VeCare: Statistical Acoustic Sensing for Automotive In-Cabin Monitoring

Yi Zhang*,†, Weiying Hou*, Zheng Yang†, Chenshu Wu*
*The University of Hong Kong
†Tsinghua University
{zhangyithss@gmail.com, wyhou@cs.hku.hk, hmilyyz@gmail.com, chenshu@cs.hku.hk}

Abstract

On average, every 10 days a child dies from in-vehicle heatstroke. The life-threatening situation calls for an automatic Child Presence Detection (CPD) solution to prevent these tragedies. In this paper, we present VeCare, the first CPD system that leverages existing in-car audio without any hardware changes. To achieve so, we explore the fundamental properties of acoustic reflection signals and develop a novel paradigm of *statistical acoustic sensing*, which allows to detect motion, track breathing, and estimate speed in a unified model. Based on this, we build an accurate and robust CPD system by introducing a set of techniques that overcome multiple challenges concerning sound interference and sensing coverage. We implement VeCare using commodity speakers and a single microphone and conduct experiments with infant simulators and adults, as well as 15 young children for the real-world in-car study. The results demonstrate that VeCare achieves an average detection rate of 98.8% with a false alarm rate of 2.1% for 15 children in various cars, boosting the coverage by over 2.3× compared to state-of-the-art and achieving whole-car detection with no blind spot.

1 Introduction

The ability of cars to sense, and save lives, inside a car remains to be improved. One life-critical feature that is widely missing is in-vehicle *Child Presence Detection* (CPD). Every year, many children have been unintentionally and unknowingly left in parked cars, or have got stuck into a car independently. As the temperature inside a car can rise rapidly[1], especially in hot months, serious injuries or heatstroke deaths could happen to children being left alone inside a car. It takes only a matter of minutes before the heat can overwhelm a child's ability to regulate his/her internal temperature and cause injuries/deaths as a child's core temperatures increase three to five times faster than an adult's [3]. On average, around 40 children dying from hot cars have been witnessed each year (about one every 10 days), leading to over 900 pediatric vehicular heatstroke (PVH) deaths on record since 1998 in the US alone [50]. Despite remarkable advances in automobiles in recent years, unfortunately, the cases of hot car deaths are only increasing, with 2018 and 2019 being the record years of 54 and 53 deaths each [50]. All of these deaths could have been prevented, if the car can detect the unattended child timely and responsively alert concerned parties or take prompt actions to keep the car cool and the child safe (as depicted in Fig. 1).

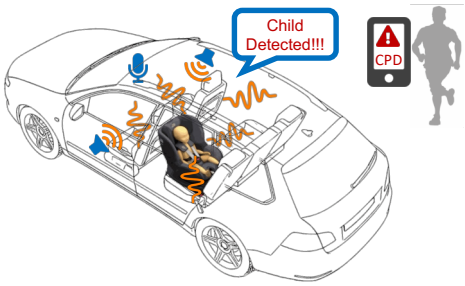


Figure 1: Application scenario of VeCare. It detects an unattended child using in-car audio and alert registered parties for immediate responses and/or activate the air conditioner to keep the child safe automatically.

The widespread and tragic problem has driven governments and the auto industry to take initiatives to make CPD a compulsion for future cars [49, 76], which fosters an expected market of $400 million by 2025 [27]. Existing solutions include early systems using special sensors such as optical/weight/pressure/ultrasonic sensors [4, 20, 21, 54], cameras [10, 13, 85], as well as recent efforts with Ultra-Wideband (UWB) or millimeter-wave (mmWave) radars [26, 28, 66], WiFi [45, 81], etc. These solutions, however, suffer from different limitations. Many works focus on adult passenger monitoring, and cannot generalize well to infants and toddlers. And the sensing coverage is mostly limited to only the seats (for special seat sensors) or a certain Field-of-View (FoV) (for

This work was done when Yi Zhang was a Research Assistant at HKU.

[1] A car can heat up by 19 degrees in just 10 minutes. Even on a mild day, the temperature in a parked car can rise to extremely dangerous and potentially fatal levels for infants and toddlers. As reported, heatstroke can occur even when outside temperatures are just 57°F [3].

cameras/UWB/mmWave radars), leading to degraded accuracy in Non-Line-Of-Sight (NLOS) scenarios and blind spots, e.g., when a child is in a rear-facing car seat, blocked by a seat, or on the car floor. More importantly, these techniques require extra hardware that is not standard offerings in today's cars to be precisely installed[2]. This not only introduces additional hardware and manufacturing costs, which are huge considering more than 80 million new cars annually, but is also backward-incompatible with most of the over one billion existing cars in the world. A truly pervasive system that requires no extra hardware and works for all cars still lacks.

In this paper, we ask the following question: *Can we build an accurate and robust in-cabin monitoring system by using only readily available in-car modules without any hardware modifications?* We present the design and implementation of such a system, named VeCare, by leveraging in-car audio systems, which are widely available in most, if not all, modern cars. As illustrated in Fig. 1, it operates by transmitting sound signals from the speakers and analyzing their reflections recorded on a microphone, which have interacted with the human body, if present. VeCare accurately and responsively detects tiny motions and extremely weak breathing of young children including newborns. It can reliably detect the presence of a child in a car, achieving whole-car detection with no blind spots. Importantly, it can be readily deployed in existing and emerging car models, offering the best ubiquity superior to the aforementioned other solutions.

Albeit acoustic sensing has been extensively studied, VeCare introduces a novel paradigm of *Statistical Acoustic Sensing* (SAS). The mainstream practice in the literature mostly focuses on *geometrical* parameters, e.g., Time of Flight (ToF) and Doppler Frequency Shift (DFS), of a few multipath reflections around the range where a target presents. Differently, inspired by recent advances in WiFi sensing [79, 88, 89], we propose to analyze the *statistical* characteristics of acoustic signals by leveraging *all* multipath reflections, which can be all affected by the target and therefore can contribute to sensing if utilized properly. Towards that end, we explore unseen properties of acoustic multipath signals and accordingly develop a novel SAS model that underpins a unified pipeline for detecting motion, estimating breathing rates, and even measuring moving speeds. The proposed SAS model truly embraces all the reflections and favors complex multipath environments, while requiring only a single microphone rather than a microphone array.

Based on the SAS model, we develop a set of techniques that overcome multiple challenges in translating SAS into a practical CPD system. First, effective acoustic channel estimation is non-trivial, mainly because of ambient sound noises, limited frequency band (up to 24kHz) on commodity devices, and multiple concurrently-transmitting speakers. In VeCare, we adopt Kasami Sequence, a pseudo-noise orthogonal sequence for channel measurement, which provides resilience to environmental noises as well as orthogonality for multi-speaker sensing. Second, acoustic sensing is known to suffer from limited coverage, e.g., typically within 1-meter range [53, 71], mainly because sound reflections off the human body are considerably weak. The problem is aggravated for young children who have even weaker motion/breathing. We boost the sensing coverage by statistically leveraging all multipaths (time diversity), optimally combining multiple subcarriers (frequency diversity), and opportunistically exploiting multiple speakers (space diversity), which ultimately allows comprehensive detection in a car. Last, CPD is a time-critical mission requiring fast response (e.g., detection within 10 seconds [49]). We design an instantaneous motion/breathing detector for CPD based on a time-domain approach, which can detect child presence rapidly (motion in a few seconds and breathing with a minimum delay slightly exceeding one breathing cycle).

We prototype an end-to-end system using commodity off-the-shelf (COTS) microphones and speakers, including car speakers and microphones. We first use infant simulators and recruit adults to systematically evaluate VeCare under various conditions both in buildings and in cars. Then we conduct a real-world study with 15 children, aged 0 to 6 except for one 10-year-old, for testing in various cars. Our results show that VeCare achieves an overall detection rate of 98.8% with a false alarm rate of 2.1%, using a single microphone. It can detect motion accurately up to 5 m, and estimate breathing at a distance of 4.5 m for an adult and 1.6 m for an infant, outperforming the state-of-the-art by 2.3×. Using in-car audio without any hardware changes, VeCare holds great potential to be widely adopted for practical CPD.

Contributions: In summary, our goal is to enable a ubiquitous solution to accurate and robust in-car CPD to prevent PVH deaths. To this end, we make three key contributions to delivering the first CPD system using accessible in-car audio: (1) We introduce a novel statistical acoustic sensing model that can detect motion, track breathing, and estimate speed by leveraging all the reflections. (2) We present a pipeline of techniques to detect motion, speed, and breathing based on the SAS model accurately and robustly, with a significantly enlarged coverage. (3) We design and implement a prototype CPD system VeCare on COTS devices and conduct extensive experiments in the real world with infant simulators and young children. Not only is VeCare a promising solution to the critical application of CPD, we also believe the proposed SAS opens a new paradigm in acoustic sensing for various applications in smart homes, healthcare, and beyond.

2 Design Space

Design Scope: Among over 900 deaths reported since 1998 [50], the primary circumstances resulting in PVH deaths include a caregiver forgetting a child in a vehicle (about 55% of

[2]WiFi is becoming popular in modern cars, but still many do not have it.

the cases), someone knowingly leaving a child in the vehicle (∼20%, e.g., running a quick errand), and the child gaining access to and getting stuck in the vehicle (about 25%). For all of these circumstances, if the car can detect the child left behind and remind/alert parents and caregivers promptly, immediate actions can be taken, by caregivers or by the car, to end these entirely preventable tragedies.

CPD systems are designed for this purpose. Typically, a CPD system is expected to run, for a short period of time, after the driver turns the engine off and locks the doors. Therefore, we mainly focus on in-cabin monitoring of a parked, closed car and do not consider a driving car. The system should then detect a child presenting anywhere inside the car quickly (e.g., within 10s [49]), and take registered actions responsively, such as alerting corresponding parties (e.g., car owners, parents, caregivers) via horn alarms and/or messages, activating the air conditioner automatically if the temperature goes high, etc. Yet how a CPD system exactly reacts is not our focus in this paper. For example, questions like how a CPD system integrates into the car system and responds to child presence (which depends on auto manufacturers, car owners, parents and caregivers), and whether the system should be a built-in component or a standalone module are out of our scope.

Why Acoustics? Although presence detection is not a new topic, existing works mostly focus on adult subjects in buildings. In-car CPD is particularly challenging because it demands a very high detection rate (any miss detection can lead to a potential tragedy) and it requires such a high accuracy for extremely tiny motion/breathing from an infant, which prior methods cannot achieve. Different modalities can be used for CPD, such as WiFi, UWB, mmWave, cameras, etc. In VeCare, we choose audio modules mainly because of the best ubiquity: Audio systems have been standard components in modern cars, which are nowadays commonly equipped with two, four, or more speakers plus *one microphone*. These speakers are most commonly installed around the dashboard, the front/back doors, and/or the rear deck, while the microphone is usually installed on the dashboard, around the rear-view mirror, or behind the steering wheel. They are placed in such a way primarily for high sound quality, which also turns out to support a good whole-car coverage for sensing.

While acoustic sensing is usually vulnerable to ambient sounds, vehicles today are designed and manufactured to provide the necessary level of safety and to muffle as much road noise as possible. Therefore, in the CPD application, the impact from the noise outside a closed car is insignificant. On the other hand, embedding sensing signals on the audible frequency band may result in shrill noises that are intrusive to human ears. Previous work has nicely modulated sensing signals into white noise [71] or on only the inaudible frequency band [9] to solve the problem. In our case, VeCare can work on either the full bandwidth (e.g., up to 24kHz) or only the inaudible band (e.g., above 18kHz), depending on practical choices. While existing works like BreathJunior [71] also monitor infants' vital signs, they are not suitable for in-car CPD as they rely on a large microphone array that is unavailable in commercial cars. Overall, acoustic signals appear to be an attractive choice for ubiquitous and practical in-car CPD, yet it entails numerous challenges to build an accurate and robust system using a single microphone.

3 Statistical Acoustic Sensing

We first present a novel statistical acoustic sensing paradigm. Our model is inspired by the success of statistical approaches in WiFi sensing [78, 79, 88, 89]. To put it briefly, this line of work treats each multipath component as a scatterer and investigates the spatial-temporal statistical properties of WiFi Channel State Information (CSI), which have been shown to imply important information such as motion [89], speed [79, 88], as well as breathing [90]. These approaches show superiority in complex environments and have been commercialized as real-world products on commercial WiFi devices [24, 25, 39]. In below, we first present our new observations on multipath propagation of acoustic signals, and then show that similar statistical properties also hold for acoustics.

Acoustic Multipaths: Unlike WiFi signals, one can effectively resolve multipath signals due to the high range resolution of acoustic signals. As a result, previous works mostly only focus on reflections from the range of interest and segment others out. However, our measurements show that a target at a certain range not only alters the reflection around that range, but also distorts multipath signals arriving later to a considerable extent. As shown in Fig. 2, while a human target contributes the strongest reflection at the Channel Impulse Response (CIR) taps corresponding to her/his range (i.e., 1.25 m), the CIR taps up to 7 m after that range are also altered remarkably. In comparison, the CIR taps are mostly noises in the empty case without human presence. Our key insight is that all these multipath distortions, if aggregated properly, can contribute useful information for sensing. The problem is, how can we truly leverage these weak and noisy multipaths?

SAS Model: CSI, a.k.a Channel Frequency Response (CFR), is the frequency-domain counterpart of CIR. CSI for an acoustic multipath channel of frequency f at time t is denoted as

$$H(f,t) = \sum_{r=1}^{R} a_r(t)\exp(-j2\pi f \tau_r(t)), \quad (1)$$

where $a_r(t)$ and $\tau_r(t)$ are the complex amplitude and propagation delay of the r-th reflection path, respectively, while R denotes the total number of paths. From a rich-scattering perspective, each reflection path can be treated as a scatterer that scatters the incoming energy back to the receiver (i.e., microphone) [22, 33, 89]. Thus, we have

$$H(f,t) = \sum_{i \in R_D} H_i(f,t) + \sum_{j \in R_S} H_j(f,t) + N(f,t), \quad (2)$$

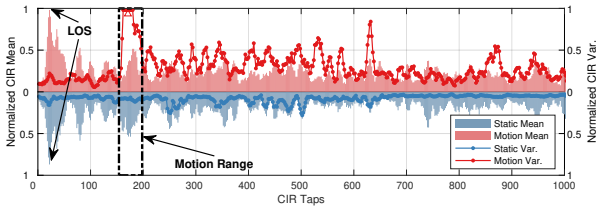


Figure 2: CIR measurements with and without human motion. The large CIR values at the motion range are truncated for the sake of visualization.

where $H_i(f,t)$ denotes the component contributed by the i-th scatterer, $N(f,t)$ is the noise term with variance σ_N^2, and R_S and R_D denote the set of static and dynamic scatterers, respectively. Assuming all scatterers are statistically independent of each other, each with the same variance $\sigma_i^2(f)$ and approximately zero means, it has been established in the context of WiFi signals [22, 79, 88], that the Autocorrelation Function (ACF) of $H(f,t)$ obeys the 0^{th}-order Bessel function of the first kind. That is, denoting $\rho_i(f,\tau)$ as the ACF of $H_i(f,t)$ with time lag τ, we have $\rho_i(f,\tau) = J_0(kv_i\tau)$, where $J_0(x) = \frac{1}{2\pi}\int_0^{2\pi}\exp(-jx\cos(\theta))d\theta$, v_i is the moving speed of $H_i(f,t)$, and k is the wavenumber. Suppose there is one single moving target, and thus all dynamic scatterers have approximately the same speed v, $v_i \approx v, \forall i \in R_D$. This assumption is realistic because, for human subjects, the torso scatterers dominate others and have a similar speed. Then the ACF $\rho(f,\tau)$ of $H(f,t)$ can be associated with the target's moving speed v as follows [79, 88]: For $\tau \neq 0$,

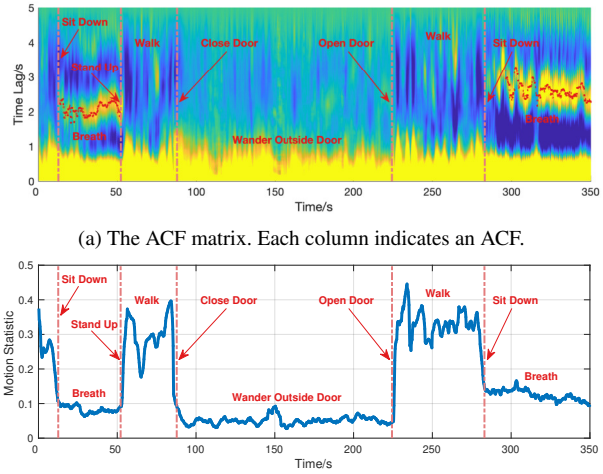
$$\begin{aligned}\rho(f,\tau) &= \frac{\sum_{i\in R_D} 2\pi\sigma_i^2(f) + \sigma_N^2(f)\delta(\tau)}{\sum_{i\in R_D} 2\pi\sigma_i^2(f) + \sigma_N^2(f)} J_0(kv\tau) \\ &\triangleq g(f)J_0(kv\tau),\end{aligned} \tag{3}$$

where $\delta(\cdot)$ is the Dirac's delta function and $g(f)$ is defined as the channel gain of $H(f,t)$. Eq. (3) bridges the ACF of the CSI with the target's moving speed. In practice, we can calculate the sample ACF, $\tilde{\rho}(f,\tau) = \rho(f,\tau) + n(f,\tau)$, from a time series of CSI measurements with a noise term $n(f,\tau)$.

While statistical approaches in WiFi sensing [79, 88–90] have demonstrated success for practical solutions and commercialized products [25, 39, 60, 67], they have not been previously explored in acoustic sensing. VECARE brings statistical sensing approaches to acoustic sensing. In the following, we perform real-world measurements to demonstrate the properties of acoustic CSI and explain how to derive motion, speed, and breathing based on the SAS model.

1) Detecting Motion: Similar to the *motion statistic* defined in [89] for WiFi CSI, we find that the defined channel gain $g(f)$ in Eq. (3) is a sensitive and robust indicator for acoustic motion detection. From Eq. (3), we have $g(f) = \lim_{\tau\to 0}\rho(f,\tau)$ since $\lim_{\tau\to 0} J_0(kv\tau) = 1$. Hence, given a sufficient CSI sampling rate F_s, we can approximate $g(f)$ as the value of the first tap of the ACF, i.e.,

$$g(f) = \tilde{\rho}(f, \tau = 1/F_s). \tag{4}$$



(a) The ACF matrix. Each column indicates an ACF.

(b) The extracted motion level

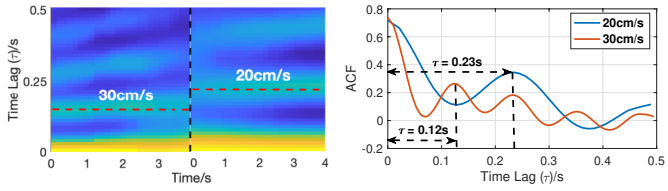
Figure 3: Human motion and breathing in a bedroom.

If any motion presents, the value of $g(f)$ is greater than zero; otherwise $g(f) \to 0$. Fig. 3 shows an example of $g(f)$ in the case of human presence and absence, respectively. As seen, there is a clear gap between the empty level and the values for motion. Additionally, $g(f)$ exhibits larger values when the motion is stronger/closer, implying that it also indicates motion strengths.

2) Tracking Breathing: ACF itself is a time-domain approach to identifying periodic signals. Therefore, we can also detect breathing signals from the ACF, which are periodic signals induced by repeated chest movements. If a breathing signal is captured by CSI, the ACF will observe a prominent peak at the time lag τ_b corresponding to the cycle time, as in Fig. 3(a). Thus, by finding time lags of these peaks over time, we can track one's breathing rates as $60/\tau_b$ BPM (breath per minute). Note that as a time-domain approach, ACF in principle is faster for breathing estimation compared to spectrum-based approaches, which usually require a much longer window to yield better frequency resolution.

3) Estimating Speed: As indicated by Eq. (3), the ACF of CSI is a function of speed v, which underpins a statistical approach entirely different from the Doppler effect for speed estimation [88]. Specifically, as shown in Fig. 4b, the shape of the ACF $\rho(f,\tau)$ resembles the Bessel function $J_0(x)$ with $x = kv\tau$, meaning that we can estimate the speed by aligning $\rho(f,\tau)$ with $J_0(x)$. Assuming x_0 is the constant value corresponding to the first peak of $J_0(x)$, then the moving speed v can be calculated as [79, 88]: $\hat{v} = \frac{x_0}{k\tau_s} = \frac{x_0\lambda(f)}{2\pi\tau_s}$, where τ_s is the time lag corresponding to the first local peak of the ACF $\rho(f,\tau)$ and $\lambda(f)$ is the wavelength of subcarrier f. Fig. 4 illustrates an example of speed estimation with the setup in Fig. 8(c), which shows that the ACF reacts to the moving speed faithfully as the above equation implies.

Remark: As seen, a peak in the ACF can either indicate a speed signal or a periodic signal. However, we notice that the peak locations for breathing (e.g., 1-5s for breathing rates



(a) ACF matrix (b) ACF

Figure 4: Speed estimation. A plate moves along a track programmed with different speeds (frequency @ 10 kHz).

60-12 BPM) are usually of magnitude longer than those for speed (e.g., <0.5s for 0.5 m/s using 10 kHz sound, and the faster the speed, the smaller the delay), a sufficient difference to determine whether to estimate breathing or speed. We do not involve speed in the current VECARE for CPD as motion and breathing will be sufficient. Yet we still present a brief description here, with an analysis of a few more issues in Appendix A.1, to show a unique approach offered by the SAS model. We keep the exploration of the full potential of SAS speed estimation for future work.

4 VECARE Design

Translating the proposed SAS model into a practical CPD system on commodity speakers and microphones still entails multiple challenges. In this section, we overcome these practical challenges and present the design of VECARE.

4.1 Acoustic Channel Estimation

The proposed SAS model leverages acoustic CSI, which demands effective channel estimation. Several unique characteristics of sound waves make it particularly challenging. First, the sound wave speed is orders of magnitude slower than that of light and EM waves, which imposes limitations on the max possible CSI sampling rate of the acoustic channel. For example, given the in-air sound speed of around 343 m/s, the propagation delay of a path of 7 meters in length will be greater than 20 ms, requiring a minimum channel measurement internal larger than 20 ms to avoid signal mixture. Second, acoustic sensing is vulnerable to environmental sound interference, especially when it is limited to a frequency band under 24 kHz on commodity devices. Ambient interference like the human voice, music, and natural sounds, can smear channel measurements for certain frequency bands. Moreover, concurrent sensing signals transmitted on multiple speakers, if used, may also interfere with each other.

In VECARE, we investigate Pseudo-Noise (PN) sequence [57] for CIR measurements. PN sequence is a set of noise-like signals and can be effectively distinguished from both a time-shifted version of itself (a.k.a, excellent auto-correlation properties) and every other signal in the set (a.k.a, excellent cross-correlation properties), which have been used in spread-spectrum communications, radar sensing, etc [52]. Among different types of PN sequences such as m-sequence [58], Golay sequence [64], GSM training sequence [86], and Zadoff-Chu (ZC) [63], we choose Kasami sequence [31] for CSI estimation because of its superior properties of orthogonality and noise tolerance. Fig. 5a shows the auto-correlation and cross-correlation of a pair of example Kasami sequences with period 2^6-1. The auto-correlation produces an impulse-like signal with minor side lobes, while the cross-correlation only produces minor values that are much smaller than the impulse of auto-correlation. Note that our approach is not limited to a particular channel estimation technique, but can work with any approach, including the widely used FMCW, that provides effective CSI.

CIR estimation with Kasami sequence: Fig. 6 shows the channel estimation process in VECARE, with two speakers as an example. We generate two orthogonal Kasami sequences s_1 and s_2 with the same length and periodically transmit them on both speakers simultaneously. The transmitted sequences undergo different time delays and attenuation before being captured by the microphone. On the receiver side, we correlate the microphone recordings with s_1 and s_2 separately to get CIR streams of the two channels, and slice them into segments with the same length as s_1 and s_2, resulting in the CIR estimates $h_1(t)$ and $h_2(t)$. We can then convert $h_1(t)$ and $h_2(t)$ into the frequency domain by performing Fourier transform and obtaining the CSI $H_1(f,t)$ and $H_2(f,t)$.

Since a correlation operation is equivalent to a conjugate multiplication in the frequency domain, the measured CSI $\tilde{H}(f)$ using Kasami sequence can be represented as

$$\begin{aligned}\tilde{H}(f) &= [S(f)\cdot H(f)+N(f)]\cdot S^*(f)\\ &= \|S(f)\|_2\cdot H(f)+N(f)\cdot S^*(f),\end{aligned} \tag{5}$$

where $H(f)$ denotes the ideal CSI, and $S(f)$ and $N(f)$ are the frequency-domain representations of Kasami sequence and sound noises respectively. $S(f)$ is a wideband signal spanning over the whole spectrum, and $S^*(f)$ is its conjugate. The term $\|S(f)\|_2\cdot H(f)$ approximates to a scaled version of $H(f)$. An example of the measured CIR is shown in Fig. 2.

We generate Pulse Coded Modulation (PCM) samples from the Kasami sequence and play them on the speaker without an extra modulation process. Several previous works have exploited more complicated signal modulation techniques to improve measurement performance including Orthogonal Frequency-Division Multiplexing (OFDM) [48] and BPSK [86], which is not necessary for VECARE.

Audibility and Interference: There are two issues with the above channel estimation process. First, the Kasami sequences composed of 1's and -1's with sharp transitions in between can be intrusive to human ears. Second, by transforming CIR, the obtained CSI spans the full spectrum, which might be polluted by the ambient sound noises, especially on the audible frequency band. To circumvent these problems, we apply a high-pass filter (HPF) on both the transmitted and received signals. The passband can be set flexibly, and

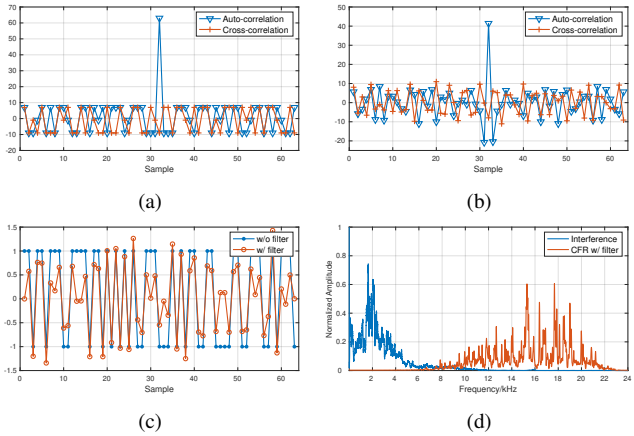


Figure 5: Kasami sequence. The auto-correlation and cross-correlation of (a) Kasami sequence and (b) Kasami sequence after applying a 10 kHz high-pass filter. (c) The Kasami sequence before and after a 10 kHz high-pass filter. (d) The spectrum of background traffic interference and CSI after the 10 kHz high-pass filter. The sound sampling rate is 48 kHz.

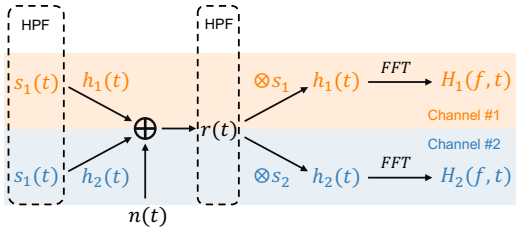


Figure 6: CSI estimation process with Kasami sequence.

VeCare can work reliably even with only the inaudible pseudo-ultrasound band, e.g., above 18kHz. Here we use an empirical passband of 10 kHz as an example for illustration and will evaluate difference choices extensively in §6.

Transmitter Filtering: When we apply the filter on the transmitter side, as shown in Fig. 5c, the binary values on the time domain signal are softened, and the output sound contains fewer intrusive bursts. The major concern is whether this filter operation breaks the auto-correlation and cross-correlation properties of Kasami sequences or not. To validate, we plot the auto-correlation and cross-correlation between the original and filtered Kasami sequence in Fig. 5b. It can be seen that, after applying the filter, the auto-correlation still observes an impulse (with a decrease in SNR) while the cross-correlation approximates the noise.

Receiver Filtering: On the receiver side, the term $N(f) \cdot S^*(f)$ in Eq. 5 is eliminated by high-pass filtering. This is because typical daily sound interference, such as traffic and human voice, mostly occurs in the frequency band below 10 kHz. Fig. 5d shows the spectrum of traffic noise and the measured CSI after applying the HPF. As seen, the noise is successfully removed. Meanwhile, we are left with fewer subcarriers for sensing because of the filtering, motivating us to maximize the sensing signals (§4.2).

Resilience to asynchronization: The speakers and microphones are connected to the same controller in our prototype and in cars as well. Due to hardware imperfections and software latency, however, they are not perfectly synchronized, which makes it difficult to measure accurate channel response. Fortunately, synchronization errors only introduce phase offsets in CSI, which does not affect VeCare because CSI is measured consecutively without blanks in between and we only use the amplitude. We will experimentally verify this in §6 and show that VeCare even works with separate speaker and microphone, while providing proof in Appendix A.2.

4.2 Sensing Signal Enhancement

Sound reflection off human bodies is considerably weak, a major reason confining the coverage of human-centric acoustic sensing [71, 74]. The problem is aggravated when the target is an infant/toddler in CPD applications. Our SAS model utilizes all multipaths for better coverage. We now present an effective technique to exploit subcarrier diversity which further boosts the sensing signals, particularly for breathing.

Subcarrier diversity is attributed to frequency selective fading, a well-known phenomenon in wireless communications. Fig. 7 demonstrates a breathing example, with the calculated ACF matrix on different subcarriers in Fig. 7a and the raw amplitude of good subcarriers (manually selected) in Fig. 7b. Some subcarriers capture dominant breathing signals, while others merely observe noises, even in such an example with strong breathing signals. We also notice that, because of the complex multipath propagation, the most sensitive subcarriers can vary over time randomly. Therefore, it is critical to dynamically find the best subsets of subcarriers and effectively combine them to maximize the signal SNR.

Like in WiFi sensing works [79, 90], we employ Maximal Ratio Combining (MRC) [6], a classical diversity combining method in wireless communications which optimizes the receiving SNR, to combine multiple subcarriers optimally. Since the noise terms on different subcarriers are statistically independent, we can maximize the signal SNR by MRC as

$$\hat{\rho}(\tau) = \sum_{f \in F} w(f)\tilde{\rho}(f,\tau), \tag{6}$$

where $\hat{\rho}(\tau)$ is the combined ACF, $w(f)$ denotes the normalized weight for combining subcarrier f (i.e., $\sum_{f\in F} w(f) = 1$), and F is the set of all subcarriers. The optimal weight $w(f)$ should be linearly proportional to the gain on each subcarrier. Following [90], we adopt the normalized $g(f)$, defined in Eq. (4), as the weight $w(f)$ in VeCare. Note that, some intuitive criteria commonly used like mean/variance of CSI amplitude cannot serve as the optimal weights for MRC, as they are not linearly proportional to the gain $g(f)$ and subcarriers with higher amplitude means/variances do not necessarily better capture the sensing signals, as shown by Fig. 7a.

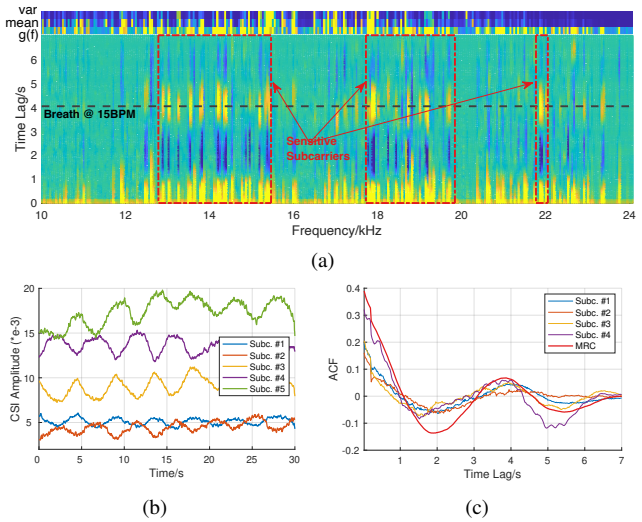


Figure 7: Breathing signals. (a) The ACF over all the subcarriers, with gain $g(f)$, mean and variance of CSI amplitude marked on the top. (b) The CSI amplitude on different subcarriers. (c) ACF on subcarriers and after MRC.

We can combine multiple subcarriers here because, by taking the ACF, the sensing signals (either breathing or speed) are synchronized across different subcarriers (Fig. 7c). It cannot be done directly on the raw amplitude due to the considerable phase offsets of breathing/speed signals on different subcarriers, as demonstrated by Fig. 7b. In case multiple speakers are available, the subcarriers on different speakers can be combined similarly, and again, asynchronization among different speakers is not an issue. By combining them, we can further boost the SNR and extend the sensing coverage.

4.3 Child Presence Detection

VECARE combines motion and breathing detection for CPD.

Motion detection: It is straightforward to detect motion. We first average the gains $g(f)$ across all subcarriers and obtain $\bar{g} = \frac{1}{|F|}\sum_{f\in F} g(f)$. Then given a preset motion threshold ε, the system detects motion at any given time t if $\bar{g}(t) > \varepsilon$; otherwise no motion presents. We use equally averaged $\bar{g}$ instead of using MRC because the averaged values across all subcarriers with equal gains will approximate zero in absence of motion, allowing us to find a generic threshold ε for different environments and cars.

Breathing Detection: To estimate the breathing rate, we first need to find whether there exists a dominant peak in the enhanced sensing signal $\hat{\rho}(\tau)$. To achieve so, we adopt similar criteria in [90] for peak finding. Basically, we first examine the peak prominence, width, and amplitude to identify potential peaks. Then we further check the peak location to sift out those beyond the typical range of human breathing rates, e.g., 10-60 BPM. We also compare the motion level $\bar{g}$ against the peak value as there will be unlikely breathing if the motion level is way larger than the peak value. Once we find the peaks corresponding to breathing, we will estimate the peak location τ_b and accordingly derive the breathing rate.

Real-time CPD: In real-time, we employ a sliding window on the continuous CSI to calculate the ACF. We employ a shorter window of CSI (e.g., 1s) for calculating the ACF for motion detection to make it more responsive while saving computation. While for breathing, a minimum window larger than a typical breathing cycle (e.g., 6s, which can be shorter for children who usually have higher breathing rates) is desired. As motion is more common and the computation is more efficient, we will only further perform breathing estimation when no motion can be detected. Note that the system can output detection decisions as fast as every CSI sample, or at a predefined lower rate, e.g., every 1 second, to save energy. Once we have the time series of motion/breathing decisions, we check them within a certain window, e.g., 5s, and child presence is claimed if there is a certain amount of motion/breathing detection, e.g., >30% of the window.

5 Implementation

Hardware: We implement VECARE using a programming audio prototype, which consists of a MiniDSP UMA-8-SP USB microphone array [44] with 7 built-in Knowles SPH1668LM4H microphones (we use only one of them) and PUI Audio AS07104PO-R speakers [5] connected to the MiniDSP board via cables. As in Fig. 8(g), we also evaluate the performance on a variety of commodity devices used in consumer electronics and cars, including JBL Stage1 621 car speaker [29] and Linhuipad car microphone [38], JBL Clip 4 speaker, Sony SRS-XB23 speaker, Razer Seiren Mini Mic, and speakers and microphones on Macbook and iPhone, etc. Again, we always use *only one single microphone* throughout our experiments, even if more are available. We connect this prototype to either a computer or Raspberry PI 3 Model B+.

Software: We implement signal generation and transmission as well as all our algorithms using MATLAB mainly for benchmark analysis. We also build an end-to-end prototype of our system running in real-time using Python3.9, which can run on embedded devices (Raspberry PI in our case).

Kasami Sequence: A longer period of Kasami Sequence allows higher SNR for channel estimation, which, however, creates an immediate conflict with sampling rates. To trade-off, we use a sequence of period $2^{10}-1$ modulated into 0.02 s, which allows a desired sampling rate of 50 Hz to use in VECARE. By default, we use 3 seconds of CSI for motion calculation and use 8 seconds for breathing rate estimation.

Handling Sharp Interference: By applying a high-pass filter, we successfully get rid of most of the daily environmental noises. However, if there are sharp and short impulse-like noises (e.g., horn honk/beep), the impacts may go above 10 kHz and cause false motion detection. We notice that these kinds of sharp noises will impose a sudden change in the CSI amplitudes, which translates into a special ACF pattern,

which linearly decreases first and then linearly increases (See Appendix A.3 for more details). Therefore, we design a detector to identify this linear decrease-then-increase pattern and skip CPD during the interfered period. By doing so, VECARE becomes immune to sharp noise like horn beep, an important feature making it more practical. Although this would reduce the effective protection time (the system is not working in presence of such noises), we argue the impact is minimal because these noises are usually short (∼1s) while VECARE detects so rapidly that it can find a period for detection.

6 Evaluation

6.1 Experimental Setup

We conduct experiments both in office environments and in cars, as Fig. 8 illustrates, which mainly consist of three parts: 1) Indoor experiments with adults to evaluate the performance in large space against various parameters. 2) Indoor and in-car experiments with infant simulators. Our evaluation involves two infant simulators. One is Laerdal SimNewB (Fig. 8(b)), a high-end model (retail price around $30,000) offered by the clinical facility in our university's medical school, which is co-created by the American Academy of Pediatrics and allows to set the breathing rate as well as move various body parts. The second one comes with a breathing motor and does not support body movements. 3) Real-world in-car experiments with children. We recruit 15 young children and perform CPD in 7 different cars including sedan and SUV.

Ground truth for adult breathing is measured by Plux piezoelectric Respiration (PZT) sensor [7]. Infant simulators have a preset fixed breathing rate. We did not record the ground truth breathing rate of children participants as it is difficult to have their cooperation. Motion and presence ground truths are manually labelled. This work does not raise any ethical issues and has been approved by our university's IRB. No sensitive data like personal identifiers were collected.

To show our performance under extreme responsiveness constraints, by default we use a 2-second window for the decision below. A higher detection rate is expected if a longer decision window is applied. Also, we use only one speaker for evaluation unless otherwise specified. Using more speakers is expected to provide larger coverage. We mainly use detection rate (DR) and false alarm rate (FAR) as the evaluation metrics for motion, breathing, and overall presence detection, while we also evaluate the mean absolute breathing rate error.

6.2 Indoor Performance

We first evaluate with comprehensive indoor experiments to validate motion detection and breathing estimation.

Motion Detection: We first evaluate the motion detection performance in a 7m×5m conference room. We set up one microphone and one speaker in the corner. The room is in an

Figure 8: Experimental setups

office building, with constant noise from the central fan and occasional footstep sounds when people pass by the outside corridor. An adult is asked to sit in a chair, at various distances from 1m to 5 m, and only move his one hand slowly to mimic the tiny motion of a child. We also test with the speaker facing different angles with respect to the subject. As shown in Fig. 9, VECARE achieves an average detection rate of 98.1%, which maintains 94.1% even when the user is 4∼5 m away from the speaker and microphone, while the false alarm rate is only 1.1%. The performance degrades slightly when the subject is at a distance and at an angle of 60°. Note that the motion detection rate is almost 100% when the user is within 3.5 m, a sufficient distance to cover a typical car.

Breath Estimation: Now we evaluate the breathing estimation performance in the same environment. First, we also test with an adult subject at different distances, sitting still in a chair. As shown in Fig. 10, we achieve a mean absolute error of 0.88 BPM within the distance of 3 m, including all orientations. More importantly, VECARE can detect breathing rate at a range as far as 4.5 m, with a slight increase in breathing rate error. We also evaluate the case when there is no Line-Of-Signt (LOS) between the speaker and the subject, as well as the case when the user wears a thick down jacket. As portrayed in Fig. 11, even when a user is wearing a thick coat, VECARE can still pick up the breathing rate at distance up to 4 m. When the speaker is blocked, the maximum range of breathing estimation decreases to 2.5 m, still more than enough to cover an entire car. Note that the accuracy under occlusion is not necessarily lower than that for LOS cases since VECARE embraces all multipath reflections to significantly enhance the NLOS scenarios, which may experience richer multipath effects.

Evaluation with SimNewB: Now we carry out a feasibility study with the SimNewB newborn simulator in a clinical facility, which features tens of beds and has continuous machinery and HVAC noises. The experimental setup is illustrated in Fig. 8(a). During the tests, the laboratory technician randomly set the breathing rate of the newborn simulator. As shown in Fig. 13, VECARE can detect the newborn simulator's very weak breathing reliably, achieving an average detection rate of 87.8% with a mean error of 3.43 BPM, which decreases to 78.0% with an increased mean error of 8.6 BPM when the newborn is covered with a blanket. Note that we didn't exhaust various breathing rates due to limited access to the fa-

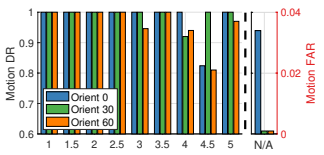


Figure 9: Indoor motion detection performance.

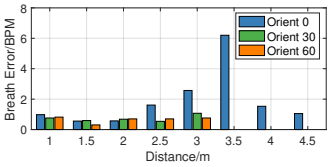


Figure 10: Indoor breath estimation performance.

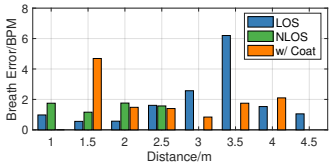


Figure 11: Breath estimation in NLOS and with coat.

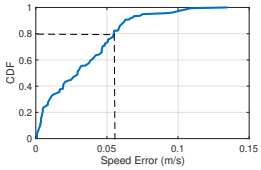


Figure 12: Speed estimation performance.

cility and SimNewB during the pandemic time, yet we believe the results already show the capability to detect a newborn's breathing at a distance. We will further study the impact of breathing rates later. In another test case, we configure the neonatal simulator to move her forearms, for which we detect the motion for 100%.

Speed Estimation: We also conduct a preliminary evaluation of our speed estimation by moving a plate back and forth at speeds from 15 cm/s to 25 cm/s along a 1.8m long straight programming track, as in Fig. 8(c). As shown in Fig. 12, VECARE achieves a considerable 80%ile accuracy of 5 cm/s. Errors mainly occur around the turning points when the plate slows to stop and starts moving again. We believe the results are encouraging and plan to further investigate statistical acoustic speed estimation for other applications in the future.

6.3 Real-World CPD Study

We conduct a real-world study with young children in different cars and parking scenarios, such as parking lots, roadside parking, garage, etc. For each child, we test different locations, with either forward-facing or rear-facing car seats as regulated. For older children who can sit/crawl independently, we also test seats without the baby car seat. All the children wear their regular winter coats. We test motion (awake) cases for every child and evaluate breathing for children who are able to get asleep (or stay very quiet) during the test. The data collection for each child lasts about 30-60 minutes. During tests, the car is parked and locked with windows closed, the typical scenario that hot-car deaths may occur. There are cars parking around and/or passing by, and parents and our experimenters talking/standing/walking around the car. There are frequent traffic noises during most of the tests, done in central downtown Hong Kong. In total, we have 15 children (aged 7 months, 12 months, 18 months, 2 (2×), 3 (4×), 4 (2×), 5 (3×), and 10 years old, respectively) tested in 7 different cars, including Lexus LS430, BMW 330, Mercedes-Benz C200, Mercedes-Benz S320, Tesla Model 3, Honda Jazz, Nissan Serena. We use one or two speakers for the real-world study, considering not all cars have four or more, and always use one single microphone. In most cases, the LOS condition is occluded, provided that the devices are installed in the front row while the kids are seated in the back. Example setups are shown in Fig. 8(d) and (e).

We mainly focus on the overall presence detection rate for this CPD test. Fig. 14 shows that VECARE achieves an average detection rate of 98.8% with an average false alarm rate of 2.1% for all age groups of children. As expected, the detection rate for infants (one 7-month-old and one 12-month-old in our experiments) is relatively lower than older kids, but is still around 90%. The high performance is consistently achieved across different cars, varying from 95% to 100% with marginal differences, as portrayed in Fig. 15. The FARs in Fig. 14 vary slightly because different kids are tested in different cars that have different enclosure materials and in-car noise levels. False negatives are most likely to occur when there is a lack of awake motion and the infant's breathing is extremely weak.

Furthermore, we analyze the performance at different in-car locations. As shown in Fig. 16, we group the results based on where the child seats, i.e., driver seat (L-F), passenger seat (L-R), two back seats (L-B and R-B), as well as the case when the child is on the back row floor (B-G). As seen, VECARE maintains a consistently high detection rate and low false alarm rate across different locations. Overall, the results demonstrate VECARE's remarkable performance in real-world scenarios, promising its potential for practical adoption.

To further understand the detection coverage in a car, we use a small toy car, as shown in Fig. 8(f), to simulate tiny motions at nine different on-seat and on-floor locations. Two speakers are installed on the left and right front doors, respectively. As depicted in Fig. 17(a), VECARE achieves a 100% detection rate for all the 9 testing locations, using either two speakers or only one single speaker on the left or the right.

Long-term Study: Besides the high accuracy, it is also critical to study false alarms, especially over a long period in diverse noisy environments like busy streets, noisy garages, etc. We first note that the above real-world experiments were conducted in noisy urban areas (including noisy garages, busy streets, parking lots next to highways, etc) in downtown Hong Kong in the presence of cars, sirens, pedestrians, etc. To further understand the performance in different environments, we carry out a relatively long-term evaluation in the busy Beijing City. We park the car, without kids inside, in a busy garage and a crowded street for about 10 hours, respectively. We report a false alarm if motion is detected for over 10% of the time for a sliding window of 2s. Our results show that VECARE observes a false alarm rate of 0.12% in the garage and 0.28% for the roadside parking case. In practice, a CPD system may not need to run for a long time, but perhaps only for a few minutes after the car is parked and locked, which will further reduce the chance of observing false alarms.

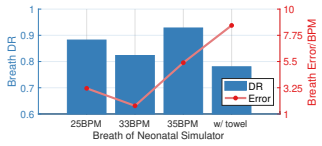


Figure 13: Evaluation with a neonatal simulator.

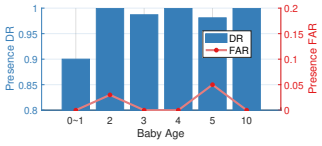


Figure 14: VECARE for children at different ages.

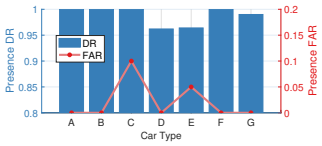


Figure 15: VECARE for children in different cars.

Figure 16: VECARE for different in-car locations.

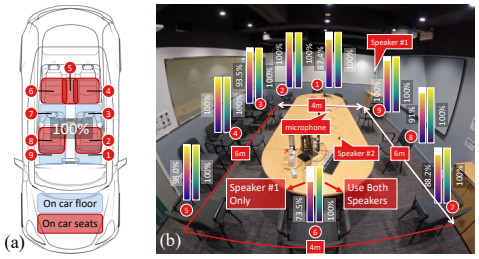


Figure 17: Motion coverage (a) in a car and (b) in a room.

System Latency: As a time-critical task, we now analyze the detection latency of VECARE. To do so, we evaluate the delay of the first decision for each test. We use a 3s window for ACF calculation for motion and an 8s window for breathing, and then use another 2s window for presence detection. Hence, the minimum delay will be 5s if motion is detected and 10s if there is no motion but breathing. With this configuration, the results show that VECARE can output the first detection within 5.7s for 81.9% of the time, 11.2s for 95.2%, and 15.2s for 98.8%. The minimum delays and thus the overall latency can be reduced by using a shorter window (e.g., 1s) for motion detection, the most common case for CPD.

6.4 Comparative Study

Baseline Comparison: We compare VECARE with the state-of-the-art approach BreathJunior [71], the closest to our work which successfully uses white noise for infant breathing monitoring. We implement BreathJunior and perform comparison experiments using an infant simulator. The results demonstrate that VECARE outperforms BreathJunior in both accuracy and coverage. As shown in Fig. 18, the maximum distance BreathJunior achieved is 70 cm (with a considerable error of 8 BPM), while VECARE goes to 1.6 meters under the same condition, which is 2.3× improvement. In addition, while BreathJunior is accurate within 0.5 m, the breathing estimation error quickly increases at a distance of 70 cm. In comparison, VECARE maintains a breathing rate error below 2 BPM at a distance of 90 cm, smaller than the error BreathJunior experiences at 60 cm.

Channel Estimation Methods: As said, VECARE can work with any channel estimation methods that output CIR. We now compare the performance of using Kasami Sequence against using different CIR estimation methods, including chirp signals (FMCW) [15], Golay Sequence [64], MLS [59], Gold Sequence [19]. As shown in Fig. 19, while all these methods produce a high detection rate above 90%, Kasami Sequence demonstrates its superior performance with the lowest breathing rate error and the highest detection rate.

Device Diversity: We now examine VECARE's performance on different devices. We are most interested in how it works on commodity car speakers and microphones. We thus evaluate it using JBL Stage1 621 car speaker and Linhuipad car microphone, both adopted in existing automobile audio systems. As shown in Fig. 20, VECARE maintains high performance and large coverage. We further test motion detection at 2m on various speaker/microphone combinations as summarized in Fig. 21, which indeed show device diversity yet good performance retains in most cases.

6.5 Benchmark Study

In this section, we evaluate the impacts of various factors and validate the robustness of VECARE. For more controllable data collection, we use the infant simulator instead of real babies for this study, and focus more on breathing estimation.

Background Interference Type: We first study the impact of background sound interference of different types, including human voices, traffic noise, rain sound, wind sound, hailstone sound, and music. To better control the experiments, we download sound files of these noises and play them through a loudspeaker around 50 dB next to the VECARE system. As shown in Fig. 22, VECARE maintains high accuracy regardless of different types of natural sound interference, with only marginal differences among them. This has also been partly verified in our real-world testing in §6.3 where we tested under real environments with all different ambient noises.

Background Interference Level: We also evaluate the performance under various background noise levels. We mainly focus on traffic noise and human voices for this test. We play sound files of noise through a loudspeaker at various powers and distances and record the actual sound level received *at the microphone*. As shown in Fig. 23, the BPM error increases with higher surrounding noises, especially over 50 dB level.

Transmitter Sound Level: The transmitting power of the speaker can affect performance. To verify this, we vary the transmitted sound from 46 dB to 53 dB and evaluate the breathing estimation error accordingly. As seen in Fig. 24, the breathing rate error quickly drops from about 7 BPM to below 2 BPM when the sound level exceeds 49 dB. Sensing sound at this level is perceived acceptable, according to our observations of the response of children participants and their parents' feedback, and users outside the car can barely hear the sound. Also, note that previous works [71] use higher sound levels (reportedly 56 dB [71] and 75 dB [70]) than

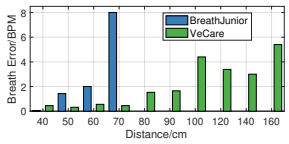


Figure 18: Performance comparison with baseline.

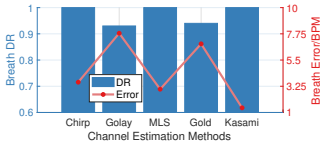
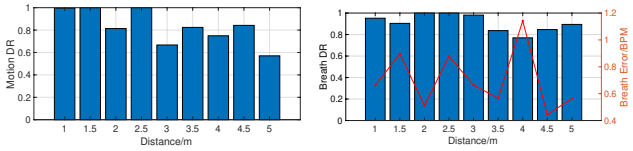


Figure 19: Comparing channel estimation methods.



(a) Motion detection. FAR=3%. (b) Breathing estimation

Figure 20: Performance using car speaker and mic.

Speaker \ Mic		Linhuipad	MiniDSP	Razer	Macbook	iPhone
JBL Stage1 Car Speaker	DR	1.00	1.00	1.00	1.00	1.00
	FAR	0.00	0.00	0.00	0.00	0.00
PUI Audio	DR	1.00	1.00	1.00	1.00	0.96
	FAR	0.00	0.00	0.00	0.00	0.18
JBL Clip4	DR	0.94	0.98	1.00	1.00	0.87
	FAR	0.07	0.00	0.00	0.00	0.03
iPhone	DR	0.92	1.00	0.87	1.00	1.00
	FAR	0.06	0.00	0.03	0.00	0.00
Sony SRS-XB23	DR	1.00	1.00	0.68	0.62	0.88
	FAR	0.00	0.00	0.16	0.41	0.28

Figure 21: Motion detection on various devices.

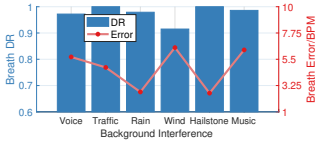


Figure 22: Impact of background interference.

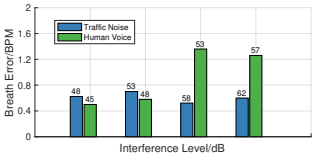


Figure 23: Impact of background interference level.

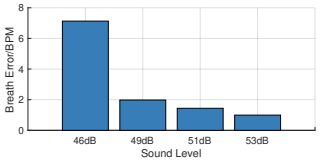


Figure 24: Impact of transmitting sound level.

VECARE. Nevertheless, a higher sound level is more favored in VECARE as a relatively high sound level could benefit CPD applications since it promises a better chance to wake up a sleeping baby for more reliable detection via awake motion. As research [16, 51] reports that a sound level higher than 75 dB will disturb the infants, we set the default sound level as 50 dB. Based on our real-world experiments with children, such a sound level appears to be tolerable to kids including infants and toddlers: We received no cases to complain about the sound intrusiveness and annoyance.

Frequency Bandwidth: We use the band above 10 kHz by default in our experiments, which may still be intrusive to human ears. We now study the performance with narrower and higher frequency bands. To do so, we adapt the passband of the high-pass filter from 10 kHz to 22 kHz with a 2 kHz step. Fig. 25 shows that VECARE retains a good performance until the passband exceeds 20 kHz. Larger bandwidths allow better performance, while VECARE still performs well using 18-24 kHz, the commonly used inaudible band in the literature. Commodity devices like Google Nest start to support acoustic frequencies up to 30 kHz [70], which we believe will become more common in the future. Such devices allow a sufficiently large and truly inaudible band across the age spectrum for non-intrusive acoustic sensing.

Impact of Temperature: As sound speed depends on temperature, we are curious how VECARE works under high temperatures. To do so, we heat up the surrounding air to about 120°F and then let it naturally cool down in a warm room of about 70°F. We keep the system running during the process and show the breathing estimation results in Fig. 26. As seen, VECARE failed to work when the devices overheat, but resumes excellent performance when they slightly cool down (after 30 seconds). We argue that a CPD system is expected to work *before* rather than *after* the car has heated up, as intervention actions are most effective right after the car is parked and locked. Therefore, we believe VECARE's CPD effectiveness will not be affected even though its performance degrades under overly high temperatures.

Multiple Speakers: Multiple speakers, if available, can further increase the sensing coverage. We present a case study with two speakers in a meeting room, where an adult sits in a chair moving one hand. As portrayed in Fig. 17(b), while the coverage with a single speaker is already good, by adding one speaker, we achieve a 100% motion detection rate through the 6m×4m area. We didn't continue with more speakers as the system already covers the entire room using two.

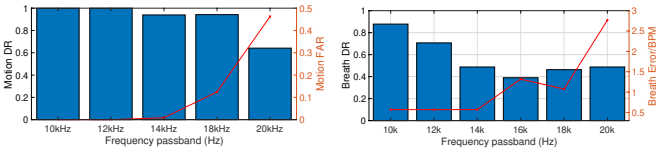
Synchronization Errors: We manually introduce large synchronization errors to show that VECARE is resilient to phase offsets. Particularly, we shift the starting point of the received signals by an amount of time ranging from 0 to 1 second with a step of 0.1 s. As shown in Fig. 27, VECARE maintains similar accuracy without being affected by the time offsets, which confirms our theoretical analysis.

Breath Intensity: Infants and toddlers usually have higher breathing rates than adults. We evaluate the performance of VECARE with respect to a range of breathing rates from 30 BPM to 60 BPM. We fix the breathing rate for each run by controlling the motor of our infant simulator. The results show insignificant differences for various breathing rates.

System Overhead: We benchmark the system overhead on a desktop (Intel i7-11700 @ 4.9GHz CPU), a MacBook Air M1, and a Raspberry Pi 3 Model B+, on which VECARE use 0.52s, 0.73s, and 3.97s respectively to process 10s of the data stream. The results show that VECARE can run in real-time on embedded devices, promising its integration into existing car control systems. The current prototype of VECARE using MiniDSP microphones introduces an extra power consumption of about 3W on Raspberry PI 3 B+, resulting in a total of 6W. The energy consumption can be optimized by improved hardware and software implementation. Additionally, the power usage is overall negligible as, again, we believe the CPD system can run only for minutes after a car is parked.

7 Discussions and Limitations

VECARE takes an important and promising step towards a ubiquitous solution to accurate and robust in-car CPD, an



(a) Motion detection (b) Breathing estimation

Figure 25: Impact of frequency band

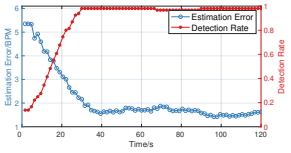


Figure 26: Impact of temperature (cooling from 120 °F)

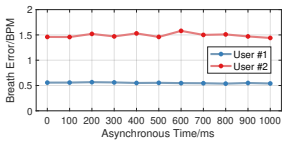


Figure 27: Impact of synchronization errors

extremely challenging task. However, there is certainly room for improvement and more to explore.

First, we cannot differentiate between an adult, a child, or a pet. Neither do we distinguish a single subject from multiple ones, as the proposed SAS currently is limited to one single user. A CPD system is expected to detect the presence of one or more children or pets, and the case of an adult being locked in a car is uncommon. While it is possible to distinguish between a child and an adult, to some extent, by examining the range of the breathing rates, we leave this task as an open challenge for the community. Second, although the proposed SAS model presents a new approach to speed estimation, the capability to estimate high speed is limited by the sampling rate of acoustic CSI. It is worth exploring how to break down this limit and enable speed tracking for normal walking speed, which will foster many applications. Third, there are more applications of the proposed SAS model in smart homes and non-contact healthcare to be explored. Towards that, one particular problem is to further improve the sound audibility, including reducing transmit sound level and shifting more to the inaudible frequency bands, e.g., 20-30 kHz used in commercial smart speakers like Amazon Echo and Google Home. Fourth, while a universal threshold ε applies to different environments and cars, we notice one-time calibration is needed for different devices due to hardware diversity (Fig. 21). Future work explores to relax it. Last, current evaluation is limited to children older than 7 months. Although we have experimented with the neonatal simulator, evaluation with real newborns is a worthwhile exploration.

VeCare can be deployed without any hardware modification: It works with legacy in-car audio systems, and we can leverage existing car control units for the computation and necessary alert functions. Alternatively, it can also be deployed as a standalone system to promote rapid adoption. Additionally, system integration with other available sensing modalities, such as seat sensors, radar, or WiFi, would promise a more reliable CPD system, but at a higher cost and lower ubiquity. These issues are, however, out of the scope of this work, and we seek to learn more about real deployment with industrial collaborations.

8 Related Work

Car Occupancy Sensing: CPD, or car occupancy detection in general, has recently gained tremendous attention with various technologies being explored. Early systems install special sensors, such as optical/weight/pressure sensors [12, 32] as well as capacitive sensors [4, 18, 54], on passenger seats/baby car seats for detection (e.g., for seat belt reminder). These systems cannot detect a left-in-car child not in a designated seat. Nor can they reliably differentiate animate targets from inanimate objects. PIR sensors [21, 55, 87] can extend the range beyond the seats, but are limited to the LOS view and sensitive to temperature changes. Camera-based systems [10, 13, 85] can be accurate given good lighting conditions[3], but cannot see through seats, in addition to being privacy-invasive and computation hungry. Radio-based systems have been recently popular. The radar industry is promoting radar systems for CPD in new car models [26, 28, 66]. UWB and mmWave radars [8, 14, 34] feature high sensing resolution, yet the coverage is limited to the FoV and the performance degrades for NLOS scenarios. Moreover, they need precise installation with wire/cable harnesses, usually on the roof of a car, to provide good coverage. With WiFi becoming prevalent for in-car connectivity, it has been exploited by the industry for CPD [45, 81]. Yet due to the innate limitations, it is challenging to detect vital signs of little infants using WiFi. Ultrasonic motion sensors have been used, e.g., in Hyundai cars [20], yet only report low detection accuracy and are being replaced [20]. Most importantly, all these solutions will incur additional dedicated hardware and/or costly installation as being non-standard offerings in cars, preventing their wide adoption, especially for old car models.

Beyond CPD, there are more works on in-car driver monitoring [68, 72, 80, 82, 83] and/or (adult) passenger detection [1, 2, 40] and even more on general human presence detection in homes especially using radio signals [30, 61, 77, 84, 89]. As the motion and vital signs of adults are of magnitude stronger than those of young children, these approaches cannot be directly applied to solve CPD. The closest to VeCare is BreathJunior [71], which nicely embeds FMCW signals into white noise for infant breathing monitoring, yet relies on a microphone array and has a limited coverage of $<$1m.

Acoustic Sensing: Acoustic sensing has been an active area in recent years. Various applications have been studied, including gesture recognition [74, 75], imaging [37, 42], localization and tracking [17, 41, 43, 56, 65, 69, 73, 86], vital sign monitoring [36, 53, 70, 71, 82], and healthcare [11, 46, 47], etc. A recent work [35] points out several practical challenges of acoustic sensing, such as audible sound leakage, affecting

[3] Infrared cameras can detect humans at night, but could still fail in dangerous cases where the car interior is already heated up.

music play and voice call, which are less concerned in our CPD application as the system is expected to only run for a short period after a car is parked. Most of the existing works explore geometrical features such as phase changes, Doppler shifts, TDoA/ToA, etc [9, 53, 74]. Many even rely on a bulky microphone array [69, 71] for phased signal processing. While these works are mostly resilient to multipaths, they do not fully leverage them. In contrast, VECARE investigates a novel statistical acoustic sensing model, which aims to leverage all reflections. Despite extensive studies of statistical WiFi sensing [23, 78, 79, 88–90] and statistical studies on acoustic communication [33, 62], none of the existing work has studied statistical acoustic sensing. We introduce statistical models to acoustic sensing, which we believe will open new directions and inspire follow-up works in the community.

9 Conclusion

We present VECARE, the first CPD system using in-car speakers and microphones. VECARE is an accurate and robust solution to the critical hot car death problem, which can be deployed on massive cars without any hardware changes. To achieve so, we introduce a novel paradigm of statistical acoustic sensing and develop a pipeline of techniques that allows motion detection, breathing estimation, and speed monitoring in a unified framework. Real-world experiments show the remarkable performance of VECARE, rendering it a promising solution in practice. The proposed SAS will inspire more exciting research in the increasingly hot acoustic sensing area.

Acknowledgments

We are grateful to all children participants in the real-world CPD study and their parents. We thank the School of Nursing at HKU for free access to the Laerdal SimNewB simulator. Thanks also go to the anonymous reviewers and to our shepherd, Nirupam Roy. This work is supported in part by NSFC under grant No. 62222216, No. 61832010 and No. 62202262, and Hong Kong RGC ECS under grant 27204522.

References

[1] Hajar Abedi, Clara Magnier, Vishvam Mazumdar, and George Shaker. Improving passenger safety in cars using novel radar signal processing. *Engineering Reports*, page e12413, 2021.

[2] Hajar Abedi, Clara Magnier, and George Shaker. Passenger monitoring using ai-powered radar. In *Proceedings of the IEEE International Symposium on Antenna Technology and Applied Electromagnetics*, pages 1–2, 2021.

[3] National Highway Traffic Safety Administration. Prevent hot car deaths. https://www.nhtsa.gov/campaign/heatstroke, 2022. Accessed: Sep 2022.

[4] Joan Albesa and Manel Gasulla. Occupancy and belt detection in removable vehicle seats via inductive power transmission. *IEEE Transactions on Vehicular Technology*, 64(8):3392–3401, 2014.

[5] PUI Audio. https://www.puiaudio.com/products/as07104por, 2022. Accessed: Jan 2022.

[6] John R. BarryEdward, A. LeeDavid, and G. Messerschmitt. *Digital Communication*. Springer, Boston, MA, 2004.

[7] Plux biosignalsplux piezoelectric Respiration (PZT) sensor. https://plux.info/sensors/316-respiration-pzt.html, 2022. Accessed: Jan 2022.

[8] A Caddemi and E Cardillo. Automotive anti-abandon systems: A millimeter-wave radar sensor for the detection of child presence. In *Proceedings of the IEEE International Conference on Advanced Technologies, Systems and Services in Telecommunications*, pages 94–97, 2019.

[9] Chao Cai, Rong Zheng, and Jun Luo. Ubiquitous acoustic sensing on commodity iot devices: A survey. *IEEE Communication Survey and Toturials*, page to appear, 2022.

[10] Haibin Cai, Donghee Lee, Hwang Joonkoo, Yinfeng Fang, Song Li, and Honghai Liu. Embedded vision based automotive interior intrusion detection system. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 2909–2914. IEEE, 2017.

[11] Justin Chan, Thomas Rea, Shyamnath Gollakota, and Jacob E Sunshine. Contactless cardiac arrest detection using smart devices. *NPJ digital medicine*, 2(1):1–8, 2019.

[12] Charles J Cole. System to detect the presence of an unattended child in a vehicle, January 30 2007. US Patent 7,170,401.

[13] Michel Devy, Alain Giralt, and Antonio Marin-Hernandez. Detection and classification of passenger seat occupancy using stereovision. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 714–719, 2000.

[14] Andreas R Diewald, Jochen Landwehr, Dimitri Tatarinov, Patrick Di Mario Cola, Claude Watgen, Catalin Mica, Mathieu Lu-Dac, Peter Larsen, Oscar Gomez, and Thierry Goniva. Rf-based child occupation detection in the vehicle interior. In *Proceedings of the IEEE International Radar Symposium*, pages 1–4, 2016.

[15] Angelo Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[16] ROLAND GÄDEKE, BERNHARD DÖRING, FRIEDRICH KELLER, and ANDRES VOGEL. The noise level in a childrens hospital and the wake-up threshold in infants. *Acta Paediatrica*, 58(2):164–170, 1969.

[17] Nakul Garg, Yang Bai, and Nirupam Roy. Owlet: Enabling spatial information in ubiquitous acoustic devices. In *Proceedings of the ACM MobiSys*, pages 255–268, 2021.

[18] Boby George, Hubert Zangl, Thomas Bretterklieber, and Georg Brasseur. Seat occupancy detection based on capacitive sensing. *IEEE Transactions on Instrumentation and Measurement*, 58(5):1487–1494, 2009.

[19] Robert Gold. Optimal binary sequences for spread spectrum multiplexing (corresp.). *IEEE Transactions on information theory*, 13(4):619–621, 1967.

[20] Hyundai Motor Group. The new radar-based occupant alert system to keep your children safe. https://tech.hyundaimotorgroup.com/article/the-new-radar-based-occupant-alert-system-to-keep-your-children-safe/, 2020. Accessed: Sep 2022.

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

[22] David A Hill. *Electromagnetic fields in cavities: deterministic and statistical theories*, volume 35. John Wiley & Sons, 2009.

[23] Yuqian Hu, Feng Zhang, Chenshu Wu, Beibei Wang, and KJ Ray Liu. Defall: Environment-independent passive fall detection using wifi. *IEEE Internet of Things Journal*, 9(11):8515–8530, 2021.

[24] Origin Wireless Inc. Hex home: Redefining smart home security. https://myhexhome.com/, 2022. Accessed: Sep 2022.

[25] Origin Wireless Inc. Origin health - remote patient monitoring. https://www.ces.tech/Innovation-Awards/Honorees/2021/Best-Of/O/Origin-Health-Remote-Patient-Monitoring.aspx, 2022. Accessed: Sep 2022.

[26] Infineon. Infineon in-cabin monitoring. https://www.infineon.com/cms/en/tools/aurix-embedded-sw/AURIX-Applications-software/in-cabin-monitoring/, 2021. Accessed: Sep 2022.

[27] UnivDatos Market Insights. Child presence detection system market: Current analysis and forecast (2019-2025). https://univdatos.com/report/child-presence-detection-system-market-current-analysis-and-forecast-2019-2025/, 2019. Accessed: Sep 2022.

[28] Texas Instruments. Vehicle occupant detection reference design. https://www.ti.com/lit/ug/tidue95a/tidue95a.pdf, 2020. Accessed: Sep 2022.

[29] JBL. Jbl stage1 621 two way car speaker. https://www.jbl.com.my/car-speakers/STAGE1+621.html, 2022. Accessed: Sep 2022.

[30] Avinash Kalyanaraman, Elahe Soltanaghaei, and Kamin Whitehouse. Doorpler: A radar-based system for real-time, low power zone occupancy sensing. In *Proceedings of the IEEE RTAS*, pages 42–53, 2019.

[31] Tadao Kasami. Weight distribution formula for some class of cyclic codes. *Coordinated Science Laboratory Report no. R-285*, 1966.

[32] K. N. Khamil, S. Rahman, and M. Gambilok. Babycare alert system for prevention of child left in a parked vehicle. *ARPN Journal of Engineering and Applied Sciences*, 10(22):17313–17319, 2015.

[33] Heinrich Kuttruff. *Room acoustics*. CRC Press/Taylor & Francis Group, Boca Raton, sixth edition edition, 2017.

[34] Antonio Lazaro, Marc Lazaro, Ramon Villarino, and David Girbau. Seat-occupancy detection system and breathing rate monitoring based on a low-cost mm-wave radar at 60 ghz. *IEEE Access*, 9:115403–115414, 2021.

[35] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. Experience: Practical problems for acoustic sensing. In *Proceedings of ACM MobiCom*, 2022.

[36] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. Lasense: Pushing the limits of fine-grained activity sensing using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–27, 2022.

[37] David B Lindell, Gordon Wetzstein, and Vladlen Koltun. Acoustic non-line-of-sight imaging. In *Proceedings of the IEEE/CVF CVPR*, pages 6780–6789, 2019.

[38] Linhuipad. Linhuipad car stereo microphone. https://www.amazon.co.uk/LINHUIPAD-Microphone-External-Compatible-Navigation-Black/dp/B08LR48PZK, 2022. Accessed: Sep 2022.

[39] LinkSys. Linksys aware: Introducing the first-ever mesh wifi motion sensing technology. https://www.linksys.com/us/linksys-aware/, 2022. Accessed: Sep 2022.

[40] Yongsen Ma, Yunze Zeng, and Vivek Jain. Carosense: Car occupancy sensing with the ultra-wideband keyless infrastructure. *Proceedings of the ACM IMWUT*, 4(3):1–28, 2020.

[41] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. Deeprange: Acoustic ranging via deep learning. *Proceedings of the ACM IWMUT*, 4(4):1–23, 2020.

[42] Wenguang Mao, Mei Wang, and Lili Qiu. Aim: Acoustic imaging on a mobile. In *Proceedings of the ACM MobiSys*, pages 468–481, 2018.

[43] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. Indoor follow me drone. In *Proceedings of the ACM MobiSys*, pages 345–358, 2017.

[44] UMA-8-SP USB mic array. https://www.minidsp.com/products/usb-audio-interface/uma-8-sp-detail, 2022. Accessed: Jan 2022.

[45] Murata. Wi-fi sensing for child presence detection (cpd). https://solution.murata.com/en-eu/technology/child-presence-detection, 2021. Accessed: Sep 2022.

[46] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Jacob E Sunshine. Opioid overdose detection using smartphones. *Science translational medicine*, 11(474):eaau8914, 2019.

[47] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 45–57, 2015.

[48] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of ACM CHI*, pages 1515–1525, 2016.

[49] European New Car Assessment Programme (Euro NCAP). Test and assessment protocol - child presence detection. https://cdn.eurocap.com/media/64101/euro-ncap-cpd-test-and-assessment-protocol-v10.pdf, 2021. Accessed: Sep 2022.

[50] noheatstroke.org. Heatstroke deaths of children in vehicles. https://www.noheatstroke.org/, 2022. Accessed: Sep 2022.

[51] M Kathleen Philbin. The influence of auditory experience on the behavior of preterm newborns. *Journal of perinatology*, 20(1):S77–S87, 2000.

[52] Markku Pukkila. Channel estimation modeling. *Nokia Research Center*, 17:66, 2000.

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

[54] Abhishek Ranjan and Boby George. A child-left-behind warning system based on capacitive sensing principle. In *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference*, pages 702–706. IEEE, 2013.

[55] Fairuz RM Rashidi and Ikhwan H Muhamad. Vehicle's interior movement detection and notification system. *Recent advances in automatic control, modelling and simulation*, pages 139–144, 2013.

[56] Mirco Rossi, Julia Seiter, Oliver Amft, Seraina Buchmeier, and Gerhard Tröster. Roomsense: an indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference*, pages 89–95, 2013.

[57] D.V. Sarwate and M.B. Pursley. Crosscorrelation properties of pseudorandom and related sequences. *Proceedings of the IEEE*, 68(5):593–619, 1980.

[58] M. R. Schroeder. Integrated-impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*, 66(2):497–500, 1979.

[59] Manfred R Schroeder. Integrated-impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*, 66(2):497–500, 1979.

[60] Signify. Wiz enables its products with new motion detection technology. https://www.signify.com/en-us/our-company/news/press-releases/2022/20220916-surprisingly-thoughtful-lighting-for-everyone-signify-brings-to-light-motion-detection-technology-spacesense, 2022. Accessed: Sep 2022.

[61] Elahe Soltanaghaei, Rahul Anand Sharma, Zehao Wang, Adarsh Chittilappilly, Anh Luong, Eric Giler, Katie Hall, Steve Elias, and Anthony Rowe. Robust and practical wifi human sensing using on-device learning with a domain adaptive model. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pages 150–159, 2020.

[62] Milica Stojanovic and James Preisig. Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE communications magazine*, 47(1):84–89, 2009.

[63] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of ACM MobiCom*, pages 591–605, 2018.

[64] Chin-Chong Tseng and C. Liu. Complementary sets of sequences. *IEEE Transactions on Information Theory*, 18(5):644–652, 1972.

[65] Yu-Chih Tung and Kang G Shin. Echotag: Accurate infrastructure-free indoor location tagging with smartphones. In *Proceedings of the ACM MobiCom*, pages 525–536, 2015.

[66] Vayyar. Vayyar child presence detection. https://vayyar.com/auto/solutions/in-cabin/cpd/, 2021. Accessed: Sep 2022.

[67] Verizon. Verizon launches new tech to monitor activity on home wifi. https://www.verizon.com/about/news/verizon-launches-new-tech-monitor-activity-home-wifi, 2023. Accessed: Feb 2023.

[68] Eric Wahlstrom, Osama Masoud, and Nikos Papanikolopoulos. Vision-based methods for driver monitoring. In *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, volume 2, pages 903–908. IEEE, 2003.

[69] Anran Wang and Shyamnath Gollakota. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the ACM CHI*, pages 1–11, 2019.

[70] Anran Wang, Dan Nguyen, Arun R Sridhar, and Shyamnath Gollakota. Using smart speakers to contactlessly monitor heart rhythms. *Communications biology*, 4(1):1–12, 2021.

[71] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. Contactless infant monitoring using white noise. In *Proceedings of the ACM MobiCom*, pages 1–16, 2019.

[72] Fengyu Wang, Xiaolu Zeng, Chenshu Wu, Beibei Wang, and KJ Ray Liu. Driver vital signs monitoring using millimeter wave radio. *IEEE Internet of Things Journal*, 2021.

[73] Mei Wang, Wei Sun, and Lili Qiu. MAVL: Multiresolution analysis of voice localization. In *Proceedings of the IEEE USENIX NSDI*, pages 845–858, 2021.

[74] Wei Wang, Alex X Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the ACM MobiCom*, pages 82–94, 2016.

[75] Yanwen Wang, Jiaxing Shen, and Yuanqing Zheng. Push the limit of acoustic gesture recognition. *IEEE Transactions on Mobile Computing*, 2020.

[76] Safe Kids Worldwide. https://www.safekids.org/, 2022. Accessed: Sep 2022.

[77] Chenshu Wu, Zheng Yang, Zimu Zhou, Xuefeng Liu, Yunhao Liu, and Jiannong Cao. Non-invasive detection of moving and stationary human with wifi. *IEEE Journal on Selected Areas in Communications*, 33(11):2329–2342, 2015.

[78] Chenshu Wu, Feng Zhang, Yusen Fan, and K. J. Ray Liu. Rf-based inertial measurement. In *ACM SIGCOMM*, 2019.

[79] Chenshu Wu, Feng Zhang, Yuqian Hu, , and K. J. Ray Liu. Gaitway: Monitoring and recognizing gait speed through the walls. In *IEEE Transactions on Mobile Computing*, pages 2186–2199, June 2021.

[80] Xiufeng Xie, Kang G Shin, Hamed Yousefi, and Suining He. Wireless csi-based head tracking in the driver seat. In *Proceedings of the ACM CoNext*, pages 112–125, 2018.

[81] Qinyi Xu, Beibei Wang, Feng Zhang, Deepika Sai Regani, Fengyu Wang, and KJ Ray Liu. Wireless ai in smart car: How smart a car can be? *IEEE Access*, 8:55091–55112, 2020.

[82] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the ACM MobiSys*, pages 54–66, 2019.

[83] Jie Yang, Simon Sidhom, Gayathri Chandrasekaran, Tam Vu, Hongbo Liu, Nicolae Cecan, Yingying Chen, Marco Gruteser, and Richard P Martin. Detecting driver phone use leveraging car speakers. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 97–108, 2011.

[84] Yuzhe Yang, Yuan Yuan, Guo Zhang, Hao Wang, Ying-Cong Chen, Yingcheng Liu, Christopher G Tarolli, Daniel Crepeau, Jan Bukartyk, Mithri R Junna, Aleksandar Videnovic, Terry D Ellis, Melissa C Lipford, Ray Dorsey, and Dina Katabi. Artificial intelligence-enabled detection and assessment of parkinson's disease using nocturnal breathing signals. *Nature medicine*, 28(10):2207–2215, 2022.

[85] Hee Jung Yoon, Ho-Kyeong Ra, Can Basaran, Sang Hyuk Son, Taejoon Park, and Jeonggil Ko. Fuzzy bin-based classification for detecting children's presence with 3d depth cameras. *ACM Transactions on Sensor Networks*, 13(3):1–28, 2017.

[86] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the ACM MobiSys*, pages 15–28, 2017.

[87] Piero Zappi, Elisabetta Farella, and Luca Benini. Tracking motion direction and distance with pyroelectric ir sensors. *IEEE Sensors Journal*, 10(9):1486–1494, 2010.

[88] Feng Zhang, Chen Chen, Beibei Wang, and KJ Ray Liu. Wispeed: A statistical electromagnetic approach for device-free indoor speed estimation. *IEEE Internet of Things Journal*, 5(3):2163–2177, 2018.

[89] Feng Zhang, Chenshu Wu, Beibei Wang, Hung-Quoc Lai, Yi Han, and K. J. Ray Liu. Widetect: Robust motion detection with a statistical electromagnetic model. In *Proceedings of the ACM IMWUT*, Sep 2019.

[90] Feng Zhang, Chenshu Wu, Beibei Wang, Min Wu, Daniel Bugos, Hangfang Zhang, , and K. J. Ray Liu. Smars: Sleep monitoring via ambient radio signal. In *IEEE Transactions on Mobile Computing*, pages 217–231, Jan 2019.

A Appendix

A.1 Speed Estimation

Sampling Rate and Speed: A sufficient sampling rate of CSI is required to estimate speed. We now discuss the relationship. Given a sound frequency f with wavelength $\lambda(f)$, a moving speed v is expected to experience a peak at the delay of $\tau = \frac{x_0\lambda(f)}{2\pi v}$. Assume we will need at least Q samples to reliably detect a peak, which corresponds to a delay of $\tau_{\min} = Q/F_s$. Then we can derive the minimum sampling rate required to measure a speed of v by $\tau = \frac{x_0\lambda(f)}{2\pi v} > \tau_{\min} = Q/F_s$, which implies $F_s > \frac{2\pi Q v}{x_0\lambda(f)}$. In other words, the maximum speed we can support can be calculated as $v < \frac{x_0\lambda(f)F_s}{2\pi Q}$, which becomes about 0.1 m/s at $f = 20$kHz (wavelength 1.7 cm), about 0.2 m/s at $f = 10$ kHz, and about 2 m/s at $f = 1$kHz, assuming $Q = 5$ and a sampling rate of about 50 Hz (considering the sound speed of $c = 343$m/s). Using lower frequencies immediately allows to support higher speed, which however may suffer more from ambient noises. How to break down the sampling rate limitations and achieve estimation of daily speed (e.g., 0.5 m/s to 2 m/s) using pseudo-ultrasound frequencies remains worthwhile direction.

Speed MRC: For breathing signals, since the periodicity is independent of subcarrier frequency, we can directly perform

MRC across subcarriers. However, a further trick is needed to combine speed signals because, for acoustic signals from 10 kHz to 24 kHz, the difference in the wavelengths cannot be neglected (the wavelength at 10 kHz is approximately twice of that at 24 kHz). Recall $\hat{v} = \frac{x_0 \lambda(f)}{2\pi\tau_s}$. Given the same speed v, the first local peaks of the ACF on different subcarriers will appear at different delays τ_s. Hence, to combine subcarriers for speed signals, we need to first compensate the linear offsets due to different wavelengths. Specifically, we can express the ACF $\tilde{\rho}(f,\tau)$ w.r.t a unit linearly proportional to $\lambda(f)$, i.e., $\mu = \frac{\tau}{\lambda(f)}$, and then average on $\tilde{\rho}(f,\mu)$. The operation is equivalent to scaling the ACF in the time lag dimension, which can be achieved by interpolation in practice.

A.2 Synchronization

Here we show a simple but stringent proof of that synchronization errors do not affect VECARE. Denote the CIR measured under synchronization offsets as $\tilde{h}(t)$:

$$\tilde{h}(t) = circshift(h(t), \tau_{off}), \tag{7}$$

where $h(t)$ is the true CIR, τ_{off} is the timing offset caused by asynchronization, and $circshift(\cdot)$ represents circular shift. The time offsets correspond to phase shifts in the frequency domain. Thus we have the asychronized CSI:

$$\tilde{H}(f) = H(f) \cdot e^{-j2\pi f \tau_{off}}, \tag{8}$$

where $H(f)$ is the true CSI. Thus we get $|\tilde{H}(f)| = |H(f)|$, meaning that VECARE is resilient to sychronization errors.

A.3 ACF Outliers

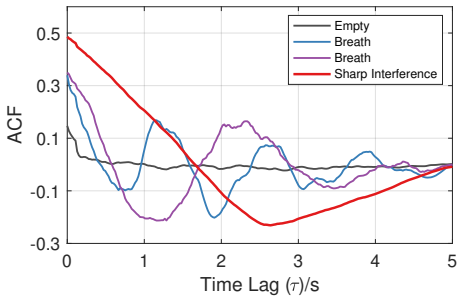


Figure 28: ACF outliers of sharp interference.

Sudden impulse-like noises will cause an abrupt status change in the CSI stream, which will smear the ACF calculation and may lead to false motion detection. We notice that ACF calculated with CSI of an abrupt status change will exhibit a special pattern, which linearly decreases and then linearly increases, as shown in Fig. 28. We also confirmed this by simulation with data stream containing a sudden change in the middle. The ACF pattern is unique and distinguishable from normal ACF in case of motion, breathing, or empty environment. Based on this observation, we design a detector to identify such abnormal ACFs and sift them out for presence detection. The idea is to examine a single valley in the ACF with linear increasing/decreasing trends on the two sides of the valley. The approach turns out to be effective and accurate.