Midori Sugaya. 2015. A mobile robot for fall detection for elderly-care. *Procedia computer science* 60 (2015), 870–880.
[46] Ke Sun, Wei Wang, Alex X Liu, and Haipeng Dai. 2018. Depth aware finger tapping on virtual displays. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 283–295.
[47] Ke Sun and Xinyu Zhang. 2021. UltraSE: single-channel speech enhancement using ultrasound. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 160–173.
[48] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 591–605.
[49] Annick AA Timmermans, Henk AM Seelen, Richard PJ Geers, Privender K Saini, Stefan Winter, Juergen Te Vrugt, and Herman Kingma. 2010. Sensor-based arm skill training in chronic stroke patients: results on treatment outcome, patient motivation, and system usability. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18, 3 (2010), 284–292.
[50] Amari Tomoya, Satoru Nakayama, Atsushi Hoshina, and Midori Sugaya. 2017. A mobile robot for following, watching and detecting falls for elderly care. *Procedia computer science* 112 (2017), 1994–2003.
[51] Anne Trafton. 2020. Robot takes contact-free measurements of patients' vital signs. https://news.mit.edu/2020/spot-robot-vital-signs-0831.
[52] Mark van Gastel, Sander Stuijk, and Gerard de Haan. 2016. Robust respiration detection from remote photoplethysmography. *Biomedical optics express* 7, 12 (2016), 4941–4957.
[53] Anran Wang and Shyamnath Gollakota. 2019. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
[54] Anran Wang, Dan Nguyen, Arun R Sridhar, and Shyamnath Gollakota. 2021. Using smart speakers to contactlessly monitor heart rhythms. *Communications biology* 4, 1 (2021), 1–12.
[55] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[56] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.
[57] Wei Wang, Alex X Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 82–94.
[58] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. 2020. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing* (2020).
[59] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. 2014. Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. 237–248.
[60] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. 2017. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th annual international conference on mobile systems, applications, and services*. 15–28.

[61] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. 2019. FarSense: Pushing the range limit of WiFi-based respiration sensing with CSI ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.

[62] Fusang Zhang, Zhaoxin Chang, Kai Niu, Jie Xiong, Beihong Jin, Qin Lv, and Daqing Zhang. 2020. Exploring lora for long-range through-wall sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–27.

[63] Fusang Zhang, Zhaoxin Chang, Jie Xiong, Rong Zheng, Junqi Ma, Kai Niu, Beihong Jin, and Daqing Zhang. 2021. Unlocking the Beamforming Potential of LoRa for Long-range Multi-target Respiration Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–25.

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

[65] Fusang Zhang, Daqing Zhang, Jie Xiong, Hao Wang, Kai Niu, Beihong Jin, and Yuxiang Wang. 2018. From fresnel diffraction model to fine-grained human respiration sensing with commodity wi-fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–23.

[66] Qian Zhang, Dong Wang, Run Zhao, and Yinggang Yu. 2021. SoundLip: Enabling Word and Sentence-level Lip Interaction for Smart Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–28.

[67] Qian Zhang, Dong Wang, Run Zhao, Yinggang Yu, and Junjie Shen. 2021. Sensing to hear: Speech enhancement for mobile devices using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–30.

[68] ZiiROBOT. 2021. Smart Cart C1. https://ziirobot.com/pages/applications.