Embracing Distributed Acoustic Sensing in Car Cabin for Children Presence Detection

YUQI SU, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, China
FUSANG ZHANG, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, China
KAI NIU, Peking University & Beijing Xiaomi Mobile Software Company Ltd., China
TIANBEN WANG, Northwest A & F University, China
BEIHONG JIN, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, China
ZHI WANG, Institute of Software, Chinese Academy of Sciences & University of Chinese Academy of Sciences, China
YALAN JIANG, Beijing Xiaomi Mobile Software Company Ltd., Beijing, China
DAQING ZHANG, Peking University, china, & Institut Polytechnique de Paris, France
LILI QIU, Microsoft Research Asia, Shanghai, China
JIE XIONG, Microsoft Research Asia, Shanghai, China & University of Massachusetts Amherst, United States

Contactless acoustic sensing has been actively exploited in the past few years to enable a large range of applications, ranging from fine-grained vital sign monitoring to coarse-grained human tracking. However, existing acoustic sensing systems mainly work on smartphone or smart speaker platforms. In this paper, we envision an exciting new acoustic sensing platform, i.e., car cabin which is inherently embedded with a large number of speakers and microphones. We propose the new concept of distributed acoustic sensing and develop novel designs leveraging the unique characteristics of rich multi-path in car cabin to enable fine-grained sensing even when the primary reflection is totally blocked. By using child presence detection as the application example, we show that child presence can be detected through body motions or even subtle breath (when the child is sleeping or in coma) at all locations in the cabin without any blind spots. We further show that the proposed system can robustly work in different car cabins, achieving an average detection accuracy of 97% and a false alarm rate always below 2% under different scenarios including those challenging ones such as rear-facing seat blockage. We believe the proposed distributed sensing modality in car cabin pushes acoustic sensing one big step towards real-life adoption.

Additional Key Words and Phrases: Distributed acoustic sensing, Smart cabin, Superposition model, Signal enhancement, Multi-path combination, Children Presence Detection

Authors' addresses: Yuqi Su, Institute of Software, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China. Fusang Zhang, State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China. Kai Niu, Peking University; Beijing Xiaomi Mobile Software Company Ltd., Beijing, China. Tianben Wang, College of Mechanical and Electronic Engineering, Northwest A & F University, Yangling, Shaanxi, China. Beihong Jin, Institute of Software, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China. Zhi Wang, State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China. Yalan Jiang, Beijing Xiaomi Mobile Software Company Ltd., Beijing, China. Daqing Zhang, Telecom SudParis, Institut Polytechnique de Paris, Evry, France; School of Computer Science, Peking University, Beijing, China. Lili Qiu, Microsoft Research Asia, Shanghai, China. Jie Xiong, University of Massachusetts Amherst, Amherst, Massachusetts; United States Microsoft Research Asia, Shanghai, China. Corresponding Authors: Fusang Zhang, Beihong Jin; E-mail: fusang@iscas.ac.cn, Beihong@iscas.ac.cn.

2474-9567/2024/3-ART16
https://doi.org/10.1145/3643548

ACM Reference Format:
Yuqi Su, Fusang Zhang, Kai Niu, Tianben Wang, Beihong Jin, Zhi Wang, Yalan Jiang, Daqing Zhang, Lili Qiu, and Jie Xiong. 2024. Embracing Distributed Acoustic Sensing in Car Cabin for Children Presence Detection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 1, Article 16 (March 2024), 28 pages. https://doi.org/10.1145/3643548

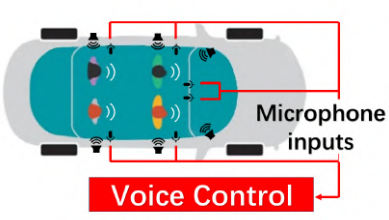
1 INTRODUCTION

The demand for Child Presence Detection (CPD) in car rapidly grows in the last few years. According to a survey from the U.S. [6], from 1990 to 2022, 1,054 children died from heat stroke after being left behind in vehicles, which could be avoided if a CPD system was installed. It is known that children's temperature can rise much more quickly (3-5 times) than adults [3]. When a child's internal temperature reaches 104 °F, major organs begin to shut down, and when the temperature reaches 107 °F, the child may die. Even on mild days, the temperature in a car can rise by 19 degrees in merely 10 minutes [3]. The Euro NCAP (New Car Assessment Program) announced that from 2022, they started award rating points of CPD for all cars [7]. The US is planning to make CPD system a mandatory equipment in motor vehicles from 2025 [2]. Therefore, CPD has drawn a significant amount of attention from automakers and customers due to the fatal consequences of leaving children alone in vehicles.

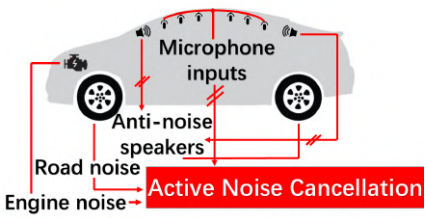
Existing CPD solutions mainly adopt cameras [19], pressure sensors [17] as well as Ultra-Wideband (UWB) [25, 56], millimeter-wave (mmWave) [4, 8] and Wi-Fi sensing [30, 61, 62]. However, cameras raise privacy concerns and pressure sensors fail to work when babies are not on the car seat (e.g., babies may fall off the seat and lie in the footwell region). Recent RF signal-based solutions (e.g., UWB and millimeter wave radar) [4, 25, 56] mount device on the car ceiling to monitor the presence of child. However, its coverage is insufficient and the use of multiple radar systems will result in high cost. More importantly, these approaches require dedicated radar hardware that is not typically available inside car cabin. Although Wi-Fi sensing [63][65] is popular in smart home environment, modern cars do not have Wi-Fi access point installed. Besides, due to its low channel bandwidth, Wi-Fi signals are easily affected by interference inside and outside of the car.

In this paper, we propose to exploit acoustic signals for sensing in car cabin. We observe that more and more cars are equipped with a large number of acoustic modules (i.e., speakers and microphones). From 2021, Tesla Model S and X are equipped with a new automotive sound system which consists of 22 speakers [5]. Traditional automakers are also following this trend. BMW develops the Bowers & Wilkins diamond surround sound system with more than 30 speakers and Mercedes-Benz equips the Burmester audio system with 31 speakers [12]. Cadillac Escalade is even equipped with 36 speakers [16], as shown in Fig. 1. Furthermore, there are also multiple microphones equipped in modern cars [57], as shown in Fig. 2. They are placed on the rearview mirror, beside the handle grips, on the control panel, etc. Currently, the audio system in the car cabin is mainly used for voice control [21], active noise cancellation (ANC) [58], hands-free voice call [46] and in-car entertainment.

Fig. 1. 2021 Cadillac Escalade has 36 speakers distributed in the car cabin.



(a) Voice control system



(b) Active noise cancellation (ANC)

Fig. 2. Multiple microphones in car audio system.

In this paper, we propose to leverage the large number of speakers/microphones in car cabin for child presence detection. While car-based acoustic sensing shares a lot of similarities with conventional acoustic sensing, the key difference is the *distributed* nature of the microphones and speakers in car environment. Even though a smart speaker has multiple microphones, the microphones in the smart speaker are all co-located, which is completely different from the scenario in car environment: the acoustic modules are distributed throughout the interior of the car, including the dashboard, pillar, car roof and rear deck as shown in Fig. 1. To effectively utilize all the acoustic modules in car cabin for sensing, we propose the concept of *distributed acoustic sensing*. We envision that the new distributed sensing modality could push acoustic sensing one big step towards real-life adoption. Specifically, distributed acoustic sensing in car cabin can provide the following advantages:

- **Low hardware cost.** As acoustic modules have become a norm in cars, we do not require deploying other dedicated hardware for sensing.
- **High density deployment.** It is rare to have such a high-density deployment in other real-world settings. For example, usually only one Wi-Fi access point is available in one apartment for Wi-Fi sensing. On the contrary, there are more than 20 acoustic modules in a small cabin space (2-3.5 m^3) in car environment. This high-density deployment presents unique opportunities to achieve not just highly accurate sensing but also full-coverage sensing, i.e., covering all the locations in the car cabin without any dead zones.

We believe the acoustic modules in car cabins provide unique opportunities to realize child presence detection. The fine-grained sensing capability of acoustic sensing can help detect baby motions and subtle breath even when a baby is in a coma. The distributed nature of the acoustic modules helps achieve coverage of all the locations in a car. This is critical when babies fall off the seats. We need to address a number of challenges in order to turn the idea into a working system.

- The first issue is that the acoustic speakers are designed for in-car entertainment and all the speakers emit signals simultaneously. On the other hand, if we would like signals emitted from distributed speakers to constructively combine at a particular location, we need to precisely control the delay of each speaker. It is thus challenging to ensure good sensing performance without controlling the delay of each individual speaker.
- The second challenge is that even with a lot of speakers, we still face non-line-of-sight (NLoS) issues in CPD. There are two reasons behind this. The first reason is that although there are $18-30$ speakers in the car cabin, not all of them are capable of emitting ultrasound. To achieve the best entertainment experience, the speakers are grouped into three categories, i.e., high-pitched speakers, medium-pitched speakers, and low-pitched speakers. Therefore, only 1/3 of the speakers can be utilized for sensing. The second reason is that the situation in a real car cabin is quite complicated. The rear-facing baby seat can attenuate and even completely block the signal. Babies can be covered by blanket and the blanket can severely affect breath sensing when the signal is weak. Note that we utilize signals reflected from the baby for sensing. When there is no LoS path from the transmitter or receiver to the target, we need to use the signal reflected twice (i.e., from the target and the car) which is extremely weak for sensing.

To address the first challenge, we propose to use signal phase amplification technique for sensing. The key intuition is that the phase change indicating target motion displacement can be accumulated. To address the second challenge, we propose to leverage the unique characteristics of extremely rich multipath in car cabin. We utilize not just the primary reflection but also the secondary reflection for sensing. By fusing multiple weak secondary reflections, we are able to achieve full-coverage in-cabin sensing even when the primary reflection is 100% blocked.

The proposed distributed acoustic sensing system has been deployed in various types of vehicles for CPD by detecting baby motions and subtle breaths. We collected a large data set (i.e., 130 hours of acoustic signal recordings with each monitoring section lasting for 1-2 minutes) with/without a child in a car. A high-fidelity child

simulator [13] with realistic newborn traits is used in our system to serve as the target. Through comprehensive experiments, we demonstrate the advantages of the proposed system: 1) Full coverage without blind spots. The experiments verify that child presence can be detected at all possible locations (front and back seats, footwell zone, etc.) in different car models including both sedan and SUV. 2) Sensing under blockage. When the child is covered with a blanket and blocked by rear-facing seat, our system can still extract subtle signal variation to accurately detect child's presence. 3) Strong anti-interference capability. Our system is robust against noise such as music play and human voice. The interference from pedestrians outside does not increase the false positive rate. The main contributions of our work are summarized as follows.

- We are the first to propose the concept of distributed acoustic sensing, bringing full-coverage sensing to car cabin platform.
- By fully leveraging the unique properties of in-cabin environment, we propose novel designs to enable highly accurate in-cabin sensing without the need of modifying the existing hardware. Specifically, we propose to use signal phase for distributed sensing and adopt secondary reflections to address the primary reflection blockage issue.
- We build a real-time child presence detection system without the need for any training. We compare our system with RF-based solutions (e.g., mmWave radar and UWB radar) to demonstrate the superior performance of the proposed system in terms of sensing accuracy and sensing coverage. The demo video can be found at https://youtu.be/dz_eXa41swo.

2 BACKGROUND

In this section, we first present the transmitted baseband signals. Then we introduce traditional acoustic sensing using a single pair of speaker and microphone.

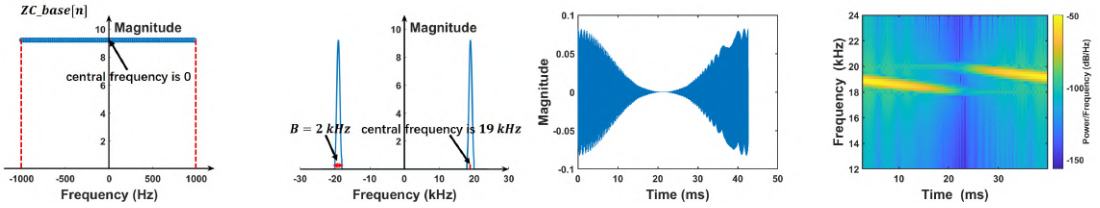
2.1 Transmitted acoustic signal

Acoustic signals have been extensively utilized for contactless sensing. In this study, we consider various types of waveforms as transmitted signals, including Pure Tone, Orthogonal Frequency Division Multiplexing (OFDM), Chirp and Pseudo-Noise Sequence (PN sequence). Note that Pure Tone and OFDM signals are relatively vulnerable to interference, while Chirp signal exhibits audible leakage [33]. In contrast, PN sequences generate pseudo-random binary sequences, which possess optimal cross-correlation property and also exhibit robustness against interference. Consequently, we choose the Zadoff-Chu (ZC) sequence [41] as our baseband signal, which is one type of widely used PN sequence. The signal can be represented as follows:

$$zc\,[n] = \exp\left[-j\frac{\pi \cdot u \cdot n\,(n+1+2q)}{N_{zc}}\right], \tag{1}$$

where u is the root sequence, $u \in [1, (N_{zc}-1)]$, N_{zc} is the length of the ZC sequence, u and N_{zc} are mutually prime, i.e., $\gcd(u, N_{zc}) = 1$. n denotes the sequence position, $n = 0, 1, \ldots, N_{zc}-1$. q is a constant ($q \in Z$ and Z denotes the set of integers). With the generated baseband signal $zc_{base}[n]$ of length N_{zc}, we expand the sequence to length L using frequency domain interpolation (where L is the frame length of the transmitted signal). Subsequently, we move the baseband sequence to the carrier frequency band with a center frequency of f_c. This results in the transmitted signal occupying a bandwidth of $B = \frac{N_{zc}}{L} \times f_s$, with the occupied frequency band defined as $(f_c - \frac{B}{2}, f_c + \frac{B}{2})$. To obtain the complete signal from the modulated signal, we set the negative frequency component as the conjugate counterpart of the positive frequency component. Finally, we perform an Inverse Fast Fourier Transform (IFFT) operation on the frequency domain signal to obtain the time domain

signal for transmission $zc_T[n]$.[1] Fig. 3 shows the key intermediate steps of the signal modulation process. Fig. 3a represents a ZC baseband signal occupying a 2 kHz bandwidth in the frequency domain. As the rectangular window can result in significant side lobes when transformed into a sinc function in the time domain, causing audible leakage, it is necessary to apply a Hamming window to smooth the signal as shown in Fig. 3b. Fig. 3c and Fig. 3d represent the time domain and spectral representation of the modulated transmitted signal, respectively.



(a) Baseband signal in frequency domain

(b) Modulated signal in frequency domain

(c) Transmitted signal in time domain

(d) Time-frequency spectrum of transmitted signal

Fig. 3. The process of generating the transmitted signal.

2.2 Traditional acoustic sensing

Now we introduce traditional acoustic sensing system with only one pair of speaker and microphone. As shown in Fig. 4, once the acoustic signal is transmitted from the speaker, the microphone captures the signal from the direct path as well as the reflection paths from the surrounding environment. The received signal is represented as:

$$zc_R[n] = \sum_{i=1}^{P} A_i e^{-j\varphi_i} zc_T\left[n - \frac{d_i}{c \cdot f_s}\right], \tag{2}$$

where P is the number of paths, A_i is the attenuation coefficient of path i and d_i denotes the length of path i. φ_i is the phase shift caused by the propagation/reflection of path i and $\varphi_i = 2\pi\frac{d_i}{\lambda_c}$ (λ_c is the signal wavelength). The received signal $zc_R[n]$ can be considered as a cyclic convolution $cir[n] * zc_T[n] \Leftrightarrow CIR[n] \times ZC_T[n]$ ($\Leftrightarrow$ denotes the Fourier Transform Pair symbol). Thus, the multi-path propagation Channel Impulse Response (CIR) of an acoustic sensing system can be expressed as:

$$cir[n] = IFFT\left(\frac{ZC_R[n]}{ZC_T[n]}\right) = \sum_{i=1}^{P} A_i e^{-j\varphi_i}\delta\left[n - \frac{d_i}{c \cdot f_s}\right], \tag{3}$$

where $\delta[n]$ is a discrete Dirac's delta function, which is non-zero only when $n = \frac{d_i}{c \cdot f_s}$. As the received signal is extracted by a frequency rectangular gate function with bandwidth B, the corresponding time domain CIR is the convolution of the *sinc* function with the impulse response:

$$cir[n] = \sum_{i=1}^{P} A_i e^{-j\varphi_i}\delta\left[n - \frac{d_i}{c \cdot f_s}\right] * sinc[n] = \sum_{i=1}^{P} A_i e^{-j\varphi_i} sinc\left[n - \frac{d_i}{c \cdot f_s}\right]. \tag{4}$$

[1]In our design, the ZC baseband sequence length $N_{zc} = 85$, the transmitted signal frame length $L = 2048$, the bandwidth $B = \frac{N_{zc}}{L} * f_s = \frac{85}{2048} * 48kHz = 2kHz$, the center frequency $f_c = 19kHz$, and the signal occupied frequency is between $18kHz$ and $20kHz$.

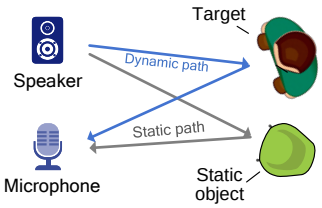


Fig. 4. Traditional acoustic sensing scenario using a single pair of speaker and microphone.

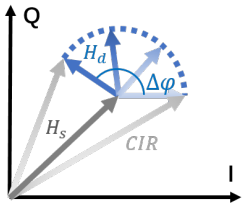


Fig. 5. The dynamic signal vector (H_d) rotates with respect to the static signal vector (H_s).

We select the path d_p that contains the target information in the CIR. Assume the target performs subtle motions with a displacement of $\Delta d_p(t)$, the signal can be divided into dynamic component and static component, which are represented as:

$$cir_t[n] = \underbrace{A_{sta}e^{-j2\pi\frac{d_p}{\lambda_c}}sinc\left(n-\frac{d_p}{c\cdot f_s}\right)}_{Static\ component} + \underbrace{A_{mov}e^{-j2\pi\frac{d_p+\triangle d_p(t)}{\lambda_c}}sinc\left(n-\frac{d_p+\triangle d_p(t)}{c\cdot f_s}\right)}_{Dynamic\ component}, \tag{5}$$

where A_{sta} and A_{mov} are the attenuation coefficient of static component and dynamic component, respectively. Here, the static component reflected from static object does not change with time and is denoted as static vector (H_s) on the I-Q plane. The dynamic vector (H_d) indicates the signal reflected from the target. The target movement causes the signal phase to change with time and the phase shift is $\triangle\varphi = 2\pi\frac{\triangle d_p(t)}{\lambda_c}$. As shown in Fig. 5, the dynamic vector rotates with respect to the static vector, and the phase shift caused by a subtle target motion $\triangle\varphi$ is less than 2π, forming a segment of arc on the I-Q plane. By obtaining the signal phase change, the amount of motion displacement can be calculated.

3 DESIGN OF DISTRIBUTED ACOUSTIC SENSING

In this section, we introduce a new sensing modality leveraging multiple speakers and microphones in a car cabin. We propose the signal superposition model for distributed speakers located at different locations. Subsequently, we derive the Channel Impulse Response (CIR) of the multi-speaker system and highlight its benefits in enhancing the sensing performance. Finally, to amplify the target motion-caused signal variation and increase the sensing sensitivity, we combine the received signals from multiple microphones to further improve the performance.

3.1 Signal Superposition Model for Distributed Speakers

Now we model the superposition of signals from multiple distributed speakers. In a car audio system, the system typically consists of high-pitched speakers, medium-pitched speakers and low-pitched speakers. Because only the high-pitched speakers can produce sounds at frequencies that are inaudible to humans ($> 18\ kHz$), we consider a typical audio system[2] consisting of six high-pitched speakers and eight microphones as shown in Fig. 6. The six speakers are installed in the panels below the windshield, A-pillar, and B-pillar of the left/right frame respectively, while the eight microphones are located above the rearview mirror and in the B-pillar of the left/right frame. The ZC sequence signal is transmitted simultaneously and continuously by all speakers, which propagates through various paths to reach the target (e.g., a child), before reaching the microphones.

[2]The locations and numbers of the speakers and microphones follow the typical design in many modern car models according to our survey of car surround audio system [26, 39] and the solution of automotive audio bus [31].

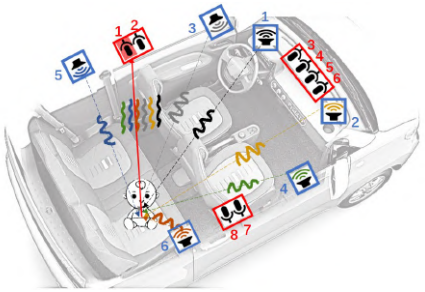


Fig. 6. Typical deployment of acoustic modules in car cabin.

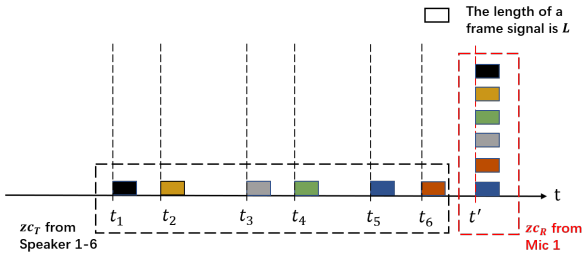


Fig. 7. Sequential time chart of received signals.

3.1.1 Modeling Received Signal from Multiple Speakers. Now we mathematically model the signal received at a microphone from distributed speakers. As shown in Fig. 7, at a random timestamp t', we have signals from six speakers received at the microphone. Due to the path length difference, the signals captured at timestamp t' are emitted out at different timestamps (i.e., $t_1 - t_6$) from the distributed speakers. Note that this does not contradict with the reality that the six speakers can only transmit at the same time as long as the speakers continuously transmit. To simplify the representation, we first model a frame[3] of the signal received at the microphone by considering one path from each speaker:

$$\begin{aligned} zc_R\left[\frac{t'}{f_s} : \frac{t'}{f_s} + L\right] = & A_1 e^{-j\varphi_1} zc_T\left[\frac{t_1}{f_s} : \frac{t_1}{f_s} + L\right] + A_2 e^{-j\varphi_2} zc_T\left[\frac{t_2}{f_s} : \frac{t_2}{f_s} + L\right] + A_3 e^{-j\varphi_3} zc_T\left[\frac{t_3}{f_s} : \frac{t_3}{f_s} + L\right] + \\ & A_4 e^{-j\varphi_4} zc_T\left[\frac{t_4}{f_s} : \frac{t_4}{f_s} + L\right] + A_5 e^{-j\varphi_5} zc_T\left[\frac{t_5}{f_s} : \frac{t_5}{f_s} + L\right] + A_6 e^{-j\varphi_6} zc_T\left[\frac{t_6}{f_s} : \frac{t_6}{f_s} + L\right], \end{aligned} \tag{6}$$

where $\frac{t_m}{f_s}$ is the start sample index of a signal frame. Next, we use the start sample $zc_T\left[\frac{t_m}{f_s}\right]$ to indicate a frame of signal $zc_T\left[\frac{t_m}{f_s} : \frac{t_m}{f_s} + L\right]$. Thus, we can rewrite Eq. (6) as:

$$zc_R\left[\frac{t'}{f_s}\right] = \sum_{m=1}^{6} A_i e^{-j\varphi_m} zc_T\left[\frac{t_m}{f_s}\right]. \tag{7}$$

Then, we use n to denote the sample index of the received signal, i.e., let $n = \frac{t'}{f_s}$. Thus, the sample index of the transmitted signal is $\frac{t_m}{f_s} = \frac{t'}{f_s} + \frac{t_m}{f_s} - \frac{t'}{f_s} = n - \frac{t'-t_m}{f_s} = n - \frac{d_m}{c \cdot f_s}$, where $\frac{d_m}{c} = t' - t_m$ denotes the time required for the acoustic signal to propagate for a path length of d_m from *Speaker* m to *Mic* 1. So the above equation can be rewritten as:

$$zc_R[n] = \sum_{m=1}^{6} A_m e^{-j\varphi_m} zc_T\left[n - \frac{d_m}{c \cdot f_s}\right]. \tag{8}$$

In practice, there are actually multiple paths from each speaker to the microphone, so the received signal for a multi-speaker system can be represented as:

$$zc_R[n] = \sum_{m=1}^{6} \sum_{i_m=1}^{P_m} A_{i_m} e^{-j\varphi_{i_m}} zc_T\left[n - \frac{d_{i_m}}{c \cdot f_s}\right], \tag{9}$$

[3]We have 2048 samples in a frame.

where P_m $(m = 1, 2, \ldots, 6)$ denotes the number of paths from *Speaker m* to *Mic* 1, and i_m denotes the index of path from *Speaker m*. Without loss of generality, if the number of speakers in the system is M, the final form of the received signal of the multi-speaker system can be represented as:

$$zc_R[n] = \sum_{m=1}^{M} \sum_{i_m=1}^{P_m} A_{i_m} e^{-j\varphi_{i_m}} zc_T \left[n - \frac{d_{i_m}}{c \cdot f_s} \right]. \tag{10}$$

In a single-speaker system with *Speaker m*, the number of signal propagation paths is P_m. If the system has M speakers, the total number of paths (P) can be obtained by summing P_m for all speakers, resulting in $P = \sum_{m=1}^{M} P_m$. Deploying multiple speakers in the environment can thus significantly increase the number of signal propagation paths and this brings a great opportunity to achieve full-coverage sensing.

3.1.2 Understanding the superposition of multi-speaker signals. The signal processing of ZC sequence usually happens in frequency domain owing to a much lighter computational load [50]. To obtain the channel information, i.e., channel impulse response (CIR) for sensing, we perform a cross-correlation operation between a frame of the received signal and the transmitted signal in frequency domain. The consecutive frames of CIR are then combined to create the *Range-Time Map*. Here we derive the CIR of the multiple-speaker system, which can be represented as:

$$cir[n] = \sum_{m=1}^{M} \sum_{i_m=1}^{P_m} A_{i_m} e^{-j\varphi_{i_m}} sinc \left[n - \frac{d_{i_m}}{c \cdot f_s} \right]. \tag{11}$$

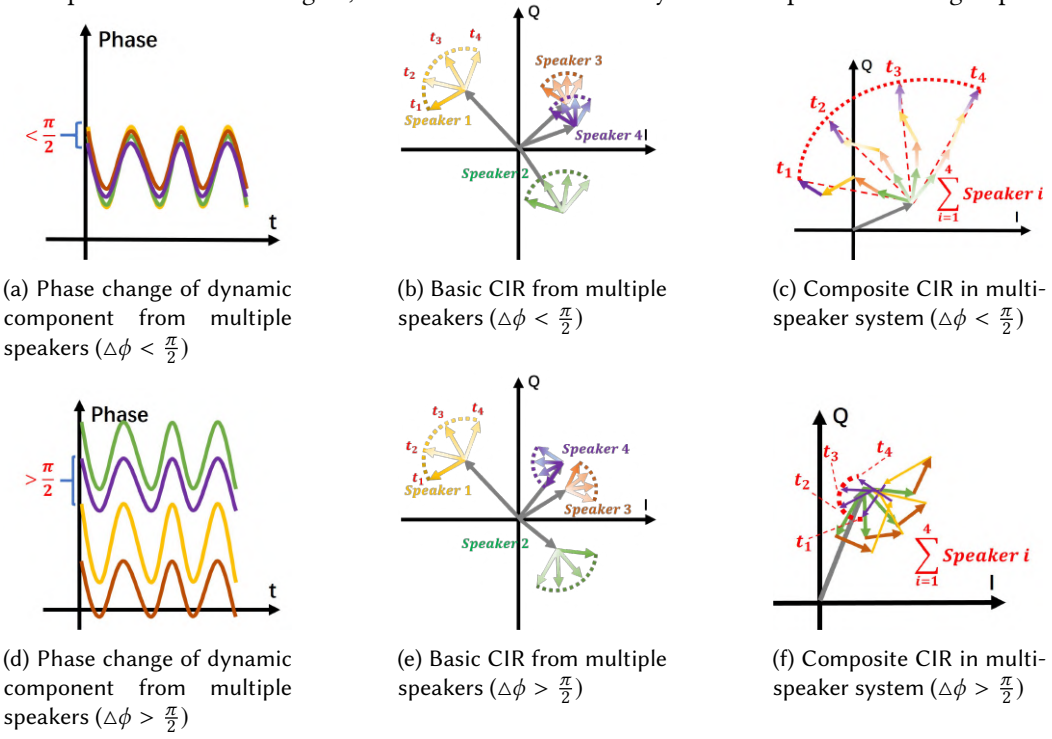
The above equation shows that the signal sent by the m-th speaker has a path length of d_{i_m} to the microphone, which results in an impulse response at the sampling point $\frac{d_{i_m}}{c \cdot f_s}$. When there are signals from multiple speakers with similar path lengths, i.e., with path length difference smaller than the range bin resolution, these signals can not be separated and will be superimposed as a single composite signal. We group the composite signal into dynamic component and static component according to the presence or absence of a moving target on the path. Consequently, the CIR of the composite signals at time t can be represented as:

$$cir_c(n, t) = \underbrace{\sum_{m=1}^{M} A_{mSTA} e^{-j2\pi \frac{d_m}{\lambda_c}} sinc \left(n - \frac{d_m}{c \cdot f_s} \right)}_{Static\ component} + \underbrace{\sum_{m=1}^{M} A_{mMOV} e^{-j2\pi \frac{d_m + \triangle d_m(t)}{\lambda_c}} sinc \left(n - \frac{d_m + \triangle d_m(t)}{c \cdot f_s} \right)}_{Dynamic\ component}, \tag{12}$$

where d_m denotes the path length of the reflected signal from the m-th speaker and $\triangle d_m(t)$ denotes the path length change with time. A_{mSTA} and A_{mMOV} denote the overall attenuation coefficient of the static component and dynamic component from the m-th speaker, respectively.

For the static component, it does not change with time, and the superposition follows the parallelogram rule which is expressed as a vector with fixed magnitude and direction on the I-Q plane. It has no impact on the motion information extraction. For the dynamic component, as the change of the signal from each speaker is caused by the same target motion, the phase change $\triangle\varphi_m(t) = 2\pi \frac{\triangle d_m(t)}{\lambda_c}$ follows the same time-varying pattern as shown in Fig. 8a and Fig. 8d. When the dynamic vector phase difference $\triangle\phi$ between multiple speakers is less than $\frac{\pi}{2}$ as shown in Fig. 8a, we select four moments of the CIR from each speaker and show them in Fig. 8b. The dynamic vectors of each moment on the I-Q plane have different phases but they all have the same overall direction. As shown in Fig. 8c, the energy of the composite dynamic component is given by

$\left|\sum_{m=1}^{M} A_{mMOV} e^{-j2\pi \frac{d_m+\triangle d_m(t)}{\lambda_c}} sinc\left(n - \frac{d_m+\triangle d_m(t)}{c \cdot f_s}\right)\right|$, which represents the superposition of the dynamic components of M speakers and is larger than the dynamic component of any single speaker. However, when the dynamic vector phase difference $\triangle\phi$ between multiple speaker is larger than $\frac{\pi}{2}$ (as shown in Fig. 8e), the composite dynamic component is shown in Fig. 8f, which is smaller than the dynamic component of a single speaker.



(a) Phase change of dynamic component from multiple speakers ($\triangle\phi < \frac{\pi}{2}$)

(b) Basic CIR from multiple speakers ($\triangle\phi < \frac{\pi}{2}$)

(c) Composite CIR in multi-speaker system ($\triangle\phi < \frac{\pi}{2}$)

(d) Phase change of dynamic component from multiple speakers ($\triangle\phi > \frac{\pi}{2}$)

(e) Basic CIR from multiple speakers ($\triangle\phi > \frac{\pi}{2}$)

(f) Composite CIR in multi-speaker system ($\triangle\phi > \frac{\pi}{2}$)

Fig. 8. Superposition relation of multi-speaker signals.

Because the superimposed signal may vary due to constructive or destructive combination, we quantitatively examine the probability that the signal becomes stronger when more speakers are included. We observe that this problem can be modeled as the well known *random flights* problem, which was first proposed by Karl Pearson in 1905 [32]. To simplify the derivation, we consider the case of M speakers. Given M random unit vectors which are uniformly distributed on the I-Q plane, we could prove that $P(|\sum_{m=1}^{M} e^{-j2\pi \frac{d_m}{\lambda_c}}| > 1) = \frac{M}{M+1}$. To understand this probability, we simulate the process of signal superposition. When signals from two speakers with the same amplitude and random phase directions are superimposed, the probability of obtaining a stronger signal is 66.7%. We simulate superimposition of different number of signals by repeating the process a large number of times and counting the probability of stronger signal. As shown in Fig. 9, with more speakers, there is a higher chance that the superimposed signal is stronger than the original signal. For a typical six-speaker deployment, the probability that a superimposed signal can be increased to 85.68%.

Verification. To validate the effect of multi-speaker superposition, we conduct experiments to quantitatively compare the multi-speaker system with the single-speaker system. As shown in Fig. 10a, we utilize four speakers to detect human breath in the first 40 seconds, while only one speaker is activated afterwards within time intervals of 40 – 58 seconds, 60 – 83 seconds, 85 – 98 seconds, and 100 – 120 seconds, respectively. In our experiment setup, *Speaker* 1 and *Speaker* 3 are located closer to the sensing target, while *Speaker* 2 and *Speaker* 4 are

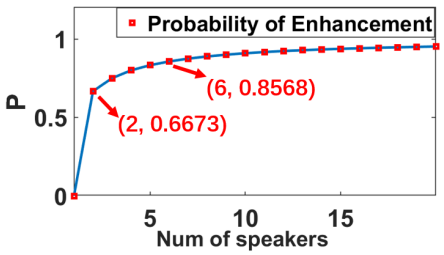
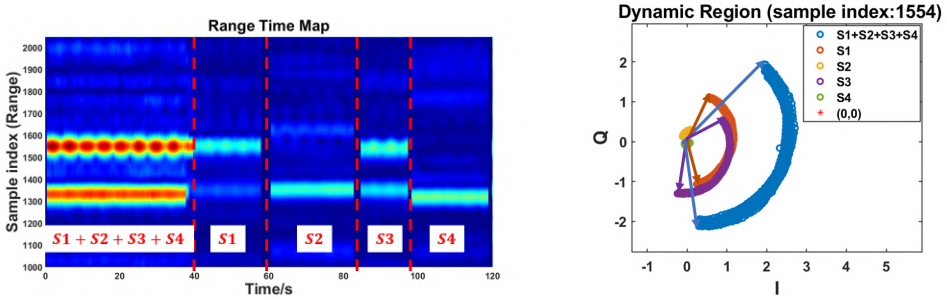


Fig. 9. Probability of energy increase with increasing number of speakers.

placed relatively further away from the target. The Range-Time Map reveals a significant energy increase in the region spanning sample index 1300 – 1400 and 1500 – 1600. By selecting a particular path from each of these two regions, the superposition relationship between the signals can be analyzed on the I-Q plane. We focus on the dynamic component region (e.g., sample index 1554) as it contains the target information. The dynamic vectors (Fig. 10b) of the four speakers exhibit varying energy strengths. We can see that the energy of the composite dynamic vector is significantly increased, making it easier to detect subtle motion.



(a) Comparison of Range-Time Map for multi-speaker system and single-speaker system

(b) Superposition of multi-speaker CIR

Fig. 10. Multi-speaker signal superposition.

3.2 Multi-microphone Combination

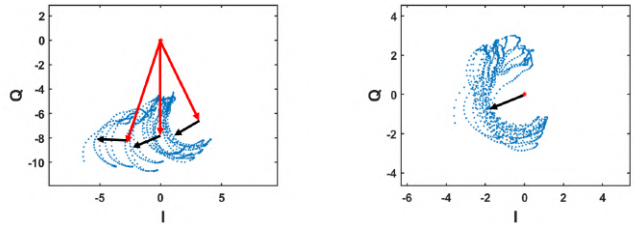
In this subsection, we propose to combine signal from multiple microphones to increase the signal phase change for target sensing. For an in-car distributed multi-microphone system, microphones are usually distributed in three places (as shown in Fig. 6). For each place, two to four microphones are typically embedded. The reflected signals captured by these microphones exhibit similar Range-Time Map and distinct spatial diversity. By combining the same motion information observed from two microphones, sensing performance can be enhanced.

3.2.1 Static Interference Removal. To improve sensing performance, we first need to remove the effect of static reflection in the environment. The small space causes complex signal reflections in a car cabin. For instance, reflections from the seat can be stronger than reflections from the target. Besides, reflections from these static objects can even obscure dynamic reflections from the target, making it difficult to extract the dynamic component required for sensing.

Existing methods for removing static components involve obtaining the average signal over a time window and subtracting it. However, the static component drifts over time and the static component is rendered non-constant in a large time window. On the contrary, utilizing a small time window results in an inaccurate estimation of the static component. To remove the slowly time-varying static components, we employ a

loopback filter for static component removal [15]. The static component of the system is defined as follows: $static\,(t) = \beta \cdot static\,(t-1) + (1-\beta) \cdot cir(t)$. Here, β is the weight coefficient of loopback filter. The dynamic component can be calculated by removing the static component as follows: $dynamic\,(t) = cir\,(t) - static(t)$.

As an illustrative example, we consider the detection of a periodic subtle motions (e.g., breaths). As shown in Fig. 11a, the static component experiences changes within a large time window, while the dynamic arc caused by target motion rotates together with the static vector on the I-Q plane (indicated by the red arrow). We employ a loopback filter to help eliminate the static component in order to obtain a clean dynamic reflection signal. The resultant signal is shown in Fig. 11b, where it shows an arc trace with the static component removed.



(a) Original signal

(b) Signal after static component removal

Fig. 11. Static component removal.

3.2.2 Target Dynamic Signal Enhancement. After removing the effect of static component, we now combine the signals at multiple microphones to enlarge the dynamic signal. We note that there are at least two microphones placed in close proximity in a distributed multi-microphone system. The signals captured by these microphones exhibit similarity, with the signal received at one microphone being a delayed version of the signal at the other microphone. Therefore, we propose to align the CIR received at two microphones by cross-correlating the signals. This alignment ensures that the CIR shares the same dynamic component phase change, enabling us to utilize the information from multiple microphones for sensing. Specifically, we can enlarge the phase change by combining the information extracted from microphones placed at the same location, which further enhances the capability of detecting the subtle motion of a target. Here, we present the method of dynamic signal enhancement in the following three steps.

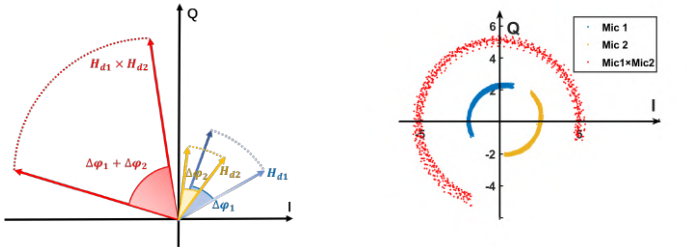
Step 1: CIR Alignment. We first select two frames from the microphone Range-Time Map at the same location. The time-delay offset between these two frames is calculated using cross-correlation, and then they are aligned by compensating the delay. Every time the microphones are turned on, the offset remain a constant until they are turned off and on again. So the offset just needs to be calculated once and this offset is used to align all subsequent data.

Step 2: Static Component Removal. We remove the impact of static components using the proposed method in Sec. 3.2.1.

Step 3: CIR Multiplication. We observe that the dynamic phase shifts in the signal at the aligned microphones exhibit the same pattern (reaching peak/valley at the same time). We can thus multiply these components to increase the signal phase change caused by the target motion. Here, assume that two dynamic components are $H_{d1}\,(n,t)$ and $H_{d2}\,(n,t)$, we obtain the result of signal multiplication H_{CMM} as:

$$\begin{aligned} H_{CMM}\,(n,t) &= H_{d1}\,(n,t) \times H_{d2}\,(n,t) \\ &= A_{d1} e^{-j2\pi \frac{d_1 + \triangle d_1(t)}{\lambda_c}} \times A_{d2} e^{-j2\pi \frac{d_2 + \triangle d_2(t)}{\lambda_c}} \\ &= A_{d1} A_{d2} e^{-j2\pi \frac{d_1 + d_2 + \triangle d_1(t) + \triangle d_2(t)}{\lambda_c}}. \end{aligned} \tag{13}$$

As shown in Fig. 12a, the multiplication of the dynamic components of microphones results in a phase change that is the sum of the dynamic phase changes. This process amplifies the phase change and increases the dynamic energy, thereby reducing the impact of noise on dynamic vectors. The multiplication operation essentially adds the phase changes (i.e., $\Delta\varphi_1(t)$ and $\Delta\varphi_2(t)$) from two microphones, thus virtually increasing the displacement of the target motion. Fig. 12b presents an I-Q plot using data collected from real hardware. The multiplication of the dynamic vectors from *Mic* 1 and *Mic* 2 results in a significant increase of the phase change and the energy of the dynamic component (indicated by the increased radius of the arc).



(a) Multiplication of two dynamic components

(b) Results of the multiplication

Fig. 12. CIR multiplication of two microphones.

4 MULTI-PATH ENHANCED CHILD PRESENCE DETECTION

Our proposed child presence detection scheme is built upon the design of multiple speakers and microphones. Based on the child's state, we divide the detection process into two stages: motion detection and breath detection. For motion detection, we detect an awake child with body motions such as arm/leg movements. This kind of motion usually involves large limb displacement which is easy to be detected. For a child that remains stationary (e.g., sleeping), we need to detect the child by sensing the breath. In this stage, we have some challenging cases, i.e., the child is covered with a blanket or blocked by the rear-facing seat. We deal with these challenges by leveraging rich secondary multi-path reflections to sense child. Next, we introduce the details.

4.1 Motion Detection

As the motion performed by child (e.g., arm/leg motions) introduces significant variations in the propagation channel of acoustic signals compared to when there is no child, it is straightforward to set a threshold for motion detection. When the signal variations are larger than the threshold, a child is detected. However, considering the possible complex environments (e.g., noisy parking lot, etc.), further signal processing is necessary before setting a threshold. To reduce the impact of the environment, we first normalize the CIR signals. After normalization, the CIR signals now have approximately the same energy range in different environments. Then the signal enhancement algorithm introduced in Sec. 3.2 is leveraged to strengthen the dynamic signal induced by the child's motion. With the enhanced dynamic signal, we calculate the signal variance for each path of every microphone within a specific time window (i.e., 2 s for 46 samples) at a step size of 1 s. To fully characterize the motion using multipath and multiple microphones, we sum up the signal variances and the final signal variance is represented as:

$$\delta(t) = \sum_{l=1}^{L} \sum_{c=1}^{C} \delta(l, c, t), \tag{14}$$

where $\delta(l, c, t)$ is the variance for the l-th microphone and c-th path at time t. To select the appropriate threshold, we collect a large number of CIR samples in both empty car and car with child motion. Fig. 13 shows that the signal variances in both scenarios follow a Normal Distribution and there is no overlap between the two distributions. According to the three-sigma rule [42], 99.73% of the data are within three standard deviations from the mean (i.e., between $\mu - 3\sigma$ and $\mu + 3\sigma$). To reduce the false alarm rate and increase the detection rate, we choose five times of the upper three-sigma range as the threshold, i.e., $threshold = 0.08 \times 5 = 0.4$. Fig. 14 shows one example to demonstrate the motion detection result using the proposed approach. We can see that the three motions can all be accurately detected.

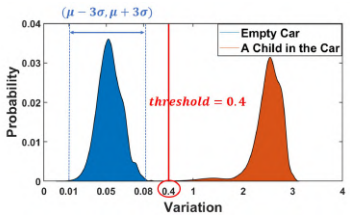


Fig. 13. Distribution of signal variances with and without a child in the car.

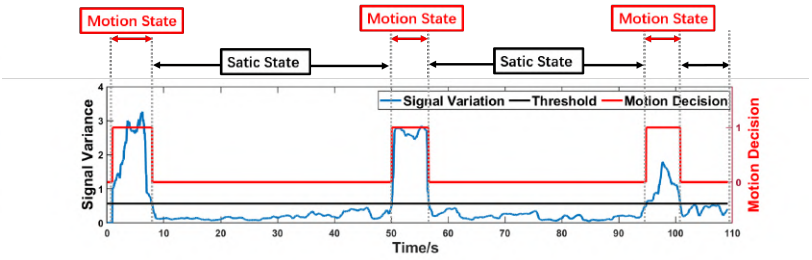


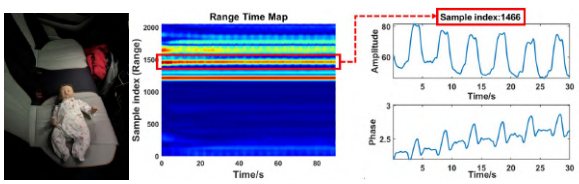
Fig. 14. Motion detection based on signal variance.

4.2 Breath Detection

Compared with relatively large body motions, child's breath is much weaker and there is only subtle chest movement. The information contained in the reflection signal can thus be easily interfered by environmental noise. This is particularly true in challenging scenarios such as the child is blocked by a rear-facing seat. In this section, we first show the case studies for breath detection in different scenarios. Then we present a feasibility analysis of using secondary reflection signals to sense the subtle breath motion. Finally, a multi-path combination algorithm is proposed to address the challenging cases.

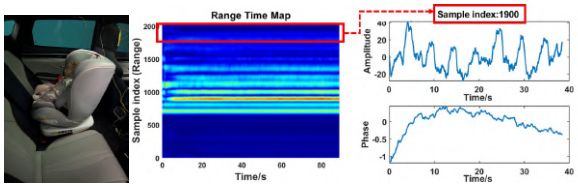
4.2.1 Case Studies for Breath Detection. We identify the presence of a stationary child in a car by sensing his/her breath. We first conduct case studies to sense child breath in two commonly-seen scenarios (i.e., lying on the seat and sitting in the rear-facing seat). As shown in Fig. 15a, when the child is lying on the seat, acoustic signals can directly reach the child's chest, making it easy to detect the child's breath. Fig. 15b shows the detection result based on the Range-Time Map, where multiple paths containing the breathing pattern can be identified. In Fig. 15b, we show the signal amplitude and phase of one path. We can see the periodicity corresponding to inhalation and exhalation of the child's breath. Because the chest movement is periodic, we employ the auto-correlation coefficient to determine whether breathing pattern is present in each path.

In contrast, when the child is sitting in the rear-facing seat (as shown in Fig. 16a), there is no direct path signal to reach the target. The breathing pattern is now difficult to be identified by searching the high auto-correlation coefficients on the Range-Time Map. As shown in Fig. 16b, we plot the signal of a secondary reflection path with relatively large auto-correlation coefficient. We can see that the breathing pattern is noisy. Owing to the small space, we observe richer multi-paths in car cabin and this observation provides us a new opportunity, i.e., besides the primary reflection, we can utilize the secondary reflection for sensing. Our insight is that the child's weak breath signal could be enhanced using secondary reflection paths. We employ the auto-correlation coefficient to select the candidate paths and combine multiple secondary reflection paths for sensing performance enhancement.



(a) Scenario (b) Detection results of good case

Fig. 15. Lying on the seat for breath detection.



(a) Scenario (b) Detection results of bad case

Fig. 16. Siting in the rear-facing seat for breath detection.

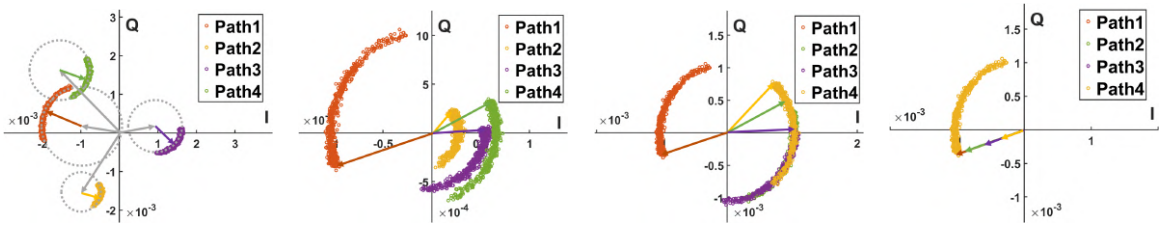
4.2.2 Feasibility Analysis of Multi-path Signal Combination. The child's breath is detected through sensing the subtle chest movement. The signal channel impulse response consists of static component, dynamic component and ambient noise. Eq.(12) can be rewritten as follows:

$$cir_c(n,t) = \underbrace{A_{nSTA}e^{-j2\pi\frac{d_n}{\lambda_c}}}_{Static\ component} + \underbrace{A_{nMOV}e^{-j2\pi\frac{d_n+\Delta d_n(t)}{\lambda_c}}}_{Dynamic\ component} + noise(n,t), \tag{15}$$

where $A_{nSTA} = \sum_{m=1}^{M} A_{mSTA} sinc\left(n - \frac{d_m}{c \cdot f_s}\right)$, $A_{nMOV} = \sum_{m=1}^{M} A_{mMOV} e^{-j2\pi\frac{d_m+\Delta d_m(t)}{\lambda_c}} sinc\left(n - \frac{d_m+\Delta d_m(t)}{c \cdot f_s}\right)$, d_n indicates the length of the path corresponding to range bin n, $\Delta d_n(t)$ denotes the path length change due to the subtle chest movement, and $noise(n,t)$ denotes the ambient noise.

Static Component: The path length of each individual multi-path exhibits difference, resulting in distinct static components, as shown in Fig. 17a. The static component has a constant magnitude and phase over a period of time. To remove the static interference, we apply a loopback filter to the signals of candidate paths as introduced in Sec. 3.2.1.

Dynamic Component: The phase shift $\varphi_n(t) = 2\pi\frac{d_n(t)}{\lambda_c}$ caused by breath motion exhibits similar pattern across different multi-paths. However, the dynamic vector's initial phase $\varphi_n = 2\pi\frac{d_n}{\lambda_c}$ and the signal phase change $\triangle\varphi_n(t) = 2\pi\frac{\triangle d_n(t)}{\lambda_c}$ are different among different paths as shown in Fig. 17b. To effectively combine the phase change of multiple paths, it is essential to normalize the phase change and also align the starting phase (Fig. 17d).



(a) Original signals (b) Signals after static removal (c) Signals after phase normalization (d) Signals after phase alignment

Fig. 17. Key steps in multi-path signal combination.

Ambient Noise: Previous studies [49] demonstrated that the ambient noise, denoted as $noise(n,t)$, exhibits the properly of isotropy and follows a random distribution with a mean of zero and a variance of σ^2. Based on our experiments, the ambient noise follows a normal distribution $N(\mu,\sigma^2)$. Combining the ambient noise of n multi-path signals results in the average noise, denoted as $\overline{noise}$ ($\frac{1}{n}\sum_{i=1}^{n} noise_i$), which also follows a normal

distribution with a mean of μ and a variance of $\frac{\sigma^2}{n}$. This suggests that as the number of paths increases, the variance of the ambient noise decreases.

4.2.3 Multi-path Combination Algorithm. Based on the feasibility analysis, we can combine the phase information from multiple dynamic paths to enhance the desired signal variation, making it possible to extract subtle breathing information. The proposed *Multi-path Combination Algorithm* consists of the following steps:

(1) Phase normalization: normalizing the phase change and strength of each dynamic multi-path signal.

(2) Phase alignment: selecting one multi-path signal as the reference and align the other signals' starting phase to that of the reference signal.

(3) Multi-path superposition: superposing secondary reflection signals to enhance the signal variation.

The details of each step are described as below.

Phase Normalization. We have obtained enlarged dynamic components after CIR multiplication in the motion detection stage. We now employ $H_{CMM}(n,t)$ within a time window of T' for normalization, which can be represented as follows:

$$H_{norm}(n,t) = \frac{H_{CMM}(n,t)}{MEAN\left(\left[H_{CMM}(n,t), H_{CMM}(n,t+T')\right]\right)}. \tag{16}$$

The results are shown in Figure 17c, where all dynamic arcs are normalized to the same size, i.e., they have the same amount of phase change and the same energy.

Phase Alignment. After the phase has been normalized, we need to align the signals of different paths to facilitate combination. Due to noise, data samples on one phase arc are slight different from samples on other arcs and the corresponding samples are unable to be perfectly matched. To align the two arcs, we select one of the arcs as a reference, and iterate through all possible angles to identify the optimal angle that minimizes the sum of distances between all corresponding data samples in the two arcs:

$$\arg\min_{\theta_j} Dis(i,j) = \sum_{t=1}^{T} \left| H_{norm}(i,t) - H_{norm}(j,t)\, e^{j\theta_j} \right|, \tag{17}$$

where $Dis(i,j)$ is the sum of the distances between all corresponding data samples in the two signals i and j, and θ_j is the rotation angle. Once we decide the optimal rotation angle for each multi-path signal j, the rotated dynamic vector can be represented as:

$$H_{rot}(n,t) = H_{norm}(n,t)\, e^{j\theta_j} = A_{MOV} \cdot e^{-j\varphi_{Ini}} \cdot e^{-j2\pi\frac{\triangle d(t)}{\lambda_c}} + noise'(n,t), \tag{18}$$

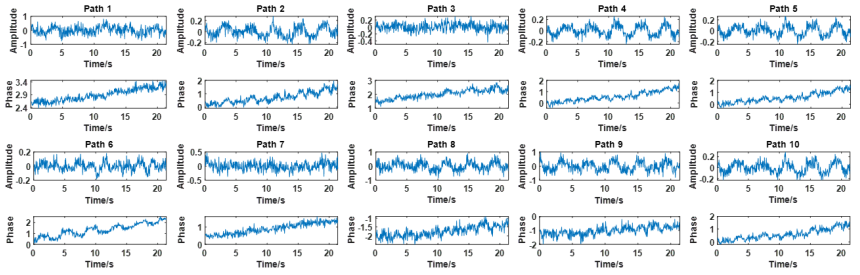
where A_{MOV} and $2\pi\frac{\triangle d(t)}{\lambda_c}$ denote the normalized dynamic energy and phase change for all candidate paths after the phase normalization operation, $e^{-j\varphi_{Ini}}$ indicates the common initial phase, where the dynamic arcs are aligned to the same position as shown in Fig. 17d, $noise'(n,t)$ represents the ambient noise after the normalization and alignment, which remains an isotropic random variable and still conforms to normal distribution with a zero mean.

Multi-path Superposition. After normalization and alignment, we can now combine the multi-path signals to create a new signal:

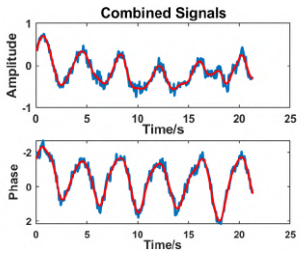
$$\begin{aligned} cir_e(t) &= \frac{1}{N}\sum_{i=1}^{N} H_{rot}(n,t) \\ &= \underbrace{\frac{1}{N}\sum_{i=1}^{N} A_{MOV}\cdot e^{-j\varphi_{Ini}}\cdot e^{-j2\pi\frac{\Delta d(t)}{\lambda_c}}}_{Dynamic\ component} + \underbrace{\frac{1}{N}\sum_{i=1}^{N} noise'(n,t)}_{Noise\ component}, \end{aligned} \tag{19}$$

where N indicates the number of candidate multi-path signals. $cir_e(t)$ is the enhanced CIR, which consists of two elements: the dynamic component and the noise component. The target's motion information is contained in the dynamic component and is denoted as $\Delta d(t)$. For noise component, the combination of N multi-path signals reduces the noise variance.

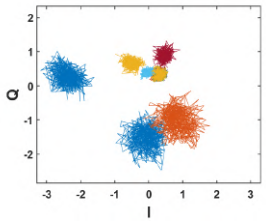
Verification. We deploy our distributed acoustic sensing system in a car, and the infant simulator is placed in the rear-facing seat. Fig. 18a shows ten candidate secondary multi-path reflection signals, which are noisy and difficult to see clear breathing pattern. We further plot the ten multi-path signals on the I-Q plane in Fig. 18c. We can see that their positions on the I-Q plane are different due to the difference of the static vector and the initial phase. As a contrast, the combined signal after being processed by our algorithm shows clear breathing pattern in Fig. 18b. Fig. 18d shows the distribution of the ten multi-path signals after phase normalization and alignment. The superposition of the aligned signals is shown in Fig. 18e. The blue line represents the result of



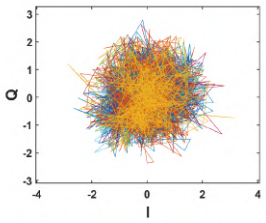
(a) Raw amplitude and phase of ten multi-path signals of breathing



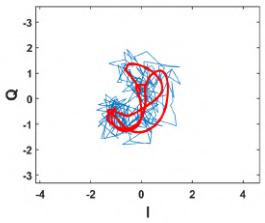
(b) Signal after applying the combination algorithm



(c) Ten multi-path signals on the I-Q plane



(d) Ten multi-path signals after phase normalization and alignment



(e) Ten multi-path signals after superposition

Fig. 18. Illustration of multi-path signals and the combination result.

the original superposition and the red line represents the result after signal filtering. The arc pattern on the I-Q plane corresponds to the breathing motion. After combining the dynamic vector of various secondary reflection multi-paths, the strength of the combined signal and the amount of phase change are increased, which lay the foundation for the proposed system to capture weak signal variations and accordingly sense subtle motions.

5 SYSTEM IMPLEMENTATION

Fig. 19 presents the overview of the CPD system implementation, which includes three key modules: data acquisition and preprocessing, motion detection and breath detection. The microphones collect acoustic signals from the distributed speakers and CIRs are extracted to perform motion detection. Once motion is detected, the alarm is raised. Otherwise, breath detection module is activated to further check whether a stationary child is present. If there is no child's breath, the system outputs "Child Not Detected". A video demo is available at: https://youtu.be/dz_eXa41swo. We present the detailed implementation as follows.

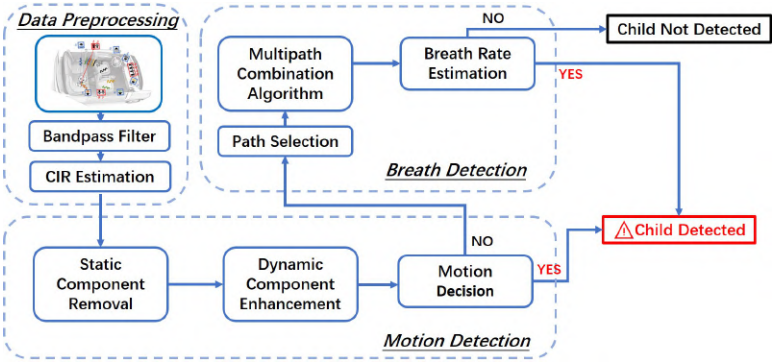


Fig. 19. Overview of the proposed distributed acoustic sensing system for child presence detection.

Data acquisition and preprocessing: When the proposed CPD system is enabled, distributed speakers in the car cabin continuously play the signals at a rate of 23 ZC-sequences per second, and the microphones record the signals. Then the laptop configured with an Intel i5-9300H processor and 16 GB memory processes the latest 23 sequences (i.e., data collected within one second) from each microphone in real time. To minimize the influence of music and ambient low-frequency noise, we utilize a bandpass filter to retain signals within the frequency range of 18 kHz to 20 kHz (the central frequency of the transmitted signal is 19 kHz). Then cross-correlation operation between the received acoustic signals and the transmitted signal is performed to estimate CIRs.

Motion Detection: A sliding window size of 5 s with a step size of 1 s is set to detect the child's motion in our system. For each window, we first normalize CIR signals with the average amplitude of the CIR signals in the window. A loopback filter is used to eliminate the impact of the static component. Then the dynamic component enhancement algorithm introduced in Section 3.2 is employed to enlarge the signal variation induced by the child's motion. When a motion is detected, the system sends the alarm of "Child Detected" to the user.

Breath Detection: A smaller window size will result in higher false alarm rate for breath detection. For breath detection, we choose a sliding window of 20 s with a step size of 1 s to balance false alarm rate and latency. We calculate the auto-correlation coefficient for every path in each window to choose the candidate paths whose auto-correlation coefficients are higher than a predetermined empirical threshold. The *Multi-path Combination Algorithm* chooses the candidate paths with high auto-correlation coefficients for combination. Then we estimate the breath rate of the enhanced signal to determine whether there is a child. If the child's breath rate is within

the range of 12 BPM~40 BPM [9], the alarm "Child Detected" is triggered and the breath rate is also reported which may help determine the child's health condition.

6 EVALUATION

We implement the child presence detection system on real in-car acoustic modules and conduct comprehensive experiments to evaluate the system performance.

6.1 Experiment Setup

We implement our system on the audio modules embedded in a real car,[4] which consists of six Bose speakers [11] and eight microphones. As shown in Fig. 20, for a typical deployment, the six speakers are located in the panel under the windshield (2x), A-pillar (2x), and B-pillar (2x). The eight microphones are located at three positions. A four-element linear microphone array ($Mic\ 3-6$) with a spacing of 1 cm is located at the top of the rearview mirror position. The other microphones ($Mic\ 1-2$ and $7-8$) are located at B-pillar near the speaker. In order to send the ZC-sequence signals, the speakers are connected to a sound card (PXUA216MB-DL2-M/XMOS) to play the sound data. The sound card is connected to a computer to receive and process the collected acoustic signals. The ZC-sequence signals are generated with parameters $f_c = 19\ kHz$, $B = 2\ kHz$, and $T = 0.0427s$. The receiver employs a sampling rate of 48 kHz and processes the signals in real time to detect child presence. We conduct experiments with two infant simulators (SimNewB newborn female and male simulators) [13], which have a preset breathing rate. This high-fidelity simulator with realistic newborn traits provides lifelike clinical feedback. We evaluate the overall child presence detection performance using *detection rate* and *false alarm rate* as the metrics, while we also evaluate the accuracy of child breath rate estimation using mean absolute error.

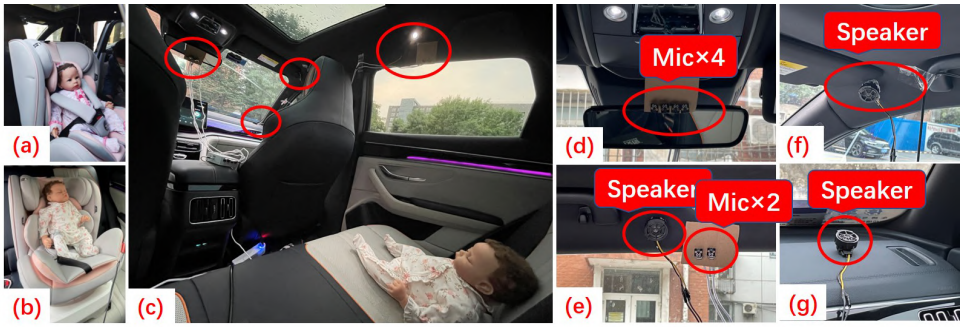


Fig. 20. Experiment setup. (a, b) show the infant simulators we use. (c) show the deployment locations of the speakers/microphones, and (d, e, f, g) zoom in to show the acoustic modules.

Data collection: we collect data with child present and absent in the car to evaluate the detection rate and false alarm rate, respectively. When the car is powered off and the door is locked, the distributed acoustic sensing modules are started and each monitoring section lasts for 1-2 minutes. We record a total amount of 130 hours data in different types of vehicles (including SUV and Sedan) and under varying scenarios (e.g., street parking lot, underground parking lot, etc.).

[4]We acquired permission from a major car manufacturer (the brand name is not disclosed due to a non-disclosure agreement) to conduct experiments using their car's built-in speakers and microphones. To conduct experiments in other car models, the same audio system is taken out and deployed in other cars.

6.2 Real-time Motion and Breath Detection

To demonstrate the effectiveness of real-time child presence detection, the infant simulator is set to one of the two states: body motion state and stationary state. We collect the acoustic signals for 10 minutes, in which the child's continuous body motion (e.g., leg movement) occurred several times in two time periods. Note that the baby simulator itself can not move the arms or legs, we thus attach a motor to the child's legs and use the remote controller to trigger the movement of the child's leg with the help of the motor. For the stationary state, the default breath rate of the infant simulator is 14 breaths per minute. Fig. 21 shows the detection results of body motion and breath. Six and five body movements can be detected in the two time intervals, i.e., 16 – 222 seconds and 397 – 548 seconds. Under the stationary state, the breath rate can be accurately measured, matching the ground truth very well with a low error rate of 0.13 BPM.

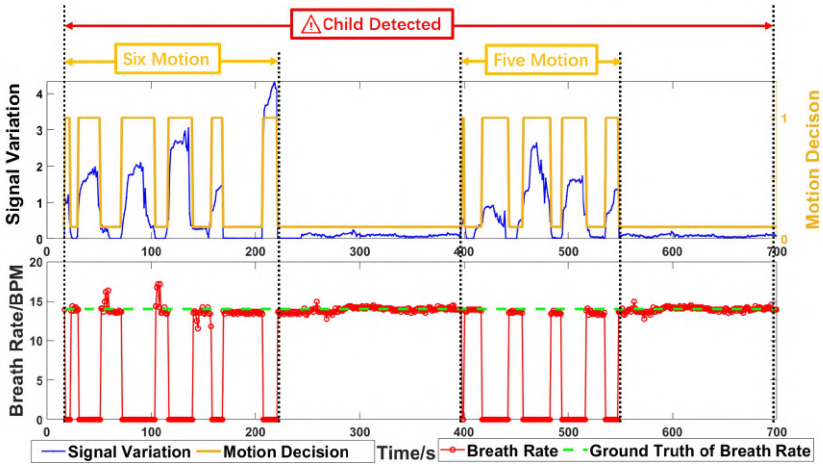


Fig. 21. 10-minute results of real-time motion and breath detection.

6.3 Child Presence Detection at Different Locations in the Car Cabin

The European New Car Assessment Programme (NCAP) specifies the coverage area for child presence detection, including the seat and other possible areas a child may hide. To evaluate the coverage performance of the proposed system, we consider ten different locations inside the car, as shown in Fig. 22. Specifically, the locations include the car seats (i.e., #3, #4, #8, #9, #10), and other areas (i.e., #1, #2, #5, #6, #7) because a child may fall off the seat. At each location, we collect 10 groups of data, each lasting for 2 minutes. As shown in Fig. 23a, the average

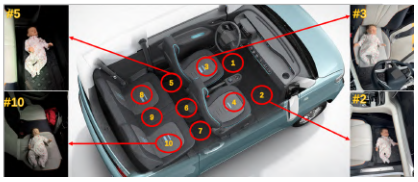
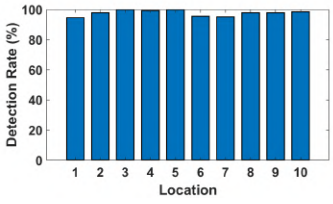
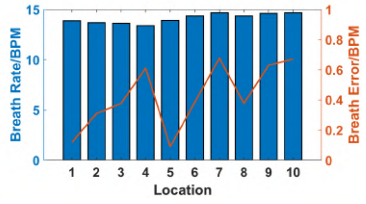


Fig. 22. Child placed at ten different locations inside the car.



(a) Detection rate at different locations



(b) Breath rate error at different locations

Fig. 23. Results of child presence detection at different locations inside the car.

child presence detection rate is 97.69% across the 10 locations. Among them, location #1 has a relatively low detection rate of 94.60% compared to the other locations due to the fact that the driver' footwell zone is blocked by the seat and the steering wheel. Fig. 23b shows that the error of breath rate estimation is below 0.7 BPM at all 10 locations. These results show that the distributed acoustic sensing system can detect a child in the car cabin without blind spots.

6.4 Impact of Car Model

As the size and interior structure vary across car models, we evaluate the sensing performance in different cars, including a BYD SUV (Car #1), a Volkswagen Tacoua SUV (Car #2), a Nissan Sylphy (Car #3), a Tesla Model S (Car #4), and a Mercedes-Benz C200 (Car #5), as shown in Fig. 24. Fig. 25 shows the detection rate of our system in five different cars. We can see that the highest detection rate (98.03%) is achieved in Car #1 and the lowest detection rate (95.19%) is achieved in Car #2. The differences in size and structure result in different multi-path distributions and accordingly slightly varying detection performance.

Fig. 24. The proposed distributed acoustic sensing system is deployed in different cars.

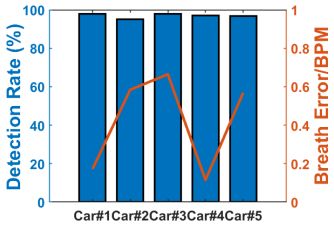


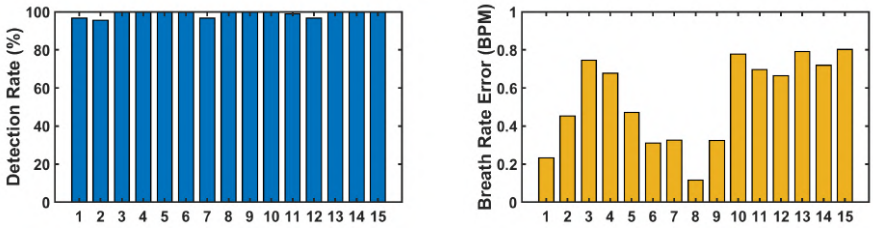
Fig. 25. Detection result in different cars.

6.5 Impact of Seating Capacity

We further conduct experiments to evaluate the performance of the proposed system in a 7-seater vehicle with a large cabin space. We evaluate the performance of the proposed system in a 7-seater Buick GL8 as shown in Figure 26a, which has eight Bose speakers and twelve microphones. As shown in Figure 26b, the eight speakers are located in the panel under the windshield (2x), A-pillar (2x), B-pillar (2x) and C-pillar (2x). The twelve microphones are located at five positions. We choose 15 locations covering all possible child presence locations in the car as shown in Figure 26c to evaluate the performance of child presence detection. The experiment result is shown in Figure 27. At these 15 locations, our system achieves an average child detection rate of 98.99% and an average breath rate estimation error of 0.54 BPM. Owing to a larger space with less obstructions and two additional speakers, we actually achieve a higher child presence detection rate in this bigger 7-seater car.

(a) Exterior of Buick GL8 (b) Deployment of speakers and microphones (c) Child at 15 different locations

Fig. 26. Deploy our CPD system in the 7-seater car (Buick GL8).



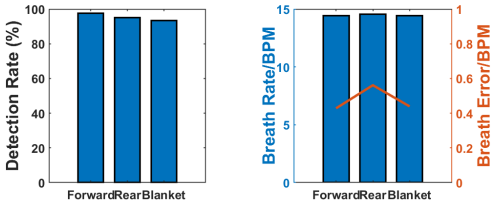
(a) Detection rate of child presence (b) Breath rate estimation error

Fig. 27. Performance of child presence detection in the 7-seater car.

6.6 Impact of NLoS Scenarios

In real-life scenarios, the rear-facing baby seat can block the direct path of the speaker to reach the child. In addition, the child may also be covered with blankets while sleeping. In this section, we evaluate the detection accuracy under challenging non-line-of-sight (NLoS) scenarios. As shown in Fig. 28, we consider both forward-facing and rear-facing seats,[5] as well as blanket covering. Fig. 29 shows the detection results under the three scenarios. We can see that while an accuracy of 97.62% can be achieved for forward-facing baby eat, slightly lower accuracy is achieved for rear-facing (95.14%) and blanket covering (93.38%). The error of breath rate estimation for the three cases are all below 0.6 BPM. These results demonstrate that even when there is no direct path to reach the target, the proposed system can still accurately detect child presence through secondary reflections.

Fig. 28. NLoS challenging scenarios: forward-facing seat, rear-facing seat and blanket covering.



(a) Detection results at different scenarios (b) Breath rate at different scenarios

Fig. 29. Results of child presence detection in three NLoS challenging scenarios.

6.7 Impact of Interference

To evaluate the performance of distributed acoustic sensing in the presence of interference, we consider three types of interference, i.e., background sound noises (e.g., human voice, traffic noise and music), surrounding motion noise (human walking, roadside vehicle passing by, tapping on the window) and weather noise (e.g., rain drops falling on the car). To flexibly control the sound noise, we record human voice, traffic noise and music. We then play them through a smartphone at a sound pressure level of 50 dB in the cabin. For motion interference, we park the car in a roadside parking lot and let another vehicle pass by. We also ask volunteers to tap on the car window to create motion interference. We also park the car in an open lot when it is raining. Fig. 30 shows our system maintains high accuracy in all the interference scenarios, achieving a detection rate of 100%, 94.24%,

[5]Note that children under 4 years old usually use rear-facing seats. Rear-facing baby seats provide better protection to child's head, neck and back.

99.26%, 95.32%, 99.62%, 97.84%, and 93.62% respectively. The error in breath rate sensing is always smaller than 0.7 BPM, demonstrating the system's robustness against interference.

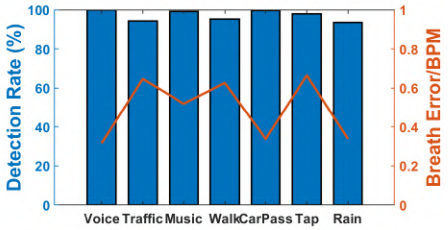


Fig. 30. Impact of various interference.

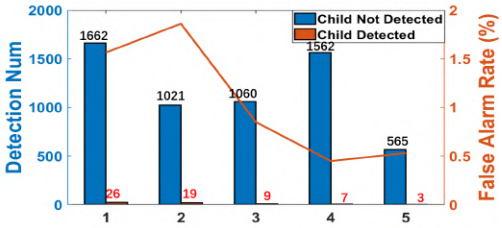
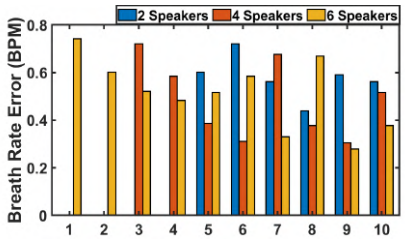


Fig. 31. False alarm rate in different environments.

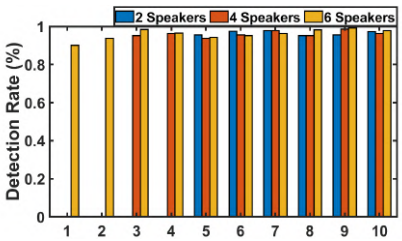
We further conduct experiments without a baby in the car to study the false alarm rate of our system. We park the car at five different locations, including an underground parking lot (Scenario 1), a surface parking lot (Scenario 2), parking on a street (Scenario 3), parking next to a construction site (Scenario 4), and parking next to a subway station (Scenario 5). We collect acoustic signals for one minute and run the proposed system for child presence detection. We collect 98 hours data in total in the five environments at different times throughout the day. As shown in Fig. 31, the false alarm rates are 1.56%, 1.86%, 0.85%, 0.45% and 0.53%, respectively. To further reduce the false alarm rate, we can run the proposed system multiple times consecutively.

6.8 Impact of Number of Speakers

The proposed sensing system is implemented on the audio system of a commercial car without hardware modification. Six speakers are available with this audio system. We further conduct experiments to evaluate the detection performance with smaller numbers of speakers. The results are shown in Fig. 32. When only 2 speakers located on the B-pillars in the rear are employed, the front zone becomes blind spots, i.e., child's presence can not be detected at locations #1, #2, #3, and #4, as shown in Figure 22. Subsequently, with 4 speakers deployed including two additional speakers located on the A-pillars in the front as shown in Fig. 20f, besides location 5-10, child's presence can also be detected at location #3 and location #4 now. However, location #1 and location #2 remain blind spots for detection. When six speakers are deployed as shown in Fig. 20c, reliable detection at all 10 locations can be achieved. The above experiment shows that the typical deployment of 6 speakers in car cabin meet the requirement for full-coverage sensing.



(a) Detection rate under different number of speakers at 10 locations

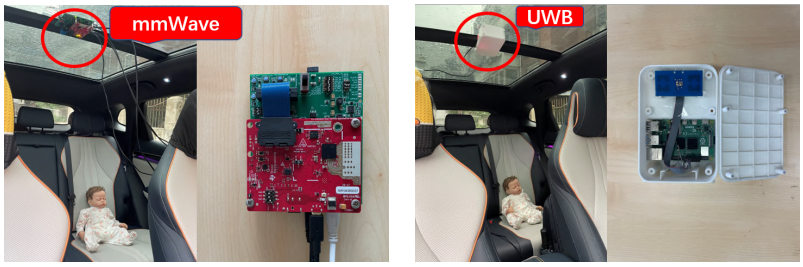


(b) Breath rate estimation error under different number of speakers at 10 locations

Fig. 32. Comparison of system performance with 2, 4 and 6 speakers deployed in the car.

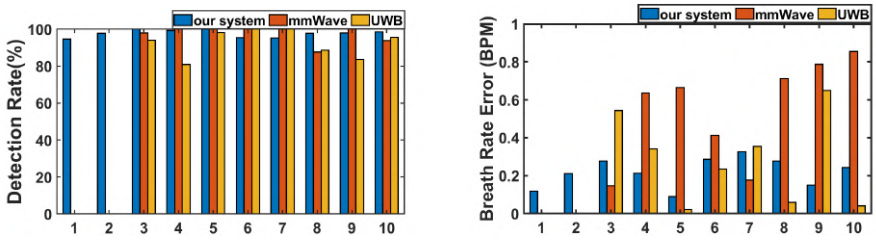
6.9 Comparison with Other Sensing Technologies

We now conduct experiments to compare the sensing performance between the proposed distributed acoustic sensing system and dedicated millimeter wave (mmWave) and ultra-wideband (UWB) based sensing systems. For the mmWave-based system, we employ the TI FMCW mmWave IWR1443 radar [1] operating at 77-81 GHz, which costs around 900$. The radar has three transmission (TX) antennas and four reception (RX) antennas. We implement the IR-UWB radar-based sensing system using XETHRU radar module X4M05 [14]. The radar has a bandwidth of 1.4 GHz centered at the carrier frequency of 7.3 GHz. The radar board is connected to a Raspberry Pi via Serial Peripheral Interface (SPI). As shown in Fig. 33, both radar devices are mounted on the ceiling of the car to cover the sensing area as much as possible. The infant simulator is placed at ten different locations as shown in Fig. 22. Figure 34a shows that our system can cover all the locations in the car (i.e., full coverage), demonstrating superior performance with high detection rates (>95%) and low breath rate estimation error (<0.4 BPM). In comparison, the mmWave radar and UWB radar based methods fail to detect the presence of child at location #1 and #2, the footwell zones of the driver and front passenger. We also notice that due to the small wavelength, mmWave radar is sensitive to interference in the environment, which leads to high errors in breath rate estimation in Figure 34b.



(a) MmWave device and deployment (b) UWB device and deployment

Fig. 33. Deployments of RF-based dedicated devices.



(a) Detection rate of different technologies at 10 locations

(b) Breath rate estimation error of different technologies at 10 locations

Fig. 34. Comparison with RF-based solutions at 10 different locations.

7 RELATED WORK

Acoustic sensing. Acoustic sensing enables a large range of applications ranging from vital sign monitoring [37, 45, 51, 53], gesture recognition [18, 34, 55] to gait recognition [35, 54], and location tracking [20, 38, 48, 52].

In the smart home scenario, C-FMCW [53] realizes human respiration sensing using a commodity speaker. Zhang et al. [64] further extract subtle heartbeat information using a smart speaker, separating heartbeat information from respiration with empirical modal decomposition. Acousticcardiogram [43] eliminates direct power leakage using two microphones available on a smartphone to achieve respiration and heartbeat monitoring. However, the achieved accuracy is not fine enough and the working range is limited. Also on a smartphone, VSkin [47] employs both airborne and structure-borne sounds to detect finger tapping and movements. To identify gait, AcousticID [54] captures the Doppler information of various body parts while walking, and then extracts fine-grained gait features for user authentication. Existing research on in-cabin sensing mainly focuses on driver sensing. SymListener [59] places a smartphone on the dashboard facing the driver to monitor the driver's vital sign. BlinkListener [36] utilizes a smartphone to sense the driver's eye blinking activity for fatigue driving detection. Smartphone and smart speaker-based acoustic sensing has a limited sensing coverage. The proposed system employs distributed in-cabin speakers to significantly improve the sensing coverage. We believe the proposed distributed sensing modality can enable a large range of new acoustic sensing applications in car cabin.

Child presence detection. Child presence detection will become mandatory in a large range of new cars [7, 10]. Existing methods rely on various types of sensors such as heat and pressure sensors mounted on the seats for detection [17, 27]. When the child is in the safety seat, the performance of these sensors degrades because they are not in direct contact with the child and the weight of child is small. Camera-based solutions [22, 60] are commonly used for fatigue driving detection. However, for child presence detection, one single camera can not cover all the area and blind spots easily occur. Besides, camera-based solutions are susceptible to the lighting condition and also raise privacy concern. RF-based solutions include the use of UWB and millimeter wave radar [4, 25, 56] for sensing. However, a single radar device still faces the issue of dead spots and the use of multiple devices results in high cost. Wi-Fi based CPD solutions [29, 61, 62] are also proposed. However, multiple Wi-Fi transmitters and receivers are required which are not available in today's vehicles. The latest acoustic solution, Vecare [66], deploys two speakers and one microphone in car to detect child's body motion and breath. However, this system does not consider the practical deployment of microphones and speakers in modern car cabin. In this paper, we propose the concept of distributed acoustic sensing and exploit existing speakers and microphones in the car for CPD.

8 DISCUSSION AND FUTURE WORK

Lessons learned. We learned a lot of lessons during the process of deploying our system in real car cabins. We briefly discuss two here. (i) Rear-facing baby seat makes child presence detection much more challenging. To best protect the baby's delicate head and neck, the rear-facing baby seat is designed to have a thick steady back which significantly attenuates or even totally blocks the direct-path acoustic signal. In this case, only weak secondary-reflection signals can be utilized for sensing. This is also the motivation why we need to fuse signals from multiple distributed speakers to strengthen the signal to the level which is capable of being used for sensing. (ii) Different car models/brands have completely different speaker/microphone placement strategies. This makes training-based approaches difficult to generalize to different car models. This motivates us to rely on signal processing techniques to achieve full-coverage CPD in car cabins.

Computational time. The computational time, i.e., the time duration takes for our signal processing algorithm to complete is around 0.7 s. To achieve reliable breath detection, we collect signal samples corresponding to 2-3 cycles of breath. In our implementation, we collect 400 samples corresponding to a duration of 17 s to detect breath. The total time duration taken is thus 17.7 s starting from data collection. According to a recent research [28], leaving a child in the car alone for over 15 minutes can be dangerous. As a result, our system satisfies the requirements for timely detecting the presence of child.

Energy consumption. It is known that the computational power consumption is much lower than the transmission/reception power [40]. The power of car speakers is usually in the range of 14 to 20 watts [44]. For six speakers, the power consumption is around 0.12 kW. Compared to 20 kW power consumption (at a speed of 60 mph) of an electrical car, the power consumption of six speakers is negligible. Based on the 2023 EV Database [23], electric vehicles have an average battery capacity of about 68.7 kilowatt-hours which can support around 572 hours of continuous audio emission. As a result, energy consumption of the proposed child presence detection system accounts for a very small portion of the energy used by the car.

Multiple children detection. For motion detection, as long as one child has motions, the proposed system can detect it and trigger an alert. For multiple children in the stationary state, signal processing techniques such as Variational Mode Decomposition (VMD) [24] can be applied to separate the breath information from multiple targets.

Limitation. One limitation of our signal processing-based method is that the detection accuracy can be affected by car motions. If the parents leave the car with the engine on, the vibration of engine can interfere with breath sensing as engine also causes signal variations. In this case, deep learning methods may be able to capture the patterns buried by interference for sensing.

Future directions. While our work mainly focuses on child presence detection, the novel distributed acoustic sensing modality can be applied to enable a broader range of in-car applications. For example, we can sense the driver's vital signs together with other activities such as eye blinks and hand movements to achieve robust fatigue driving detection. Furthermore, by utilizing distributed acoustic modules, it is possible to sense the driver and passengers at the same time.

9 CONCLUSIONS

In this paper, we present distributed acoustic sensing to detect the child's presence in a contact-free manner. We mathematically model the relationship between superimposed signals from multiple distributed speakers and target motion. To address the blockage issue, we leverage the rich secondary multipath reflections in car cabin to enhance subtle motion detection. We implement the proposed child presence detection system on existing audio system of a car to evaluate the sensing performance in real-world settings. Comprehensive experiments demonstrate that the proposed system is capable of achieving full-coverage sensing and is robust against blockage and interference. We believe distributed acoustic sensing in car cabin is a powerful new sensing modality which can enable a large range of applications.

ACKNOWLEDGMENTS

This work is partially supported by the Beijing Natural Science Foundation (L223034), the Beijing Nova Program, the National Natural Science Foundation of China (No.62172394, No. 62072450), the National Natural Science Foundation of China A3 Foresight Program (No.62061146001), the Youth Innovation Promotion Association, Chinese Academy of Sciences (No. 2020109).

REFERENCES

[1] 2018. IWR1443 Single-Chip 76- to 81-GHz mmWave Sensor. https://www.ti.com/product/IWR1443.

[2] 2019. Auto Industry Agrees to Put Rear-Seat Reminder Systems in Most New Cars by 2025. https://www.consumerreports.org/car-safety/auto-industry-agrees-to-put-rear-seat-reminder-systems-in-most-new-cars-by-2025/.

[3] 2020. *AS TEMPERATURES REACH DEADLY LEVELS, SAFETY ADVOCATES JOIN TOGETHER TO WARN ABOUT DANGERS OF CHILDREN DYING IN HOT CARS.* https://www.safekids.org/press-release/temperatures-reach-deadly-levels-safety-advocates-join-together-warn-about-dangers.

[4] 2021. Infineon. Infineon in-cabin monitoring. https://www.infineon.com/cms/en/tools/aurix-embedded-sw/AURIX-Applications-software/in-cabin-monitoring/.

[5] 2022. Audio Systems for the Tesla Model S and Model X. https://teslatap.com/articles/audio-systems-for-the-tesla-model-s-and-model-x/.

[6] 2022. How you can help prevent hot car deaths as temperatures continue to rise. https://www.wafb.com/2023/06/14/how-you-can-help-prevent-hot-car-deaths-temperatures-continue-rise/.

[7] 2022. TEST AND ASSESSMENT PROTOCOL – CHILD PRESENCE DETECTION. https://cdn.euroncap.com/media/75474/euro-ncap-cpd-test-and-assessment-protocol-v11.pdf.

[8] 2022. Texas Instruments. Vehicle occupant detection reference design. https://www.ti.com/lit/ug/tidue95a/tidue95a.pdf.

[9] 2022. What You Need to Know About Infant and Children's Vital Signs. https://health.clevelandclinic.org/pediatric-vital-signs/.

[10] 2023. Global child presence detection system market grows tremendously at a CAGR. https://straitsresearch.com/press-release/global-child-presence-detection-system-market-analysis.

[11] 2023. *in-cabin-experiences*. https://automotive.bose.com/road-ahead/in-cabin-experiences

[12] 2023. Mercedes-Benz S-Class Burmester Sound System: Are 31 Speakers Enough? https://thedailymotor.com/sound-systems/2021-mercedes-benz-s-class-burmester-sound-system-are-31-speakers-enough/.

[13] 2023. SimNewB. https://laerdal.com/us/products/simulation-training/obstetrics-pediatrics/simnewb/.

[14] 2023. Xethru impulse uwb radar chip. Xethruimpulseuwbradarchip,https://www.xethru.com/x4-radarchip.html/.

[15] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 2014. 3D tracking via body radio reflections. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. 317–329.

[16] AKG. 2021. 2021 Cadillac Escalade with AKG Studio Reference System. https://www.akg.com/ForCadillac.html.

[17] Joan Albesa and Manel Gasulla. 2015. Occupancy and Belt Detection in Removable Vehicle Seats Via Inductive Power Transmission. *IEEE Transactions on Vehicular Technology* 64, 8 (2015), 3392–3401. https://doi.org/10.1109/TVT.2014.2356443

[18] Takashi Amesaka, Hiroki Watanabe, Masanori Sugimoto, and Buntarou Shizuki. 2022. Gesture Recognition Method Using Acoustic Sensing on Usual Garment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 41 (jul 2022), 27 pages. https://doi.org/10.1145/3534579

[19] Haibin Cai, Donghee Lee, Hwang Joonkoo, Yinfeng Fang, Song Li, and Honghai Liu. 2017. Embedded vision based automotive interior intrusion detection system. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2909–2914. https://doi.org/10.1109/SMC.2017.8123069

[20] Shirui Cao, Dong Li, Sunghoon Ivan Lee, and Jie Xiong. 2023. *PowerPhone: Unleashing the Acoustic Sensing Capability of Smartphones*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3570361.3613270

[21] CodeTD. 2019. Market car stereo microphone placement and noise reduction solutions. https://www.codetd.com/en/article/7713766.

[22] R.C. Coetzer and G.P. Hancke. 2011. Eye detection for a real-time vehicle driver fatigue monitoring system. In *2011 IEEE Intelligent Vehicles Symposium (IV)*. 66–71. https://doi.org/10.1109/IVS.2011.5940406

[23] Electric Vehicle Database. 2023. Useable battery capacity of full electric vehicles. https://ev-database.org/cheatsheet/useable-battery-capacity-electric-car.

[24] Konstantin Dragomiretskiy, Dominique Zosso, Luminita Vese, and Andrea Bertozzi. [n. d.]. Variational Mode Decomposition. ([n. d.]).

[25] Gert Freiberger and Helmut Schreiber. 2022. Modelling Child Life Presence Detection with Ultra-Wideband Radars for an Automotive Environment. In *2022 International Conference on Broadband Communications for Next Generation Networks and Multimedia Applications (CoBCom)*. 1–5. https://doi.org/10.1109/CoBCom55489.2022.9880618

[26] Jacob Green. 2022. Dolby Atmos Best Speaker Setup Practices In the Home. https://www.audioholics.com/audio-technologies/dolby-atmos-best-setup-practices.

[27] NMZ Hashim, HH Basri, A Jaafar, MZAA Aziz, A Salleh, and AS Ja. 2014. Child in car alarm system using various sensors. *ARPN Journal of Engineering and Applied Sciences* 9, 9 (2014), 1653–1658.

[28] Emmarie Huetteman. 2022. Heat Waves Affect Children More Severely. https://www.scientificamerican.com/article/heat-waves-affect-children-more-severely/.

[29] Sakila S Jayaweera, Beibei Wang, Xiaolu Zeng, Wei-Hsiang Wang, and KJ Ray Liu. 2023. WIFI-Based Robust Child Presence Detection for Smart Cars. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[30] Sakila S. Jayaweera, Beibei Wang, Xiaolu Zeng, Wei-Hsiang Wang, and K. J. Ray Liu. 2023. WIFI-Based Robust Child Presence Detection for Smart Cars. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. https://doi.org/10.1109/ICASSP49357.2023.10096908

[31] Dietmar Ruwisch Ken Waurin and Yu Du. 2021. How A2B Technology and Digital Microphones Enable Superior Performance in Emerging Automotive Applications. https://www.analog.com/media/en/analog-dialogue/volume-55/number-2/how-a2b-technology-and-digital-microphones-enable-superior-performance-in-emerging-automotive-applications.pdf.

[32] J.C. Kluyver. 1905. A local probability problem. https://dwc.knaw.nl/DL/publications/PU00013859.pdf.

[33] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: Practical Problems for Acoustic Sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking* (Sydney, NSW, Australia) *(MobiCom '22)*. Association for Computing Machinery, New York, NY, USA, 381–390. https://doi.org/10.1145/3495243.3560527

[34] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2023. Room-Scale Hand Gesture Recognition Using Smart Speakers. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (Boston, Massachusetts) *(SenSys '22)*. Association for Computing Machinery, New York, NY, USA, 462–475. https://doi.org/10.1145/3560905.3568528

[35] Jie Lian, Xu Yuan, Ming Li, and Nian-Feng Tzeng. 2021. Fall Detection via Inaudible Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 114 (sep 2021), 21 pages. https://doi.org/10.1145/3478094

[36] Jialin Liu, Dong Li, Lei Wang, and Jie Xiong. 2021. BlinkListener: "Listen" to Your Eye Blink Using Your Smartphone. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 73 (jun 2021), 27 pages. https://doi.org/10.1145/3463521

[37] 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 (sep 2022), 27 pages. https://doi.org/10.1145/3550329

[38] 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* (Los Cabos, Mexico) *(MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 38, 16 pages. https://doi.org/10.1145/3300061.3345439

[39] Martin. 2023. How to Install Car Surround Sound System? https://improvecaraudio.com/car-surround-sound-system/.

[40] Rahul Murmuria, Jeffrey Medsger, Angelos Stavrou, and Jeffrey M. Voas. 2012. Mobile Application and Device Power Usage Measurements. In *2012 IEEE Sixth International Conference on Software Security and Reliability*. 147–156. https://doi.org/10.1109/SERE.2012.19

[41] Branislav M Popovic. 1992. Generalized chirp-like polyphase sequences with optimum correlation properties. *IEEE Transactions on Information Theory* 38, 4 (1992), 1406–1409.

[42] Friedrich Pukelsheim. 1994. The Three Sigma Rule. *The American Statistician* 48, 2 (1994), 88–91. https://doi.org/10.1080/00031305.1994.10476030 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/00031305.1994.10476030

[43] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring Heartbeats using Acoustic Signals on Smart Devices. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*. 1574–1582. https://doi.org/10.1109/INFOCOM.2018.8485978

[44] 1 sixty8 media. 2022. How Much Power Does a Car Radio Really Make? https://www.bestcaraudio.com/how-much-power-does-a-car-radio-really-make/.

[45] 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* (London, United Kingdom) *(MobiCom '20)*. Association for Computing Machinery, New York, NY, USA, Article 52, 14 pages. https://doi.org/10.1145/3372224.3419209

[46] King State. 2023. Hands free car mic module design. https://www.kingstate.com.tw/index.php?do=technology&id=62.

[47] 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 (MobiCom '18)*. Association for Computing Machinery, New York, NY, USA, 591–605. https://doi.org/10.1145/3241539.3241568

[48] Yimiao Sun, Weiguo Wang, Luca Mottola, Ruijin Wang, and Yuan He. 2023. AIM: Acoustic Inertial Measurement for Indoor Drone Localization and Tracking. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (Boston, Massachusetts) *(SenSys '22)*. Association for Computing Machinery, New York, NY, USA, 476–488. https://doi.org/10.1145/3560905.3568499

[49] David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge university press.

[50] Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. 2021. RespTracker: Multi-user room-scale respiration tracking with commercial acoustic devices. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 1–10.

[51] Haoran Wan, Shuyu Shi, Wenyu Cao, Wei Wang, and Guihai Chen. 2023. Multi-User Room-Scale Respiration Tracking Using COTS Acoustic Devices. *ACM Trans. Sen. Netw.* 19, 4, Article 85 (jun 2023), 28 pages. https://doi.org/10.1145/3594220

[52] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300248

[53] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 170 (jan 2018), 20 pages. https://doi.org/10.1145/3161188

[54] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 363–373. https://doi.org/10.1145/2971648.2971670

[55] 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* (New York City, New York) *(MobiCom '16)*. Association for Computing Machinery, New York, NY, USA, 82–94. https://doi.org/10.1145/2973750.2973764

[56] Zhi Wang, Beihong Jin, Siheng Li, Fusang Zhang, and Wenbo Zhang. 2023. ECG-Grained Cardiac Monitoring Using UWB Signals. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 186 (jan 2023), 25 pages. https://doi.org/10.1145/3569503

[57] James Williams. 2023. Where Is The Microphones In My Car? https://techpenny.com/microphones-cars/#Where_is_the_Microphone_Located_in_my_Mercedes-Benz.
[58] TI Staff Writer. 2020. 4 audio trends transforming the automotive industry. https://e2e.ti.com/blogs_/b/behind_the_wheel/posts/automotive-audio-design-trends.
[59] Yue Wu, Fan Li, Yadong Xie, Yu Wang, and Zheng Yang. 2023. SymListener: Detecting Respiratory Symptoms via Acoustic Sensing in Driving Environments. *ACM Trans. Sen. Netw.* 19, 1, Article 3 (jan 2023), 21 pages. https://doi.org/10.1145/3517014
[60] Hee Jung Yoon, Ho-Kyeong RA, Can Basaran, Sang Hyuk Son, Taejoon Park, and Jeonggil Ko. 2017. Fuzzy Bin-Based Classification for Detecting Children's Presence with 3D Depth Cameras. *ACM Trans. Sen. Netw.* 13, 3, Article 21 (aug 2017), 28 pages. https://doi.org/10.1145/3079764
[61] Xiaolu Zeng, Beibei Wang, Chenshu Wu, Sai Deepika Regani, and KJ Ray Liu. 2022. Intelligent Wi-Fi based child presence detection system. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 11–15.
[62] Xiaolu Zeng, Beibei Wang, Chenshu Wu, Sai Deepika Regani, and KJ Ray Liu. 2022. WiCPD: Wireless child presence detection system for smart cars. *IEEE Internet of Things Journal* 9, 24 (2022), 24866–24881.
[63] Fusang Zhang, Kai Niu, Jie Xiong, Beihong Jin, Tao Gu, Yuhang Jiang, and Daqing Zhang. 2019. Towards a Diffraction-Based Sensing Approach on Human Activity Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 1, Article 33 (March 2019), 25 pages.
[64] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can "Hear" Your Heartbeat! *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 161 (dec 2020), 24 pages.
[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. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 53 (March 2018), 23 pages.
[66] Yi Zhang, Weiying Hou, Zheng Yang, and Chenshu Wu. 2023. VeCare: Statistical Acoustic Sensing for Automotive In-Cabin Monitoring. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 1185–1200.