Published: 18 June 2025

RESEARCH-ARTICLE

WindDancer: Understanding Acoustic Sensing under Ambient Airflow

KUANG YUAN, Carnegie Mellon University, Pittsburgh, PA, United States

DONG LI, University of Maryland, Baltimore County (UMBC), Baltimore, MD, United States

HAO ZHOU, Pennsylvania State University, University Park, PA, United States

ZHEHAO LI, Stanford University, Stanford, CA, United States

LILI QIU, The University of Texas at Austin, Austin, TX, United States

SWARUN KUMAR, Carnegie Mellon University, Pittsburgh, PA, United States

View all

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 9, Issue 2 (June 2025)
https://doi.org/10.1145/3729469
EISSN: 2474-9567

WindDancer: Understanding Acoustic Sensing under Ambient Airflow

KUANG YUAN, Carnegie Mellon University, USA
DONG LI, University of Maryland, Baltimore County, USA
HAO ZHOU, The Pennsylvania State University, USA
ZHEHAO LI, Stanford University, USA
LILI QIU, Microsoft Research Asia, China and UT Austin, USA
SWARUN KUMAR, Carnegie Mellon University, USA
JIE XIONG, Nanyang Technological University, Singapore

Acoustic sensing has recently garnered significant interest for a wide range of applications ranging from motion tracking to health monitoring. However, prior works overlooked an important real-world factor affecting acoustic sensing systems—the instability of the propagation medium due to ambient airflow. Airflow introduces rapid and random fluctuations in the speed of sound, leading to performance degradation in acoustic sensing tasks. This paper presents *WindDancer*, the first comprehensive framework to understand how ambient airflow influences existing acoustic sensing systems, as well as provides solutions to enhance systems performance in the presence of airflow. Specifically, our work includes a mechanistic understanding of airflow interference, modeling of sound speed variations, and analysis of how several key real-world factors interact with airflow. Furthermore, we provide practical recommendations and signal processing solutions to improve the resilience of acoustic sensing systems for real-world deployment. We envision that *WindDancer* establishes a theoretical foundation for understanding the impact of airflow on acoustic sensing, and advances the reliability of acoustic sensing technologies for broader adoption.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Computer systems organization → Sensors and actuators**.

Additional Key Words and Phrases: Wireless Sensing, Audio Signal Processing, Environmental Modeling, Airflow

ACM Reference Format:
Kuang Yuan, Dong Li, Hao Zhou, Zhehao Li, Lili Qiu, Swarun Kumar, and Jie Xiong. 2025. WindDancer: Understanding Acoustic Sensing under Ambient Airflow. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 2, Article 61 (June 2025), 25 pages. https://doi.org/10.1145/3729469

1 Introduction

Acoustic sensing has gained extensive attention from both academic and industrial communities over the past few years. Prior research has pushed the accuracy of acoustic sensing to sub-millimeter level, and enabled various applications from basic motion tracking [7, 31, 36] to more intricate tasks such as lip reading [29, 41], respiration and heartbeat monitoring [47, 50, 53, 57]. Beyond academia, emerging industrial products including Amazon Echo [1] and Google Nest [44] have incorporated basic acoustic sensing functions such as human motion sensing

Authors' Contact Information: Kuang Yuan, Carnegie Mellon University, Pittsburgh, USA, kuangy@andrew.cmu.edu; Dong Li, University of Maryland, Baltimore County, Baltimore, USA, dli@umbc.edu; Hao Zhou, The Pennsylvania State University, State college, USA, hfz5190@psu.edu; Zhehao Li, Stanford University, Stanford, USA, zhehaoli@stanford.edu; Lili Qiu, Microsoft Research Asia, Shanghai, China and UT Austin, Austin, USA, lili@cs.utexas.edu; Swarun Kumar, Carnegie Mellon University, Pittsburgh, USA, swarun@cmu.edu; Jie Xiong, Nanyang Technological University, Singapore, Singapore, jie.xiong@ntu.edu.sg.

ACM 2474-9567/2025/6-ART61
https://doi.org/10.1145/3729469

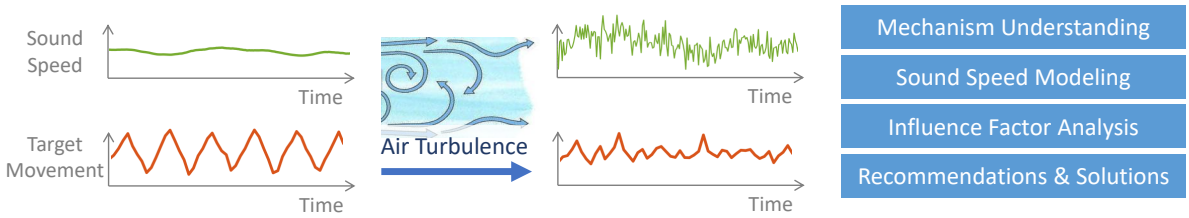


Fig. 1. Turbulence flow in the air introduces rapid and random fluctuations to sound speech, presenting unique challenges for existing acoustic sensing systems. In *WindDancer*, we take four key steps to approach the problem and tackle the challenge.

in them. Though promising, most of the current acoustic sensing systems are mainly evaluated in constrained lab settings and lack validation in real-world scenarios. Recent papers identify multiple real-world factors that affect acoustic sensing performance, such as device mobility [28], multi-target interference [20], and audible leakage [19]. While efforts have been made to mitigate these issues, in this paper, we study an unexplored intrinsic challenge of acoustic sensing for the first time—the impact of ambient airflow (i.e., wind), which results in an unstable propagation medium for acoustic sensing.

In fact, the disruptive effects of airflow on acoustics are commonly experienced in everyday life. Anyone making a phone call outdoors on a windy day has likely experienced degradation in audio quality. While prior solutions [13, 30, 34, 54] have studied and mitigated the effect of wind noise on voice, little attention has been paid to airflow's effects on the propagation of ultrasonic acoustic signals—the foundation of the state-of-the-art acoustic sensing systems. In a typical acoustic sensing system, an ultrasound wave is generated from a speaker, propagates through the air as the medium, gets reflected by the surrounding objects, and finally hits the microphone. The signal received is then processed to estimate the movement of the target, which usually is the reflector. Existing acoustic sensing systems assume that the propagation medium is stable and the speed of sound is generally a constant (about 343 m/s at room temperature).

However, as we transition acoustic sensing systems from lab settings to real-world applications, we find that this assumption does not always hold true. Instead of a stable medium, airflow, or wind, the movement of air can easily make the propagation medium of acoustic waves unstable and dynamic. Intuitively, if the acoustic wave is traveling in the same direction as the airflow, it will travel faster than its original speed. Conversely, if the sound is traveling against the wind, it will travel slower. More importantly, irregular air turbulence is induced when airflow interacts with the environment. Such turbulence can further introduce rapid and random fluctuations of the sound speed in an unpredictable manner.

In everyday life, airflow is ubiquitous in both indoor and outdoor environments, presenting unique challenges for acoustic sensing systems. In indoor environments, fans and air conditioners are widely used, which potentially degrade the performance of a wide range of acoustic sensing systems in smart home settings [1, 22, 33, 44] and health monitoring applications [47, 50, 53, 57]. Our study in Sec. 6 shows that a regular fan can increase the sensing error by 92% for object detection and by 172% for fine-grained movement tracking. One surprising yet interesting observation is that the extent of performance degradation is influenced not only by the wind but also strongly related to surrounding objects in the environment. For instance, the effect of airflow is significantly amplified in environments with strong reflectors.

To understand this intrinsic and unexplored challenge of wind effect on acoustic sensing and push the acoustic sensing systems from lab settings to real-world adoption, we propose *WindDancer*. In this paper, we do not present a narrowly focused system that simply "fixes" the effect of airflow for a particular acoustic sensing application. Instead, our goal is broader—we aim to develop a general framework that explains how ambient airflow impacts a range of existing acoustic sensing systems across various application domains, and also provide

solutions to enhance their resilience to airflow. We take the following steps to approach the problem and tackle the challenges:

Mechanistic Analysis of Interference: At first glance, the impact of typical household airflow on acoustic sensing seems counterintuitive. Air conditioners and fans generally produce wind speeds below 5 m/s—merely 1.5% of the sound speed. We are curious about how this level of wind can cause significant enough interference to affect acoustic sensing systems. To understand the phenomenon, we dive into the details of the mechanism of the commonly used FMCW-based acoustic sensing pipeline.

Most existing acoustic sensing systems rely on background subtraction to remove noise and track target movement [20, 33], based on a key assumption: signals reflected from static objects remain consistent and can be effectively canceled out. However, airflow—even at low speeds—creates turbulence that leads to rapid fluctuations in the speed of sound. While these fluctuations may not be huge in magnitude, they lead to continuously varying reflection signals from static objects. This violates the assumption to remove the effects of static reflectors and becomes the primary source of errors in both coarse-grained ranging and fine-grained motion tracking.

Modeling Speed of Sound Variations: To accurately model and analyze the impact of ambient airflow on acoustic sensing, we first conduct experiments to measure variations in the speed of sound under different wind conditions. From the results, we observe two key factors: the structured shift in the mean sound speed caused by laminar flow, and the unstructured fluctuations in sound speed due to turbulent flow. The latter component, turbulence, is the primary factor degrading acoustic sensing performance. Although this fluctuating sequence appears random, we found that its temporal behavior can be effectively characterized by an Autoregressive (AR) model [5]. The AR model offers a tool to parameterize the characteristics of airflow profiles, enabling a quantitative analysis of how these airflow characteristics affect acoustic sensing. We further employ physics-based modeling to validate the physical accuracy of using the AR model to characterize turbulent flow.

Influence Factor Analysis: We then conduct a comprehensive quantitative analysis of how various real-world factors interact with airflow to impact acoustic sensing performance, combining experimental measurements with simulations based on our model. Specifically, we examine various airflow characteristics, environmental conditions, system configurations, and sensing tasks. Our analysis reveals several counterintuitive findings. For example, we observe that wind speed is not the primary factor affecting acoustic sensing. Instead, the effects are primarily driven by the irregularity of turbulence and static reflectors in the environment. Other counterintuitive findings include that fine-grained acoustic sensing is less affected by ambient airflow compared to coarse-grained sensing. We detail these observations in Sec. 6.

Practical Recommendations and Technical Solutions: Based on the findings from our analysis of influencing factors, we provide practical recommendations for deploying acoustic sensing systems in real-world environments with ambient airflow. Moreover, to further mitigate the effects of airflow, we develop signal processing solutions to enhance the performance of coarse-grained and fine-grained sensing. Our proposed algorithm leverages a key observation: Although the speed of sound keeps fluctuating, these variations exhibit bounded magnitude and temporal correlation, as characterized by the AR model. This property results in a relatively stable and structured spectral pattern in the channel frequency response from static reflectors over time. Building on this observation, we develop two specialized algorithms: a spectral variation thresholding method to enhance the performance of coarse-grained sensing, and a Variational Mode Decomposition (VMD) approach to improve the performance of fine-grained sensing. To summarize, *WindDancer* makes the following contributions:

- For the first time, we identify ambient airflow as an unexplored and important noise source for acoustic sensing systems. We conducted a mechanism analysis on the FMCW-based signal processing pipeline to identify the key error source that degrades the sensing performance of existing acoustic sensing systems.

- We effectively model the speed of sound variations in the presence of airflow using the AR model and validate its physical fidelity through physics-based analysis. The AR model provides us a valuable tool to characterize the airflow profile and enable our simulation framework.
- We conduct a comprehensive analysis of influencing factors to understand the impact of various real-world factors. We distill our findings into four key observations and offer corresponding practical recommendations for the real-world deployment of acoustic sensing systems.
- We develop signal processing algorithms to enhance the performance of acoustic sensing under airflow. Our evaluation results show that the algorithms achieve 67.3% and 52.5% performance improvements in coarse-grained and fine-grained sensing respectively.
- To facilitate broader research on ambient airflow's effects on diverse applications and future acoustic sensing systems, we make our simulation frameworks and signal processing algorithms publicly available[1].

In the following sections, we first present the related works in Sec. 2, then discuss the background of the FMCW-based acoustic sensing pipeline in Sec. 3. We dive deeper into the pipeline and present the mechanism of how airflow affects acoustic sensing in Sec. 4. After that, Sec. 5 presents the speed of sound measurement experiments and modeling, which is the prerequisite of the detailed influencing analysis in Sec. 6 and the proposed solutions in Sec. 7. We discuss the limitations and conclude the work in Sec. 8 and Sec. 9 respectively.

2 Related Work

The research community has successfully leveraged acoustic signals emitted from commodity devices to capture detailed contextual information about both human activities and environmental states. Acoustic sensing has enabled a broad spectrum of applications, including contactless motion tracking [20, 33], gesture recognition [22, 25], health monitoring [9, 55, 57], eye blink detection [11, 27], gaze tracking [23], speech enhancement [41, 56], and even non-invasive measurements of environmental parameters such as temperature [6, 45]. These advancements allow for increasingly sophisticated and varied sensing capabilities, expanding the scope of acoustic sensing across domains such as healthcare, smart homes, human-computer interaction, and environmental monitoring.

In recent years, researchers have dedicated considerable effort to advancing acoustic sensing by enhancing the sensing accuracy [7, 46], extending the sensing range [21, 32, 33], and improving the sensing granularity [8, 10]. This progress has been driven by the development of advanced signal processing algorithms, machine learning models, and hardware innovations, all of which help minimize noise interference and more effectively isolate meaningful acoustic features. While some studies have addressed practical challenges, such as audible sound leakage [19] and device mobility [28], The impact of airflow on sound wave propagation, and consequently on acoustic sensing, remains relatively underexplored. Airflow can cause deviations in the direction, amplitude, and frequency of sound waves, potentially introducing variability and compromising the reliability of acoustic measurements [35]. This airflow-induced disruption presents unique challenges, especially in enclosed environments with air conditioning or ventilation systems, where fluctuations in airflow can complicate accurate acoustic sensing. Our work addresses this gap by exploring the impact of airflow on acoustic sensing and developing adaptive models to mitigate its effects on sensing performance.

Existing work has studied the use of ultrasound for airflow and turbulence measurement. Barth et al. demonstrated that ultrasound can be used to effectively capture wind speed profiles [3]. Wang et al. leveraged ultrasound to detect exhaled airflow patterns associated with breathing [49]. In addition, Yuan et al. showed that ultrasound can help reduce noise from airflow, particularly in turbulent conditions, thereby enhancing the stability of audio recordings [54]. However, these studies did not examine how measured airflow and turbulence impact acoustic sensing or propose solutions to mitigate airflow-related disruptions in acoustic sensing applications.

[1] https://github.com/kuangyuan-cmu/WindDancer

3 The Basics of FMCW-based Acoustic Sensing

Frequency Modulated Continuous Wave (FMCW) signals have been extensively used in acoustic sensing systems [20, 26, 33, 57]. This section introduces the basics of the FMCW signal processing pipeline, covering both coarse-grained and fine-grained sensing through mathematical formulations.

An FMCW signal, also known as a chirp signal, is a sine wave whose frequency sweeps linearly over time. As shown in Fig. 2a, its frequency can be expressed as $f = f_0 + \frac{B}{T}t$, where f_0 is the initial frequency, B is the chirp bandwidth and T is the duration of a single chirp. The chirp signal transmitted by a speaker can be denoted as:

$$x(t) = \exp\left(-j2\pi\left(f_0 t + \frac{B}{2T}t^2\right)\right). \tag{1}$$

The transmitted signal propagates through the air, gets reflected by surrounding objects, and is finally received by a microphone. As illustrated in Fig. 2b, since there exist multiple reflectors in the environment, the received signal $y(t)$ can be denoted as a summation of all the reflected chirps:

$$y(t) = \sum_{i=1}^{P} \alpha_i \exp\left(-j2\pi\left(f_0(t - \frac{d_i}{c}) + \frac{B}{2T}(t - \frac{d_i}{c})^2\right)\right), \tag{2}$$

where α_i and d_i are the signal attenuation and the length of the i-th reflection path, respectively. c is the speed of sound in the air, and P is the total number of reflection paths. To convert the received signal to baseband for further processing, we multiply the received signal $y(t)$ with the transmitted signal $x(t)$ and apply a low-pass filter. Fig. 2c illustrates the resulted mixed signal [20, 57]:

$$m(t) = \sum_{i=1}^{P} \alpha_i \exp\left(-j2\pi\left(\frac{Bd_i}{cT}t + \frac{f_0 d_i}{c}\right)\right). \tag{3}$$

We then apply Fast Fourier Transform (FFT) on the mixed signal $m(t)$ to transform it into Channel Frequency Response (CFR), which allows us to separate the reflections from different distances into various range bins [21]:

$$M(f_d) = \mathcal{F}\{m(t)\} = \sum_{i=1}^{P} \alpha_i \exp\left(j2\pi\frac{f_0 d_i}{c}\right), \tag{4}$$

where the frequency index $f_d = \frac{Bd}{cT}$ and the phase $\angle M(f_d) = 2\pi\frac{f_0 d}{c}$ contain the coarse-grained and fine-grained distance information of surrounding objects, respectively. In general, we use a sequence of chirp signals over a period of time to sense the target movement. We denote the CFR across multiple chirps as:

$$M[f_d, n] = \sum_{i=1}^{P} \alpha_i \exp\left(j2\pi\frac{f_0 d_i[n]}{c}\right), \tag{5}$$

where $n = 1 \cdots N$ represents the chirp index.

Coarse-grained Sensing: As visualized in Fig. 2d, the CFR magnitudes $|M[f_d, n]|$ represent the signal strengths of reflections from different range bins over time. However, we cannot directly identify the target from the raw CFR magnitude map. The reason is that the reflection from the target (e.g., hand) is typically weaker than those from surrounding static reflectors (e.g., wall or table). Futher, this issue becomes more pronounced for small-sized targets or targets located further away from the device [21]. As shown in Fig. 2e, prior systems employed the background subtraction method [20, 33] to remove the impact of static reflectors. Without loss of generality, we assume there is only one moving target whose range bin index is p. Mathematically, to perform background

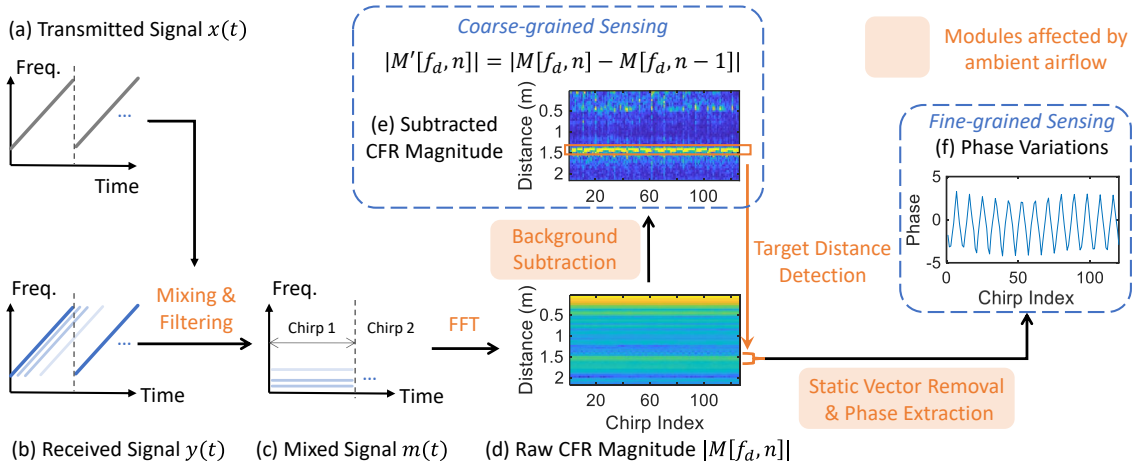


Fig. 2. The FMCW-based acoustic sensing pipeline for coarse-grained and fine-grained sensing. Two necessary modules in the pipeline would be impacted by ambient airflow.

subtraction, we calculate the difference of CFRs between two consecutive chirp signals:

$$\begin{aligned} M'[f_d, n] &= M[f_d, n] - M[f_d, n-1] \\ &= \sum_{i=1}^{P-1} \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c}\right) + \alpha_p \exp\left(j2\pi \frac{f_0 d_p[n]}{c}\right) - \sum_{i=1}^{P-1} \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c}\right) - \alpha_p \exp\left(j2\pi \frac{f_0 d_p[n-1]}{c}\right) \\ &= \alpha_p \exp\left(j2\pi \frac{f_0 d_p[n]}{c}\right) - \alpha_p \exp\left(j2\pi \frac{f_0 d_p[n-1]}{c}\right). \end{aligned} \tag{6}$$

The static reflection paths ($i = 1, \ldots, P-1$) are subtracted out based on the fact that they remain the same over time, i.e., $d_i[n] = d_i[n-1]$. As a result, the reflection from the moving target stands out in the subtracted CFR magnitude map. By selecting the range bin index corresponding to the peak along the chirp dimension, we can easily extract coarse-grained distance information, which can be used for various sensing tasks, such as hand tracking [33] and gesture recognition [22].

Fine-grained Sensing: For certain sensing tasks, such as respiration monitoring [50], heartbeat monitoring [57], and eye blinking detection [27], the scale of the target movements can be smaller than the range resolution of the CFR magnitude map.

To achieve fine-grained sensing, prior systems extracted the phase information from the raw CFRs at the target range bin [21, 27, 57]. Based on Eq. (5), the CFR at the target range bin can be viewed as a combination of multiple vectors in the I-Q space. Specifically, it is a summation of a constant vector (from static reflectors) and a dynamic vector (from the moving target), which can be denoted as:

$$M[\hat{f}_{d_p}, n] = M_{d_p}[n] + M_s = \exp\left(j2\pi \frac{f_0 d_p[n]}{c}\right) + \sum_{i=1}^{S} \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c}\right), \tag{7}$$

where $\hat{f}_{d_p}$ is the frequency index at the target distance, M_s is the combination of the vector from S static reflectors, and $M_{d_p}[n]$ is a dynamic vector induced by the target movement. By removing the static vector using methods such as circle fitting [21] or moving average [51], we can obtain the phase variations of the dynamic vector $\Delta\varphi$,

as illustrated in Fig. 2f. Then the fine-grained target displacement Δd can be derived as:

$$\Delta d = d[n] - d[n-1] = \frac{c\Delta\varphi}{2\pi f_0} = \frac{c(\angle M_{d_p}[n] - \angle M_{d_p}[n-1])}{2\pi f_0}. \tag{8}$$

4 Understanding the Mechanisms of Airflow Effects on Acoustic Sensing

We now explain how ambient airflow interferes with both coarse-grained and fine-grained acoustic sensing. Airflow, defined as the movement of air, directly influences sound propagation, which can be generated by various environmental and mechanical sources, including natural wind, air conditioners, fans, etc. Typically, both laminar and turbulent airflow are present at the same time. Laminar flow causes sound to travel faster in its direction and slower against it, while turbulent flow introduces unstructured fluctuations in the speed of sound. As a result, unlike previous systems that assume a constant sound speed c, we must treat the sound speed as a time-varying function $c(t)$ in the presence of ambient airflow.

Recall from Sec. 3 that the coarse-grained and fine-grained distances can be estimated using $d = \frac{cf_dT}{B}$ and $\Delta d = \frac{c\Delta\varphi}{2\pi f_0}$, respectively, where both estimations are proportional to the sound speed. We found that, even under strong wind conditions, with wind speeds reaching up to 10 m/s, the distance calculation errors remain below 3%.[2] Does this imply that ambient airflow will only cause issues for acoustic sensing?

In fact, airflow does not degrade acoustic sensing performance by directly causing calculation errors, but rather by disrupting two critical modules in the signal processing pipeline—background subtraction and static vector removal—illustrated by the orange blocks in Fig. 2. As shown in Eq. (6) and Eq. (7), both background subtraction and static vector removal assume that the CFRs from static objects remain constant over time. However, turbulence in the air introduces abrupt and random fluctuations in the sound speed, causing changes even between two consecutive chirps within milliseconds. Mathematically, if we only consider one static object, we have:

$$c[n] \neq c[n-1] \Rightarrow \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n]}\right) - \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n-1]}\right) \neq 0, \tag{9}$$

where $c[n]$ and $c[n-1]$ denote the sound speeds for two consecutive chirp signals. We can observe that, due to the varying sound speed, the CFRs from static objects can no longer be treated as constant over time, causing the assumptions behind background subtraction and static vector removal to no longer hold.

To intuitively illustrate the above-mentioned issue, we show an example of the coarse-grained and fine-grained sensing results with and without ambient airflow in Fig. 3. In this experiment (setup illustrated in Fig. 7), we mount a wooden board (target object) on a motion stage and control it to move forward and backward with a displacement of 2 cm. The speaker and microphone are co-located at a distance of 1.4 m from the target. A few static reflectors (such as a tabletop, book, etc.) are positioned around the target.

As shown in the first row of Fig. 3, when there is no airflow, the subtracted CFR magnitude $|M'[f_d, n]|$ consistently shows a dominant peak at 1.4 m, accurately reflecting the target's distance. If we further extract the CFRs across time at this distance for fine-grained sensing, we can observe that the I-Q variations caused by the target movement consistently rotate around a central point. The static vector pointing to the central point is produced by the static objects around the target as indicated by Eq. (7). By estimating the static vector and removing it, we can compute the phase changes caused by the target movement and further derive accurate fine-grained displacement estimates.

The second row of the figures provides an example of airflow disrupting both coarse-grained and fine-grained sensing. In the figure of the subtracted CFR magnitude $|M'[f_d, n]|$, some peaks (bright yellow pixels) appear at distances where no moving objects are present, indicating a failure in the background subtraction method.

[2]Considering that the sound speed is 343 m/s at room temperature, the calculation error can be computed as $\frac{10\ m/s}{343\ m/s} \approx 3\%$.

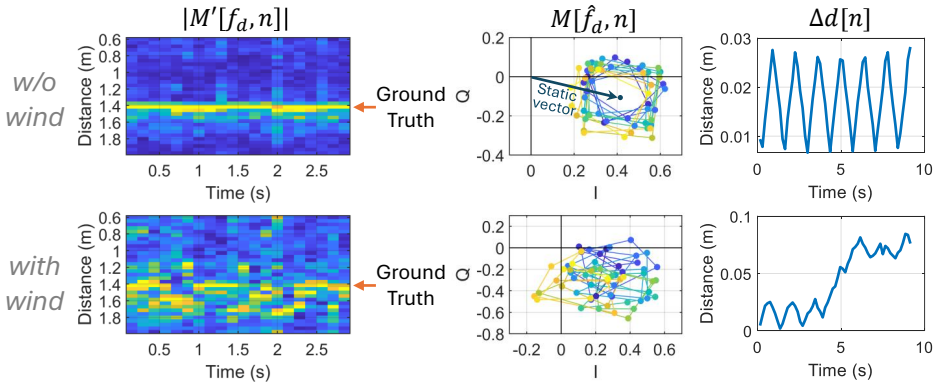


Fig. 3. The illustrative example of fine-grained and coarse-grained sensing with and without airflow (wind). A target moves forward and backward with a displacement of 2 *cm* at 1.4 *m* away from the sensing device. The figures from left to right show the coarse-grained distance estimation, I-Q variations caused by the target movement, and fine-grained distance estimation.

Essentially, even when there is only one static reflector in the range bin, the subtraction of two consecutive CFR vectors would not be a zero vector due to the fluctuations of the speed of sound, as illustrated in Eq. (9). If the magnitude of this residue vector is accidentally high for some chirps due to the abrupt fluctuation of the speed of sound, it can introduce undesired peaks at distances containing only static objects, potentially causing distance estimation errors in coarse-grained sensing. Furthermore, for fine-grained sensing, the static I-Q vector from the static reflector is no longer a constant and starts shifting over time. As a result, the I-Q vector from target movements rotates more randomly without a fixed central point. This further leads to drifts in fine-grained movement estimation as we can no longer isolate the CFR of the moving target through static vector removal.

Note that Fig. 3 shows an example where both fine-grained and coarse-grained sensing are severely affected by airflow. A more comprehensive and quantitative discussion of how airflow impacts acoustic sensing across different environmental conditions is presented in Sec. 6.

5 Understanding Speed of Sound under Airflow

In this section, we present a quantitative analysis of how airflow impacts the speed of sound. We begin by conducting experiments to measure sound speed variations under airflow, followed by the use of the Autoregressive (AR) model to characterize these variations. The resulting model serves as a valuable tool for parameterizing airflow, enabling more precise analysis of its influence and supporting the development of effective mitigation solutions.

Experiment to measure sound speed: As illustrated in Fig. 4, we place a speaker and a microphone (Mic 1) at a fixed distance d apart, and measure the Time-of-Flight (ToF) Δt between the speaker and the microphone using an FMCW signal. The speed of sound can be obtained using $c = \frac{d}{\Delta t}$. Since it is challenging to perfectly synchronize the speaker and the microphone, we connect another microphone (Mic 2), co-located with the speaker and connected to the same sound card [40] as Mic 1. This setup allows us to calculate Time-Difference-of-Arrival (TDoA) between the synchronized Mic 1 and Mic 2. Specifically, we set the distance d to 1 m and use an FMCW signal sweeps from 18 kHz to 22 kHz with a duration of 50 ms. A fan is used to generate ambient airflow. We vary the location, orientation, and speed level of the fan to create different airflow profiles. We conduct the measurements on two types of fans, including a bladeless fan producing more steady wind and a traditional fan with blades. We follow the prior work [6, 31] to estimate TDoA using FMCW signals.

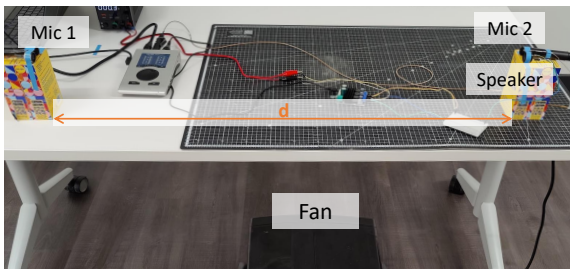


Fig. 4. The experiment setup to measure the sound speed using the TDoA between two synchronized microphones.

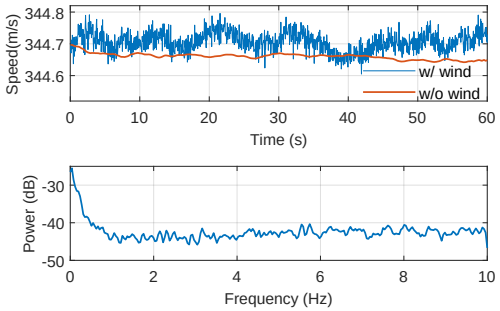


Fig. 5. The results of sound speed measurements in both time and frequency domains.

Fig. 5 illustrates the results of sound speed measurements with and without airflow (wind). We can observe that the measured sounds remain relatively stable when there is no wind. The reason is that sound speed is determined by environmental factors such as temperature, humidity, and air pressure [52]. These factors typically do not change abruptly over short time intervals. However, when wind is introduced by a fan, the measured sound speeds differ from those in the absence of wind in three aspects:

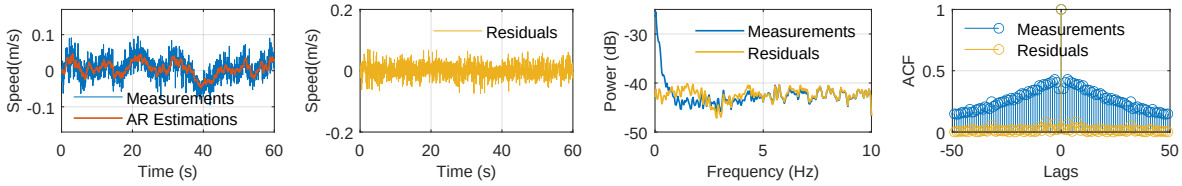
- The mean value of the measured sound speeds shifts slightly higher with the introduction of wind. This is attributed to the fan's orientation being slightly angled toward Mic 1 in our experiment, which results in increased sound speed along the wind direction.
- There exist abrupt and random fluctuations in the sound speed measurements, which are caused by unsteady and irregular turbulence flow. Numerous spikes make the sound speed measurements inconsistent even between consecutive chirps.
- The measured sound speeds exhibit short-term correlations, with identifiable macro-scale peaks and valleys. By analyzing the power spectrum of the measurements—after subtracting the mean—we observe that the energy is concentrated at low frequencies, deviating from the flat spectrum characteristic of white noise. This pattern further confirms the presence of temporal dependencies in the sound speed measurements.

Based on these observations, we propose using the Autoregressive (AR) model [5] to characterize sound speed variations under wind. The AR model is well-suited for capturing time-varying processes with temporal dependencies [2, 24] and has been widely applied across various domains to represent natural and time-dependent phenomena, including temperature fluctuations [14], river flow variations [2], and seismic activity [43]. Specifically, it models each value in a time series as a linear combination of previous values and a white noise term. The AR model $AR(p)$ of order p for sound speed measurements $c[t]$ can be defined as:

$$c[t] = \sum_{i=1}^{p} \phi_i c[t-i] + \varepsilon[t], \quad \varepsilon[t] \sim \mathcal{N}(0, \sigma^2), \tag{10}$$

where $\phi_1, \cdots, \phi_p$ are the linear model parameters, and $\varepsilon[t]$ represents white noise with variance σ^2. Based on the AR model, we can parameterize the behaviors of sound speed variations caused by a certain type of airflow. Specifically, we denote the parameters of an airflow as $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \sigma^2\}$, where $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)$.

Fig. 6 provides an illustrative example of applying the AR model to fit the speed of sound measurements. Note that the mean value is subtracted out beforehand as our analysis only focuses on the fluctuations. The red curve in Fig. 6a presents the estimations calculated by the linear combination term in the AR model, capturing the macro-scale trends in the raw measurements. The residuals, obtained by subtracting the AR estimations from the raw measurements, exhibit white noise characteristics in both the time (Fig. 6b) and frequency domains (a



(a) AR estimations (b) Residuals (c) Power Spectrum (d) Autocorrelation

Fig. 6. An illustrative example of using the AR model to fit the sound speed measurements. (a) The raw measurements and AR estimations. (b) The residuals after removing AR estimations from raw measurements. (c) The spectrum of raw measurements and residuals. (d) The absolute values of the autocorrelation for both raw measurements and residuals. (b)-(d) validates the effectiveness of the AR modeling as the characteristics of the residuals are close to white noise.

	Distance (m)	Wind Speed (m/s)	**Parameters θ**		ACF Error of Residuals (MSE)
			ϕ	σ	
Fan #1	0.5	7.5	(0.11, 0.26, 0.23, 0.16, 0.13)	0.017	5.6×10^{-4}
	1.1	5.4	(0.19, 0.05, 0.06, 0.11, 0.12)	0.033	6.8×10^{-4}
	1.7	3.6	(0.39)	0.034	6.5×10^{-4}
	2.3	2.4	(0.60)	0.030	6.4×10^{-4}
Fan #2	1.7	2.4	(0.76)	0.049	9.5×10^{-4}

Table 1. The results fitted by AR models under various experiment setups. The generated turbulence flows are different even when the wind speeds generated by two fans are the same. Small ACF errors of residuals mean good fitness of AR models, compared with the corresponding errors from raw measurements (around 1×10^{-2}).

steady power spectrum in Fig. 6c). Fig. 6d shows the absolute values of the Autocorrelation Function (ACF) of both the raw measurements and residuals. The ACF values of the raw measurements increase when the lags are close to zero, indicating the short-term dependencies of the raw measurements. In contrast, the ACF values of the residuals are consistently close to zero at non-zero lags, suggesting that the residuals do not have correlated components. This example validates the AR model's ability to effectively capture sound speed variations.

Table. 1 presents the parameters estimated from the raw sound speed measurements where fans are placed at different distances away from the signal path between the speaker and Mic 1. Different airflow profiles, modeled by $\boldsymbol{\theta}$, are observed across setups. Even with identical wind speeds, i.e., Fan #1 and Fan #2 both at 2.4 m/s, the parameters reflect distinct turbulence characteristics. We also evaluate the AR model's fit quality by calculating the Mean Square Error (MSE) of the normalized Autocorrelation Function (ACF) of the residuals, comparing it to the ACF of an ideal white noise. All of he computed errors are around $5 \times 10^{-4} \sim 9 \times 10^{-4}$, which is two orders of magnitude smaller than that from the raw measurements (over 1×10^{-2}), indicating a strong fit of the AR model to the sound speed measurements.

Note that Table. 1 only shows a limited subset of our airflow characterization. We conduct comprehensive data collection across various airflow environments, including different fan speeds, diverse surroundings, and varying lengths of chirp signals. We characterize and parameterize each airflow profile as a new $\boldsymbol{\theta}$, which enables a comprehensive analysis of how different factors of wind affect acoustic sensing through parameterized simulation. We present the details and our findings in Sec. 6.

To complement the empirical results in Fig. 6 and Table. 1, we further validate the physical fidelity of our AR model through physics-based modeling of complex turbulent flow, as presented in Appendix A.

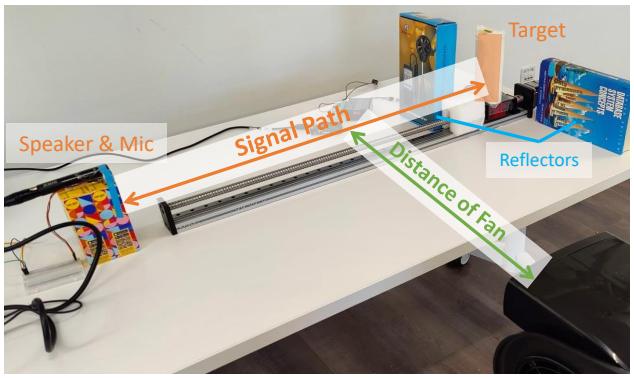


Fig. 7. The experiment setup to evaluate the performance of acoustic sensing under airflow.

6 Analyzing Real-world Factors Affecting Acoustic Sensing Under Airflow

In this section, we present a comprehensive analysis of how various real-world factors interact with airflow to influence acoustic sensing performance, based on both real-world experiments and extensive simulations. Our analysis covers various airflow characteristics, environment conditions, sensing tasks, and system configurations. By default, the chirp signal sweeps from 18 kHz to 22 kHz with a duration of 50 ms at a sampling rate of 48 kHz.

Real-world Experiments: As shown in Fig. 7, the speaker and microphone are connected to the same sound card to transmit and receive chirp signals. We mount a wooden board on a motion stage and control its movement to emulate the displacement of a target object. In addition, we vary the distance between the fan and the signal path to create different airflow profiles. Several static reflectors (e.g., a cardboard box and a book) are positioned around the target to simulate indoor multipath reflections. Unless otherwise specified, the bladeless fan is utilized to create an airflow with a speed of 4.1 m/s, and the target is located 1 m away from the speaker and microphone.

Simulations: To provide a detailed quantitative analysis of how various factors—such as airflow characteristics and object-reflected signal strengths—impact acoustic sensing, we conduct extensive simulations grounded in our problem formulation and sound speed modeling. Specifically, we simulate the CFR at each range bin based on Eq. (5) and the sound speed variation based on Eq. (10), where we parameterize the target movement, the location and reflective attenuation of each reflector, and the airflow profile defined by the AR model ($\boldsymbol{\theta} = \{\boldsymbol{\phi}, \sigma^2\}$). After generating the simulated CFRs, we apply the same processing pipeline as in our experiments for both coarse-grained and fine-grained sensing.

We present our key observations and findings below.

Observation 1: Turbulence, rather than wind speed alone, significantly affects acoustic sensing. In this experiment, we vary the fan's distance from the signal path to produce airflow at different speeds. At each distance, we use an anemometer positioned on the signal path facing the fan to measure wind speeds. The target's distance is estimated using the coarse-grained sensing pipeline. Fig. 8 presents the ranging errors when the fan is at each distance. We can observe that, as the fan moves further away, wind speed naturally decreases due to attenuation. Surprisingly, the ranging error initially increases as wind speed decreases (i.e., fan distance from 0.5 m to 1.5 m). Ranging error begins to decrease only when the distance exceeds 1.5 m. It suggests that higher wind speeds do not necessarily correlate with greater error in acoustic sensing.

To further investigate this observation, we conduct three sets of experiments:

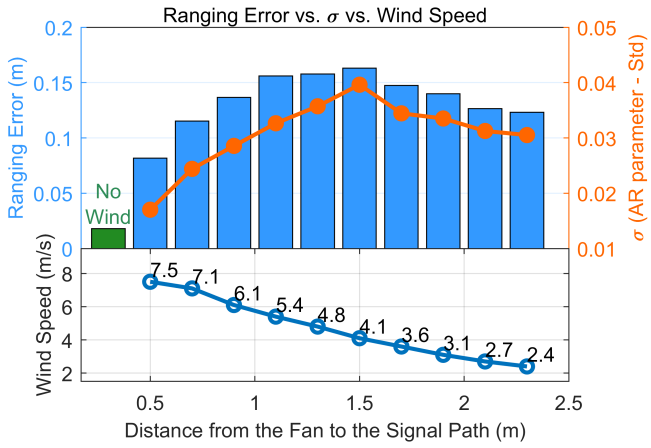


Fig. 8. The ranging errors when the fan is placed at different distances from the signal path. A higher speed of wind is created when the fan is closer. However, increased wind speed does not necessarily result in higher ranging errors. Instead, the error correlates strongly with the standard deviation parameter σ in the AR model, which characterizes turbulence randomness.

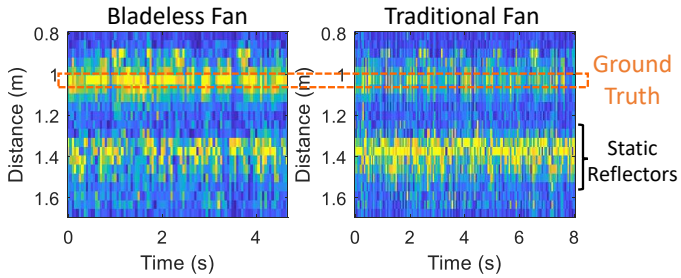


Fig. 9. An illustrative example comparing the ranging performance of two fans operating at the same wind speed. The traditional fan with blades induces more turbulence compared to the bladeless fan.

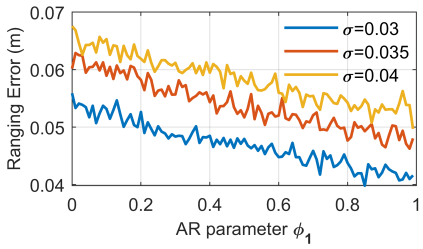
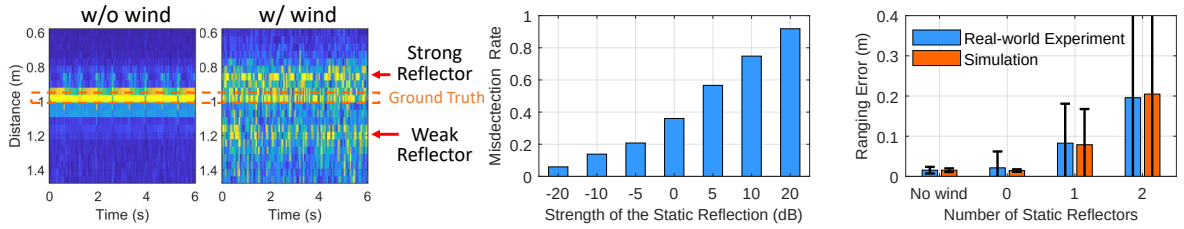


Fig. 10. The ranging errors across different airflow profiles. Lower errors are observed under lower σ and higher ϕ_1 (higher stability).

- We measure sound speed variations for each fan placement using the method outlined in Sec. 5. We fit sound speed measurements using the AR model and compute the standard deviation parameter σ of the white noise term. As visualized in the orange curve in Fig. 8, it shows a strong correlation with the ranging error. The parameter σ effectively captures the randomness in sound speed variation, with higher σ values indicating increased turbulence irregularity. As the fan moves farther from the signal path, wind speed decreases monotonically, but turbulence irregularity along the signal path first increases and then decreases beyond a certain distance.
- We test two types of fans separately: a bladeless fan and a traditional fan with blades, while keeping all other experimental settings the same. To ensure the same wind speed along the signal path (i.e., 2.4 m/s), we carefully adjust their placement distances with respect to the signal path. As shown in Fig. 9, the dominant peaks (bright yellow pixels) for the bladeless fan align well with the ground truth, while those for the traditional fan deviate towards the locations of the static reflectors. The airflow characteristics for both fans at this wind speed are detailed in Table 1, where the traditional fan (Fan #2) shows a significantly



(a) An illustrative example (b) Impact of static reflection strength (c) Impact of number of reflectors

Fig. 11. The coarse-grained sensing results across different setups for static reflectors. (a) An illustrative example demonstrating the impact of static reflector under airflow. (b) The target misdetection rate increases with the strength of the static reflector. (c) The ranging error rises as the number of static reflectors increases.

higher standard deviation $\sigma = 0.049$ compared to that of the bladeless fan (Fan #1) $\sigma = 0.030$. It indicates that the bladeless fan produces a steadier and more consistent wind with less turbulence.

- We conduct simulations to create different airflow profiles by adjusting the standard deviation parameter σ and the first linear coefficient parameter ϕ_1 in the AR model. For simplicity, we only consider the AR model of order 1 for sound speed measurements, i.e., $AR(1)$. We set σ to 0.03, 0.035, and 0.04, respectively. For a given σ, we vary ϕ_1 from 0 to 1. As shown in Fig. 10, the ranging error decreases as ϕ_1 increases and σ decreases. When ϕ_1 approaches 1, the current sound speed becomes more dependent on the previous sound speed, leading to less pronounced fluctuations.

These experiments indicate that higher σ and lower ϕ_1 lead to more irregular turbulence, which, in turn, generates a more random and less predictable sound speed. These fluctuations in sound speed hinder background subtraction from effectively removing the CFR of static reflectors, resulting in high sensing errors. In summary, turbulence—not merely wind speed—is the primary factor impacting acoustic sensing performance.

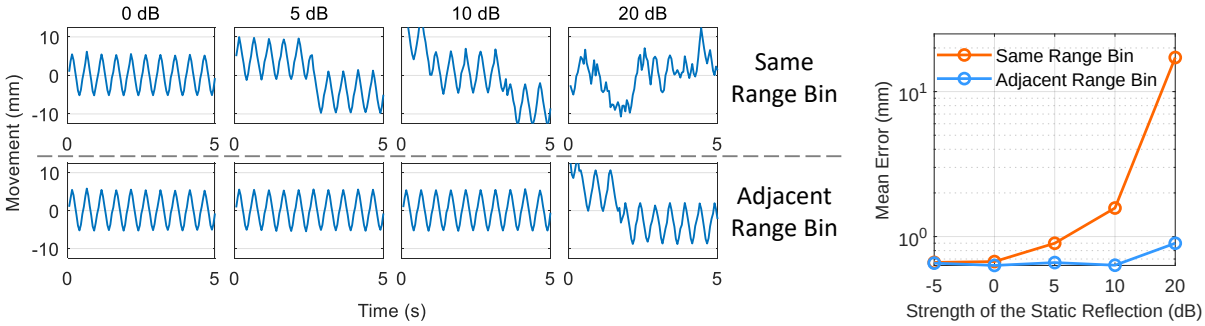
Observation 2: The impact of airflow is significantly amplified in environments with stronger and larger numbers of static reflectors. Based on the problem formulation in Eq. (9), we can denote the residual CFR of a static reflector i after background subtraction as:

$$R_i = \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n]}\right) - \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n-1]}\right) \neq 0, \tag{11}$$

where α_i is the attenuation factor, f_0 is the initial frequency of the chirp signal, and d_i is the reflection path length. $c[n]$ and $c[n-1]$ denote the sound speeds for two consecutive chirp signals. We can observe that, for a given static reflector at distance d_i, there are two factors affecting the residual R_i—sound speed variations $c[n]$ and reflection attenuation α_i. We discussed the factor of sound speed fluctuations in Observation 1. Now, we conduct experiments to investigate the impact of varying reflection attenuation on acoustic sensing.

A higher α_i value signifies a stronger static reflector, which may result in a larger residual CFR after background subtraction under airflow conditions. Fig. 11a visualizes an illustrative example of coarse-grained ranging results where a strong and a weak static reflector are placed around the target at 1 m, respectively. When the fan is off (no wind), signals from both reflectors are perfectly subtracted, allowing the target location to be clearly identified. However, when the fan is turned on (with wind), strong residual signals from both reflectors remain after background subtraction, causing significant ranging errors. Moreover, the stronger reflector generates considerably more interference than the weaker one.

To further investigate the effect of static reflector strength on sensing performance, we conduct a quantitative analysis using simulations with a target positioned at 1 m and a static reflector at 1.2 m from the device. We vary



(a) Illustrative examples of fine-grained movement tracking

(b) Fine-grained tracking errors

Fig. 12. The fine-grained tracking results where the static reflector is positioned in the same range bin and the adjacent range bin with respect to the target, respectively. Its relative signal strength to the target reflection is varied from 0 dB t0 20 dB. When a strong reflector is located in the same range bin as the target, fine-grained sensing would be severely interfered.

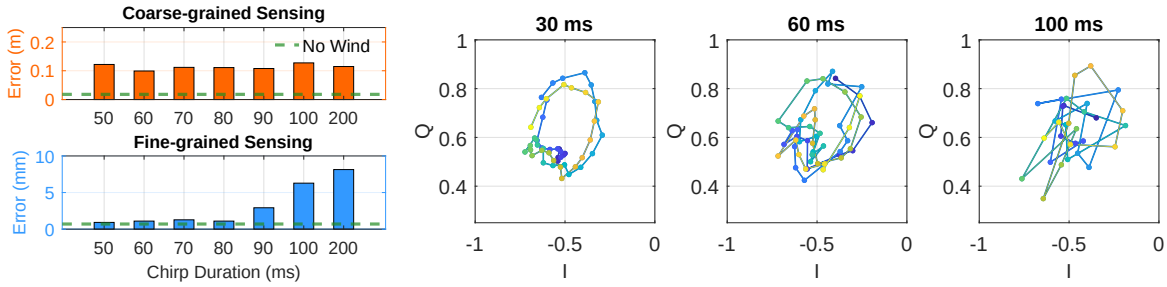
the signal strengths of reflection from the static reflector, measuring them on a decibel scale relative to the signal strength of the target. We simulate the airflow using the AR model parameters that produce the highest turbulence observed in Observation 1. Fig. 11b demonstrates the target misdetection rate, which is defined as the ratio of the number of misdetections to the total number of measurements. We can observe that, the misdetection rate is negligible when the static reflector's signal strength is below -10 dB. However, when the signal strength reaches 5 dB, the number of target misdetections exceeds that of correct detections. At even higher signal strengths, coarse-grained distance estimation nearly fails, as the target range bin becomes almost unidentifiable.

We also evaluate the effects of varying the number of static reflectors from 0 to 2, with reflectors randomly positioned between 0.5 m and 1.5 m from the device in the presence of airflow. Additionally, we conduct a set of reference experiments with two nearby static reflectors in the absence of airflow. Fig. 11c shows the ranging errors for both real-world experiments and simulations. When no static reflector is present under airflow conditions, the ranging error remains as low as in the no-airflow condition with two reflectors. The mean ranging error increases to approximately 8 cm with one reflector and 20 cm with two reflectors. It is worth noting that the results obtained from real-world experiments and simulation are consistent, further validating the fidelity of our simulation method.

Observation 3: Fine-grained sensing is less susceptible to the effects of airflow than coarse-grained sensing. Intuitively, one might assume fine-grained sensing is less robust to noise than coarse-grained sensing, given that target movements in fine-grained tasks are generally smaller. However, our results indicate that, once the correct range bin of the movement is identified, fine-grained movement tracking is actually less affected by airflow.

To investigate this unexpected outcome, we conduct simulations where the target moves back and forth with a displacement of 1 cm at 1 m away from the device. We also simulate a nearby static reflector and vary its signal strength relative to the target reflection from 0 dB to 20 dB. The airflow is simulated using the AR model parameters that produce the highest turbulence in Observation 1. Fig. 12a demonstrates illustrative examples of fine-grained movement estimation where the static reflector is positioned in the same range bin and adjacent range bin with respect to the target[3], respectively. We can observe that, when the static reflector is in the same range bin as the target, the fine-grained movement pattern is highly susceptible to distortion caused by the airflow. Even with a relative strength of only 5 dB between the static reflector and the target, the movement pattern is

[3]The length of one range bin is $\frac{c}{2B} = \frac{340}{2*4000} = 4.29$ cm in our chirp setting.



(a) Tracking errors across chirp durations

(b) I-Q plot of CFR under different chirp durations

Fig. 13. The results under different chirp durations. (a) shows both coarse-grained and fine-grained tracking accuracies, with green lines representing performance in the absence of wind. The coarse-grained tracking remains consistently erroneous, while the fine-grained tracking performance degrades with longer chirp durations, (b) shows the I-Q plot of CFR under different chirp durations. The IQ signals vary more randomly as chirp duration increases.

distorted. Notably, while a slight distortion may not impact periodic human activities (e.g., respiration), it can significantly degrade the sensing performance for non-periodic activities, such as finger tracking or eye blink detection. As the reflection strength of the static reflector increases, the distortion caused by airflow becomes more pronounced. We find that the larger the distance between the static reflector and the target, the stronger the reflection from the static reflector must be to cause distortion. For instance, when the static reflector is located in the adjacent range bin rather than the same range bin as the target, its strength relative to the target must be at least 20 dB to distort the movement pattern. We can obtain similar observations from the mean tracking errors of fine-grained movement in Fig. 12b.

Observation 4: FMCW signals with longer chirps are more prone to airflow interference in fine-grained sensing applications. Prior acoustic sensing systems [21, 33] have proposed to leverage chirp signals with longer duration to achieve extended sensing ranges. We explore the effect of chirp length, an important system parameter that may impact FMCW-based acoustic sensing performance under airflow.

We conduct simulations with a target positioned at 1 m from the device and two static reflectors located at 10 cm and 50 cm away from the target. The target is simulated to move back and forth with a displacement of 1 cm. The chirp durations are varied from 30 ms to 200 ms. The airflow is simulated using the AR model parameters that produce the highest turbulence in Observation 1. Fig. 13a shows the tracking errors for both coarse-grained and fine-grained sensing. The coarse-grained tracking error is consistently at around 11 cm across different chirp durations. However, for fine-grained sensing, we observe significant performance degradation when the chirp duration is longer than 90 ms.

To understand this phenomenon, we visualize the I-Q plot of CFR at the target range bin in Fig. 13b. Essentially, each point in the I-Q plot is a combination of a static vector (from static reflectors) pointing to the center of the circle, plus a dynamic vector induced by the moving target. The movement results for the fine-grained sensing are derived by first estimating the center of the circle in the I-Q plot, and then calculating the phase shifts. As shown in Fig. 13b, the shapes of all the curves deviate from an ideal circle, with their centers drifting over time due to airflow-induced shifts in the static vector. However, when the chirp duration is shorter (e.g., 30 ms), we have a larger number of chirps per second, resulting in denser points on the circles. In this case, we can still quite accurately estimate the center of each circle even when it is drifting. On the other hand, when we use a longer chirp duration (e.g. 100 ms), the number of chirp samples on each circle are sparse and the shape of the circle can

be barely identified in the presence of airflow. As a result, it is challenging to estimate the center of the circle, and a more significant error in fine-grained sensing is induced.

7 Practical Recommendations and Technical Solutions

In this section, we first discuss practical recommendations when deploying existing acoustic sensing systems in real-world environments with ambient airflow, based on our key observations presented in the previous section. Following that, to further push acoustic sensing for wider adoption, we propose signal processing solutions as plug-and-play modules that can be integrated into existing coarse-grained sensing and fine-grained sensing pipelines to enhance system performance under ambient airflow.

Practical Recommendations for Acoustic Sensing Deployment:

- **Choose appropriate air circulation systems:** Opt for bladeless fans or air conditioners that deliver steady airflow with minimal turbulence. To reduce turbulence-induced disruptions, minimize obstacles along the airflow path, as irregular air movement can significantly affect sensing accuracy.
- **Be mindful of environmental reflectors:** Strong reflectors in the environment amplify the impact of airflow interference. For example, deploying an acoustic gesture recognition or hand-tracking system beside a wall can be susceptible to the effects of airflow.
- **Optimize distance configuration:** For fine-grained sensing applications such as respiration or heartbeat monitoring, pre-configuring the approximate distance between the sensor and the target can enhance robustness. Since fine-grained tracking is more resilient to airflow when the correct range bin is identified, supplying the system with user-to-device distance information significantly improves its robustness under airflow interference.
- **Adjust signal parameters:** Employ shorter chirp durations in FMCW signals to enhance system robustness, especially for fine-grained sensing applications. This configuration choice helps maintain sensing accuracy under airflow interference.

In addition to practical recommendations for minimizing the impact of airflow on acoustic sensing, a more robust signal processing pipeline is always desired to handle inevitable airflow interference in real-world deployments. To this end, we propose solutions that can be integrated into existing sensing pipelines to enhance system robustness. The following subsections detail our technical approaches for coarse-grained and fine-grained sensing.

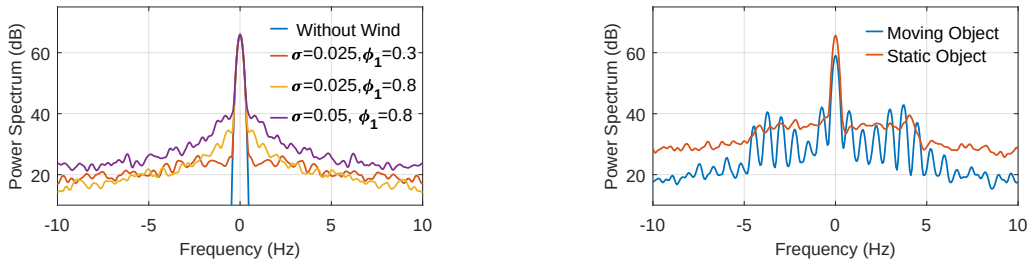
7.1 Mitigating Coarse-grained Sensing Error

Coarse-grained sensing error is induced when the CFR from static reflectors are dynamically changing due to the speed of sound fluctuations. In such cases, existing background subtraction algorithms cannot distinguish whether the variations in the Channel Frequency Response (CFR) over time are caused by object movement or fluctuations in the speed of sound. A range bin containing only static reflectors may be mistakenly identified as a target by the algorithm. To mitigate coarse-grained sensing errors under airflow, we design an algorithm that supports background subtraction by distinguishing CFR variations caused by object movement from those caused by airflow.

Mathematically, the problem can be formulated as a signal classification problem between:

$$M_t[n] = \exp\left(j2\pi \frac{f_0 d_t[n]}{c[n]}\right) \quad \text{and} \quad M_s[n] = \sum_{i=1}^{S} \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n]}\right),$$

where $M_t[n]$ represents the CFR from the range bin containing a moving target and $M_s[n]$ denotes the summation of CFRs from the range bin with S static reflectors. Our objective is to identify proper features out of a CFR sequence that can distinguish the two type of CFRs. Note that the mixed CFR, which includes contributions from



(a) Spectrum of $M_s[n]$ under different AR parameters

(b) Spectrum of $M_s[n]$ and $M_t[n]$

Fig. 14. The visualization of the spectral patterns of CFR from static reflectors ($M_s[n]$), and from moving objects ($M_t[n]$). $M_s[n]$ in (a) has a stable and structured spectral pattern under different AR model parameters, while $M_t[n]$ in (b) has more variations across frequencies, providing us with an opportunity to distinguish them.

both static reflectors and moving objects, exhibits characteristics similar to $M_t[n]$. Thus, we simplify our notation without loss of generality.

As we discussed in Sec. 5, the speed of sound variation $c[n]$ can be effectively characterized using an AR model. Rather than being purely random, the speed of sound sequence exhibits temporal correlations and the magnitude of the variation is bounded within a range, as illustrated in Fig. 5. Also, its spectral pattern is also relatively stable across a wide frequency range. Following the observation, we investigate the spectral pattern of $M_s[n]$. In Fig. 14a, we visualize the spectrum of $M_s[n]$ corresponding to different speed of sound sequences generated by different parameters in the AR model. Similar to the spectrum of $c[n]$, we also observe a consistent and structured spectral pattern—featuring a dominant peak around 0 Hz induced by the constant energy of the reflector, along with a broad and stable distribution across higher frequencies caused by sound speed variations. We further visualize a spectrum of the CFR sequence from a moving object $M_t[n]$ in Fig. 14b, where a series of peaks and valleys can be observed, due to the instantaneous velocity of the moving object. A key observation from the results is that the CFR from static reflectors exhibits a stable and structured spectral pattern, whereas the CFR from moving objects shows greater spectral variation across frequencies.

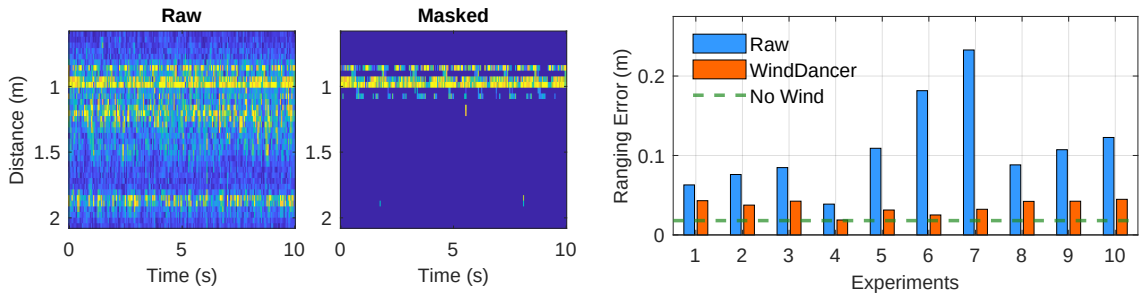
By leveraging this important observation, we develop an algorithm using **spectral variation thresholding** to identify and mask out the CFR from only static reflectors. We first calculate the magnitude of the spectrum of a CFR sequence $M[n]$ across a certain time window:

$$S_M[f] = |\mathcal{F}[M[n] - \mu_M]|, \tag{12}$$

where μ_M is the mean value of the sequence in the time window and $\mathcal{F}$ represents Fast Fourier Transform (FFT). We remove the mean value prior to spectrum calculation to eliminate the effect of constant reflective energy, which manifests as the dominant peak at 0 Hz in Fig. 14. We then calculate the normalized spectral variation V_M using $S_M[f]$:

$$V_M = \frac{\sigma^2_{S_M[f]}}{\mu_{S_M[f]}}, \tag{13}$$

where $\sigma^2_{S_M[f]}$ and $\mu_{S_M[f]}$ are the variation and mean of $S_M[f]$ across frequencies respectively. The spectral variation V_M characterizes the intensity of fluctuations on the spectrum of $S_M[f]$ across frequencies. Based on the observation we made above, V_M from the CFR of static reflectors would be lower and that from a moving object is higher. Thus, to reduce the error rate of mistakenly identifying static reflectors as the target under airflow, we propose to mask out the CFR if the V_M is lower than a threshold, before the final range bin estimation.



(a) $|M'[f_d, n]|$ before and after spectral variation thresholding (b) Ranging error across different experiments

Fig. 15. Results for mitigating coarse-grained sensing error under airflow. (a) Our algorithm can effectively mask the CFR component from static reflectors (the target is located at 1 m). (b) By leveraging the proposed algorithm, our system outperforms the traditional ranging algorithm by a large margin under ambient airflow.

In our experiments, we empirically set the threshold as 0.1 and we use a 0.5 s time window to segment the signal and calculate the spectral variation.

Fig. 15a shows an example of $|M'[f_d, n]|$ (the CFR magnitude after background subtraction across time and range bins), before and after applying our proposed masking algorithm. The target object is located at 1 m and a few static reflectors are located in the range of 0.8 m to 2 m. As shown in the masked example, our algorithm can effectively filter out most of the signal from static reflectors (e.g. at around 1.2 m and 1.9 m), while retaining a dominant signal component from the moving object.

We further evaluate the algorithm in ten different environmental settings with different reflectors and wind profiles, and the result is shown in Fig. 15b. On average, the original processing pipeline ("Raw" in the figure) has a ranging error of 11.0 cm, while our proposed spectral variation thresholding algorithm effectively reduces the mean error to 3.6 cm, demonstrating a performance improvement of 67.3 %. While the error is still slightly higher than the error without wind (1.8 cm), we believe our method provides a simple yet effective plug-and-play solution to enhance the robustness of the existing acoustic sensing pipeline under airflow.

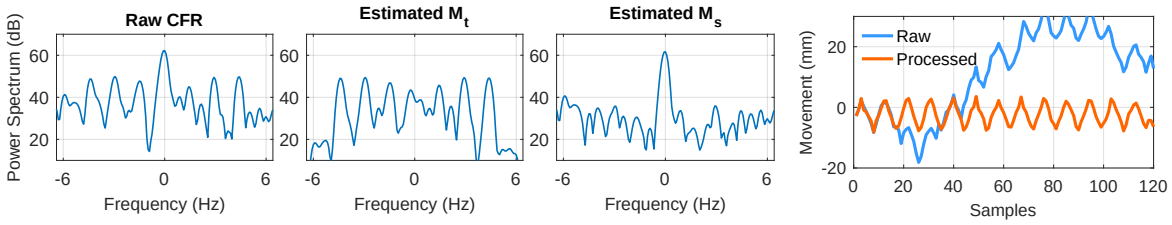
7.2 Mitigating Fine-grained Sensing Error

As we presented in Sec. 6, fine-grained sensing error would be induced when strong reflectors exist at the approximately same distance of the target away from the device. Unlike the challenge of distinguishing the correct range bin for coarse-grained sensing, this section addresses the signal separation issue when both the moving target and static reflectors are located within the same range bin. Mathematically, we consider the mixed CFR sequence:

$$M_x[n] = M_t[n] + M_s[n] = \exp\left(j2\pi \frac{f_0 d_t[n]}{c[n]}\right) + \sum_{i=1}^{S} \alpha_i \exp\left(j2\pi \frac{f_0 d_i}{c[n]}\right), \tag{14}$$

where $M_t[n]$ and $M_s[n]$ are the CFR sequence from the target object and static reflectors respectively. The objective is to develop a signal separation algorithm that can effectively separate $M_t[n]$ and estimate target movement $d_t[n]$.

However, as shown in Fig. 14b, separating the two signals is non-trivial. Both signals span a wide frequency band, and their spectra overlap, making direct filtering-based separation infeasible. To tackle this challenge, we employ Variational Mode Decomposition (VMD) [12] to first decompose the mixed signal and then reconstruct components to enable signal separation. VMD is a data-adaptive technique that decomposes the signal into



(a) Example of VMD-based separation result shown in spectrum (b) Example of the processed waveform

Fig. 16. An example showing our signal separation method based on VMD. In (a), the raw CFR consists of both components from the moving target and static reflectors. The algorithm reconstructs the first $N-1$ IMFs to estimate the target's component, and combines the last IMF and residual to estimate the static reflector's component. (b) shows the corresponding phase waveform before and after processing.

a discrete number of bandwidth-concentrated components, known as Intrinsic Mode Functions (IMFs), while preserving their physical characteristics. Prior systems leverage VMD algorithms for heartbeats and breathing monitoring [42, 58] under the interference of other signal components. We choose VMD based on two key observations from the spectral characteristics in our data:

- The static reflector's CFR sequence $M_s[n]$ exhibits a single dominant spectral peak at 0 Hz, where most of the energy is concentrated (Fig. 14a).
- The target's movement manifests as a series of distinct spectral peaks in the CFR sequence ($M_t[n]$), each corresponding to instantaneous velocities during the movement (Fig. 14b).

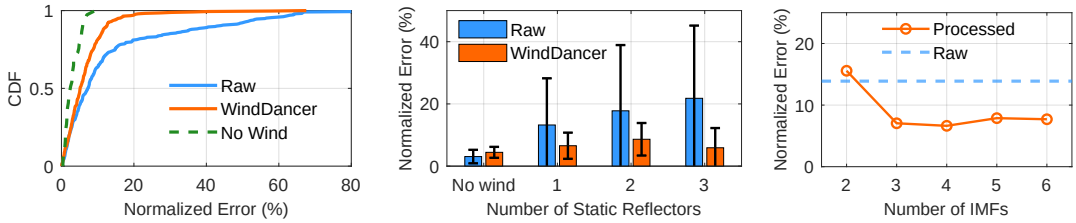
By applying VMD, we can effectively decompose the signal into a fixed number of IMFs, where each IMF captures a dominant spectral peak and its physically related frequency components.

Since VMD is designed for real-valued signals while our CFR sequences are complex-valued, we process the real and imaginary components separately before recombining them into the final complex-valued result. Besides, we empirically find that the last IMF decomposed by VMD is always a component with a dominant spectral peak at 0 Hz, associated with the CFR sequence from the static reflectors. Thus, we retain the first $N-1$ IMFs and add them up to perform signal reconstruction for $M_t[n]$, where N is the total number of IMFs. We then combine the last IMF and the residual to estimate the CFR from static reflectors. Mathematically, we denote our approach as

$$M_t[n] \approx \sum_{i=1}^{N-1} IMF_i(M_x[n]), \quad M_s[n] \approx IMF_N(M_x[n]) + R[n], \tag{15}$$

where $R[n]$ is the residual signal decomposed by VMD. Fig. 16a shows an example in the frequency domain of the signal separation. The raw CFR consists of both components from the moving target and static reflectors. After the separation algorithm, we observe the estimated $M_t[n]$ captures a series of spectral peaks originating from target movement (similar pattern shown in Fig. 14b), and the estimated $M_s[n]$ manifests similar spectral pattern as the CFR from static reflectors shown in Fig. 14a. We further visualize the corresponding extracted phase waveform of the raw CFR and the estimated $M_t[n]$ in Fig. 16b, where the significant drift in the waveform is effectively removed and we obtain a more accurate waveform reflecting the actual movement of the target.

We further evaluate the performance of our signal separation algorithm through a more comprehensive experiment. In the experiment, the target moves back and forth with a 1 *cm* displacement. Multiple static reflectors are placed in the range between 0.8 *m* to 1.2 *m* to create different settings. We use a 150 *ms* chirp duration instead of 50 *ms* used in the previous experiments such that a more significant error can be observed (discussed in Sec. 6, Observation 4).



(a) CDF error plot from all experiments (b) Different environmental settings (c) Different number of IMFs

Fig. 17. Evaluation of fine-grained sensing. (a) shows our methods reduce the mean normalized error from 13.9% to 6.6%. (b) and (c) shows under different environments or different IMF numbers, the errors after processing are consistently under 8.5%.

Fig. 17a shows the overall performance in the CDF plot of the normalized error. The normalized error is defined as the tracking error divided by the actual displacement of the target. We observe that our system significantly improves the fine-grained sensing performance, especially in challenging situations with higher errors. The error on average is reduced from 13.9% to 6.6%. Similar to the coarse-grained sensing results, the error after applying our algorithm is slightly higher compared to the performance under wind-free conditions.

Fig. 17b further shows the performance under different environmental settings. When there is no wind, our algorithm introduces a slightly higher error to the system because of the artifacts in signal decomposition. When wind exists, we evaluate the system with different numbers of reflectors in the environment. With more reflectors, the tracking error of the original pipeline gradually increases. Our algorithm can consistently reduce the error down to <8.5% under different settings.

We also evaluate the robustness of the algorithm when using different VMD configurations. We evaluate the performance when decomposing signals into different number of IMFs. When the number of IMFs is two (i.e. only use the first IMF as the target CFR), the error is even higher than the raw CFR, indicating an insufficient decomposition in the VMD algorithm. When the number of IMFs is higher than 3, we obtain a consistently satisfactory performance for error mitigation. Such a result indicates that our system performance is not sensitive to the number of IMFs, which is an important configuration in the VMD algorithm. In our evaluation, four IMFs are used as the default configuration.

8 Discussion

Computational Cost of the Solutions: We note that our proposed approaches to mitigate coarse-grained and fine-grained sensing errors are computationally lightweight and would not compromise the real-time performance of existing acoustic sensing systems. Both algorithms—based on FFT for coarse-grained sensing and VMD for fine-grained sensing—have a computational complexity of $O(n \log n)$ where n is the sequence length. Importantly, both algorithms operate on chirp sequences rather than raw audio samples, introducing negligible computational overhead compared to the more time-consuming steps in the FMCW processing pipeline, i.e., sample-level signal mixing and range-FFT. Our tests of the MATLAB-implemented algorithms on a single-thread Intel Core i5 CPU showed that the coarse-grained algorithm processed a 10-chirp window (0.5 s) in under 0.03 s, while the fine-grained algorithm handled a 100-chirp window in under 0.01 s. These results demonstrate that our solutions are suitable to be deployed as plug-and-play modules on resource-constrained edge devices.

Other Types of Signals: While this work represents the first comprehensive study of ambient airflow's impact on acoustic sensing systems, we focused our analysis on FMCW-based acoustic sensing due to its widespread adoption. Our fundamental findings regarding the speed of sound fluctuations and preliminary experiments suggest that similar airflow effects exist in other signal types, including continuous wave (CW) signals [41] and

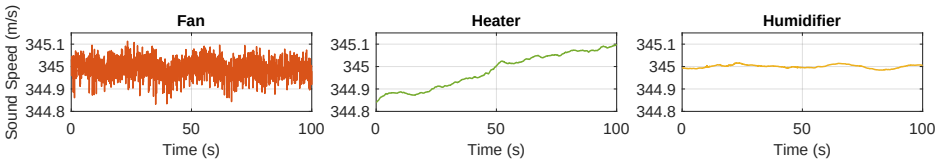


Fig. 18. Sound speed variation when using different environmental comfort devices. Only airflow from the fan induces abrupt and random fluctuations, while the heater and humidifier introduce gradual and predictable patterns.

Frequency Hopping Spectrum Spread (FHSS) signals [48]. Importantly, the sound speed variation modeling based on AR (Sec. 5) provides a generalizable framework for analyzing airflow effects across different signal types, laying thefoundation for future studies of alternative acoustic sensing approaches.

Outdoor Settings: Our airflow profile characterization focused on indoor environments where wind sources are primarily fans and air conditioners. This scope aligns with the predominant use cases of existing acoustic sensing systems, which are largely designed for indoor applications [7, 33, 53, 57]. Furthermore, as we discussed in Sec. 6, strong static reflectors are necessary factors for airflow to impact acoustic sensing. Such conditions are far less common in open outdoor spaces, making indoor environments the more critical context for understanding and mitigating the effects of airflow.

Potential Failure Modes: Our proposed solutions may face limitations under certain challenging conditions. The proposed algorithms process CFR sequences using fixed-length time windows (0.5 s in our implementation), assuming relatively stable wind profiles within each window. In environments with highly dynamic airflow patterns, such as outdoor windy conditions, this assumption may not hold, potentially resulting in degraded performance. Additionally, the direct impact of airflow on microphone membranes can cause signal clipping, resulting in temporary loss of information during these periods. As these represent edge cases outside typical indoor deployment scenarios, we leave a thorough investigation of these scenarios to future work.

Other Environmental Factors Affecting Sound Speed: Indeed, airflow is not the only factor that causes the changes in the speed of sound. Other environmental properties including temperature and relative humidity also affect the sound speed value [4]. Generally, an acoustic wave travels faster in warmer and more humid air. However, as we discussed, airflow affects acoustic sensing not by causing acoustic signals to travel faster or slower, but by introducing abrupt and random fluctuations in the sound speed. In contrast, temperature and humidity changes in indoor environments, especially those with HVAC systems, typically occur gradually and follow predictable patterns [16]. We validate this by placing a heater and humidifier close to our experiment setup (Fig. 4) and measuring the sound variance speed across 100 s. As shown in Fig. 18, the sound speed gradually increases near the heater and remains relatively stable near the humidifier. This contrasts with airflow's abrupt and random effects on sound speed, which pose more significant challenges to acoustic sensing.

9 Conclusion

This paper explores the previously unexamined challenge of ambient airflow interference in acoustic sensing systems. Our findings reveal that airflow can induce significant disturbances in speed of sound, affecting both coarse-grained and fine-grained sensing. Through analyzing the underlying mechanisms, we identify turbulence, rather than wind speed alone, as a primary factor contributing to performance degradation. By modeling sound speed variations with an Autoregressive model, we analyze the airflow's effects on sensing accuracy along with other real-world factors. Additionally, we provide practical recommendations and signal processing enhancements to improve resilience against airflow-related disruptions. This work lays the foundation for advancing acoustic sensing technologies in real-world conditions, enabling broader applicability across diverse environments.

Acknowledgment

We are grateful to anonymous reviewers for their constructive comments. We acknowledge support from the NSF (2106921, 2030154, 2007786, 1942902, 2111751), ONR, and CyLab-Enterprise. This research is also supported by NTU SUG-NAP.

References

[1] [n. d.]. Amazon.com: All Departments. https://www.amazon.com/b?ie=UTF8&node=23435461011. ([n. d.]). Accessed: 2024-09-19.
[2] Nicole H Augustin, Lindsay Beevers, and William T Sloan. 2008. Predicting river flows for future climates using an autoregressive multinomial logit model. *Water Resources Research* 44, 7 (2008).
[3] Manuela Barth and Armin Raabe. 2011. Acoustic tomographic imaging of temperature and flow fields in air. *Measurement Science and Technology* 22, 3 (2011), 035102.
[4] Dennis A Bohn. 1987. Environmental effects on the speed of sound. In *Audio Engineering Society Convention 83*. Audio Engineering Society.
[5] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
[6] Chao Cai, Zhe Chen, Henglin Pu, Liyuan Ye, Menglan Hu, and Jun Luo. 2020. AcuTe: acoustic thermometer empowered by a single smartphone. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (SenSys '20)*. Association for Computing Machinery, New York, NY, USA, 28–41. https://doi.org/10.1145/3384419.3430714
[7] Gaoshuai Cao, Kuang Yuan, Jie Xiong, Panlong Yang, Yubo Yan, Hao Zhou, and Xiang-Yang Li. 2020. Earphonetrack: involving earphones into the ecosystem of acoustic motion tracking. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 95–108.
[8] Shirui Cao, Dong Li, Sunghoon Ivan Lee, and Jie Xiong. 2023. PowerPhone: Unleashing the Acoustic Sensing Capability of Smartphones. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–16.
[9] Tao Chen, Yongjie Yang, Xiaoran Fan, Xiuzhen Guo, Jie Xiong, and Longfei Shangguan. 2024. Exploring the Feasibility of Remote Cardiac Auscultation Using Earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 357–372.
[10] Xiangru Chen, Dong Li, Yiran Chen, and Jie Xiong. 2022. Boosting the sensing granularity of acoustic signals by exploiting hardware non-linearity. In *Proceedings of the 21st ACM Workshop on Hot Topics in Networks*. 53–59.
[11] Haiming Cheng, Wei Lou, Yanni Yang, Yi-pu Chen, and Xinyu Zhang. 2023. TwinkleTwinkle: interacting with your smart devices by eye blink. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–30.
[12] Konstantin Dragomiretskiy and Dominique Zosso. 2014. Variational Mode Decomposition. *IEEE Transactions on Signal Processing* 62 (2014), 531–544. https://api.semanticscholar.org/CorpusID:16819328
[13] Omar Eldwaik and Francis F. Li. 2018. Mitigating wind induced noise in outdoor microphone signals using a singular spectral subspace method. *Technologies* 6, 1 (2018), 19.
[14] Renata Graf. 2019. A multifaceted analysis of the relationship between daily temperature of river water and air. *Acta Geophysica* 67, 3 (2019), 905–920.
[15] D.C. Haworth. 2010. Progress in probability density function methods for turbulent reacting flows. *Progress in Energy and Combustion Science* 36, 2 (2010), 168–259. https://doi.org/10.1016/j.pecs.2009.09.003
[16] Jiale Hu, Yingying Wang, Dengjia Wang, Hu Du, Jianhua Fan, Yanfeng Liu, and Xin Sun. 2023. Indices for dynamic evaluation of indoor humidity and thermal environment. *Communications Engineering* 2, 1 (2023), 59.
[17] Andrei Nikolaevich Kolmogorov. 1991. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 434 (1991), 9–13. https://doi.org/10.1098/rspa.1991.0075
[18] Steven P. Lalley. 2013. Brownian Motion. (2013). https://galton.uchicago.edu/~lalley/Courses/313/BrownianMotionCurrent.pdf Accessed: 2024-10-29.
[19] Dong Li, Shirui Cao, Sunghoon Ivan Lee, and Jie Xiong. 2022. Experience: practical problems for acoustic sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 381–390.
[20] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2020. FM-track: pushing the limits of contactless multi-target tracking using acoustic signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems (SenSys '20)*. Association for Computing Machinery, New York, NY, USA, 150–163. https://doi.org/10.1145/3384419.3430780
[21] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. Lasense: Pushing the limits of fine-grained activity sensing using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.
[22] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. 2022. Room-scale hand gesture recognition using smart speakers. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 462–475.

[23] Ke Li, Ruidong Zhang, Boao Chen, Siyuan Chen, Sicheng Yin, Saif Mahmud, Qikang Liang, François Guimbretière, and Cheng Zhang. 2024. GazeTrak: Exploring Acoustic-based Eye Tracking on a Glass Frame. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 497–512.

[24] Yunzhi Li, Vimal Mollyn, Kuang Yuan, and Patrick Carrington. 2024. WheelPoser: Sparse-IMU Based Body Pose Estimation for Wheelchair Users. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–17.

[25] Yin Li, Rohan Reddy, Cheng Zhang, and Rajalakshmi Nandakumar. 2024. Beyond-Voice: Towards Continuous 3D Hand Pose Tracking on Commercial Home Assistant Devices. In *2024 23rd ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 151–162.

[26] Xiaoxuan Liang, Zhaolong Wei, Dong Li, Jie Xiong, and Jeremy Gummeson. 2024. SONDAR: Size and Shape Measurements Using Acoustic Imaging. In *Proceedings of the Twenty-fifth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*. 361–370.

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

[28] Jialin Liu, Dong Li, Lei Wang, Fusang Zhang, and Jie Xiong. 2022. Enabling contact-free acoustic sensing under device motion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.

[29] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. Lip reading-based user authentication through acoustic sensing on smartphones. *IEEE/ACM transactions on networking* 27, 1 (2019), 447–460.

[30] Gregory W Lyons, Carl R Hart, and Richard Raspet. 2021. As the Wind Blows: Turbulent Noise on Outdoor Microphones. *Acoustics Today* 17, 4 (2021), 20–28.

[31] Wenguang Mao, Jian He, and Lili Qiu. 2016. CAT: High-precision acoustic motion tracking. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 69–81.

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

[33] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 38, 16 pages. https://doi.org/10.1145/3300061.3345439

[34] Daniele Mirabilii and Emanuël A.P. Habets. 2019. Multi-channel Wind Noise Reduction Using the Corcos Model. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 646–650. https://doi.org/10.1109/ICASSP.2019.8683873

[35] Thomas J Mueller. 2002. *Aeroacoustic measurements*. Springer Science & Business Media.

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

[37] Jean Piquet. 1999. *Turbulent Flows: Models and Physics* (revised 2nd printing ed.). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-03559-7

[38] Stephen B. Pope. 2000. *Turbulent Flows*. Cambridge University Press.

[39] Stephen B. Pope. 2011. Simple models of turbulent flows. *Physics of Fluids* 23, 1 (01 2011), 011301. https://doi.org/10.1063/1.3531744

[40] RME Audio. 2024. Babyface Pro FS - High-end USB Audio Interface. https://rme-audio.de/babyface-pro-fs.html. (2024). Accessed: 2024-10-25.

[41] Ke Sun and Xinyu Zhang. 2021. UltraSE: single-channel speech enhancement using ultrasound. In *Proceedings of the 27th annual international conference on mobile computing and networking*. 160–173.

[42] Xue Sun, Jie Xiong, Chao Feng, Wenwen Deng, Xudong Wei, Dingyi Fang, and Xiaojiang Chen. 2023. Earmonitor: In-ear Motion-resilient Acoustic Sensing Using Commodity Earphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 182 (Jan. 2023), 22 pages. https://doi.org/10.1145/3569472

[43] JB Tary, RH Herrera, and M Van der Baan. 2014. Time-varying autoregressive model for spectral analysis of microseismic experiments and long-period volcanic events. *Geophysical Journal International* 196, 1 (2014), 600–611.

[44] Ashton Udall. 2019. How Ultrasound Sensing Makes Nest Displays More Accessible. https://blog.google/products/google-nest/ultrasound-sensing/. (2019). Accessed: 2024-09-19.

[45] Haoran Wan, Lei Wang, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, and Wei Wang. 2022. VECTOR: Velocity based temperature-field monitoring with distributed acoustic devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–28.

[46] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[47] Lei Wang, Tao Gu, Wei Li, Haipeng Dai, Yong Zhang, Dongxiao Yu, Chenren Xu, and Daqing Zhang. 2023. DF-Sense: Multi-user Acoustic Sensing for Heartbeat Monitoring with Dualforming. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3581791.3596867

[48] Lei Wang, Wei Li, Ke Sun, Fusang Zhang, Tao Gu, Chenren Xu, and Daqing Zhang. 2022. LoEar: Push the range limit of acoustic sensing for vital sign monitoring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–24.
[49] Tianben Wang, Daqing Zhang, Leye Wang, Yuanqing Zheng, Tao Gu, Bernadette Dorizzi, and Xingshe Zhou. 2018. Contactless respiration monitoring using ultrasound signal with off-the-shelf audio devices. *IEEE Internet of Things Journal* 6, 2 (2018), 2959–2973.
[50] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–20.
[51] Wei Wang, Alex X. Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking (MobiCom '16)*. Association for Computing Machinery, New York, NY, USA, 82–94. https://doi.org/10.1145/2973750.2973764
[52] Wikipedia contributors. 2024. Speed of Sound. (2024). https://en.wikipedia.org/wiki/Speed_of_sound Accessed: 2024-10-25.
[53] Xiangyu Xu, Jiadi Yu, Yingying Chen, Yanmin Zhu, Linghe Kong, and Minglu Li. 2019. BreathListener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the 17th annual international conference on mobile systems, applications, and services*. 54–66.
[54] Kuang Yuan, Shuo Han, Swarun Kumar, and Bhiksha Raj. 2024. DeWinder: Single-Channel Wind Noise Reduction using Ultrasound Sensing. In *Proc. Interspeech 2024*. 627–631.
[55] Kuang Yuan, Mohamed Ibrahim, Yiwen Song, Guoxiang Deng, Robert A Nerone, Suvendra Vijayan, Akshay Gadre, and Swarun Kumar. 2024. ToMoBrush: Exploring Dental Health Sensing Using a Sonic Toothbrush. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–27.
[56] Kuang Yuan, Yifeng Wang, Xiyuxing Zhang, Chengyi Shen, Swarun Kumar, and Justin Chan. 2025. SonicSieve: Bringing Directional Speech Extraction to Smartphones Using Acoustic Microstructures. *arXiv preprint arXiv:2504.10793* (2025).
[57] Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, and Daqing Zhang. 2020. Your Smart Speaker Can "Hear" Your Heartbeat! *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 161 (dec 2020), 24 pages. https://doi.org/10.1145/3432237
[58] Tianyue Zheng, Zhe Chen, Chao Cai, Jun Luo, and Xu Zhang. 2020. V2iFi: in-Vehicle Vital Sign Monitoring via Compact RF Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 70 (June 2020), 27 pages. https://doi.org/10.1145/3397321

A Physics-based Modeling for Speed of Sound Variation

As we discussed, the AR model is a widely used representation to describe time-varying natural phenomena. In this work, we adopt it to characterize the speed of sound variations under airflow, demonstrating good fitness to the data we collected. In this section, we present a theoretical analysis supported by physics-based modeling, to validate the physical soundness of the AR model in characterizing the sound speed variations.

Considering sound propagating in a statistically stationary turbulent air medium, we start by representing the speed of sound c as the formula below, where $c[t]$ is the effective speed of sound at time t, c_0 is the original sound speed in still air, and $v[t]$ is the velocity contribution from the moving turbulent medium:

$$c[t] = c_0 + v[t]. \tag{16}$$

Based on research work on turbulent flow in the fluid physics community [15, 37–39], stochastic models are commonly used to represent turbulent properties, especially when accounting for the random, fluctuating nature of turbulence. Pope [39] proposes a Langevin model to describe the evolution of fluid properties such as velocity over time. Specifically, a stochastic differential equation for velocity is introduced in the form of:

$$\mathrm{d}v[t] = -(\frac{1}{2} + \frac{3}{4}C_0)\frac{\epsilon}{k}v[t]\mathrm{d}t + (C_0\epsilon)^{\frac{1}{2}}\mathrm{d}W(t),$$

where C_0 is a universal constant in Kolmogorov's turbulence theory [17] with a value in the range of $1.4 \sim 1.8$. k represents the average energy per unit mass contained in the turbulent motion of fluid, and ϵ is the dissipation rate at which turbulent kinetic energy is lost due to the viscosity of fluid to create more small-scale structures [37], so a higher ϵ indicates more intensity of turbulence. We further simplify this formula as:

$$\mathrm{d}v[t] = -C_1 v[t]\mathrm{d}t + C_2\mathrm{d}W(t),$$

where $C_1 = (\frac{1}{2} + \frac{3}{4}C_0)\frac{\epsilon}{k}$ is a parameter related to the characteristics of turbulent flow, $C_2 = (C_0\epsilon)^{\frac{1}{2}}$ represents the intensity of the fluctuations, and $W(t)$ is a Wiener process [18] representing random noise. We denote $\eta[t]$ as a

random variable following standard normal distribution, such that $\int_t^{t+\Delta t} \mathrm{d}W(t) \approx \sqrt{\Delta t}\eta(t)$. By a forward time integration using the derivative, we can represent the change in velocity over a small time step Δt as:

$$v[t+\Delta t] = v[t] + \int_t^{t+\Delta t} \mathrm{d}v[t] = v[t] + \left(-C_1 v[t]\Delta t + C_2\sqrt{\Delta t}\eta(t)\right) = (1 - C_1\Delta t)v[t] + C_2\sqrt{\Delta t}\eta(t), \tag{17}$$

which shows how the velocity at the current time step linearly depends on the previous value, and a random noise term $\eta(t)$ represents the stochastic component of turbulence. Based on Eq. (16), we further denote $C = C_1\Delta t$ and $\varepsilon[t] = C_2\sqrt{\Delta t}\eta(t) \sim \mathcal{N}(0, \sigma^2)$, where $\sigma = C_2\sqrt{\Delta t}$. We finally obtain:

$$v[t+\Delta t] = (1 - C)\, v[t] + \varepsilon[t], \quad \varepsilon[t] \sim \mathcal{N}(0, \sigma^2),$$

which represents a first-order Autoregressive (AR) model. Higher-order AR models can be obtained using higher-order time integration schemes in Eq. (17), incorporating additional past time steps.