

Submillimeter Acoustic Vibration Measurement and Monitoring using a Single Smartphone

Liu Yang, Xiaofei Li, Wenwu Wang, *Senior Member, IEEE*, Xinheng Wang, *Senior Member, IEEE*, Peisong Li, *Student Member, IEEE*, Guangyao Liu and Zhi Wang, *Member, IEEE*

Abstract—Accurate vibration measurement is crucial for monitoring and diagnosing industrial equipment. Existing solutions require either installing contact sensors on the equipment or using non-contact sensors such as laser. Both approaches involve complex deployment, stringent environmental conditions, and high cost. As a better alternative, we propose a sub-millimeter acoustic vibration measurement system using a single smartphone, called Mobile-Vib. Firstly, we develop a novel acoustic ranging method that builds on traditional acoustic ranging techniques, incorporating the reflection principle of acoustic signals from vibrating objects. This approach addresses the challenge of acoustic signal refresh rate in vibration measurement by employing advanced signal design and processing techniques. Secondly, we design a noise removal algorithm utilizing the dual-channel technology of smartphones to minimize multipath signals and noise interference, enabling accurate phase estimation. To mitigate the impact of unrelated human motions in real-world measurements, we implement an optimisation-based method to correct distortions and reduce errors. Finally, by clarifying the relationship between phase changes and actual displacement, we enable tracking of vibration displacement in industrial environments. We have implemented Mobile-Vib, and the extensive experimental results demonstrate an average error of approximately 0.629 mm in displacement estimation and 5.6 Hz in frequency estimation at a 1-meter distance from the vibrating object in real industrial vibration monitoring scenarios.

Index Terms—Vibration monitoring, Acoustic, Smartphone, Echo ranging

I. INTRODUCTION

VIBRATION is a common occurrence in mechanical equipment. During normal operation, a mechanical device usually exhibits minimal vibration. However, when internal components undergo changes such as wearing, misalignment, looseness, or poor sealing, vibration displacement and

frequency increase [1], [2]. As a result, vibration measurement is widely used to assess the health of equipment in industrial environments [3], [4].

Traditional methods used for vibration measurement rely on contact-based specialized sensors [3], [5], such as accelerometers [6] and piezoelectric sensors [7], which are common devices utilizing small changes in acceleration to calculate vibration amplitude. However, due to error accumulation, the vibration measurements can contain measurement deviations. In addition, they need to be deployed at multiple locations on the equipment, leading to high costs. This approach is limited by its associated challenges in sensor deployment, maintenance, and subsequent data collection. In contrast, non-contact methods offer advantages in addressing these challenges and become increasingly popular, such as vibrometers [8], [9] and vision-based [10], [11] solutions. However, vision-based solutions require favorable on-site conditions. They are not applicable in low visibility or foggy environments. Additionally, laser-based solutions are costly and still face limitations when it comes to multipoint and large-scale vibration measurements.

With the development of wireless sensing technology, several studies have used electromagnetic waves for vibration measurement. For example, Radio Frequency Identification (RFID) solutions take advantage of the reflection characteristics of RF signals and their immunity to light interference by attaching passive RFID tags to the target object and measuring the vibration of the object through the RFID system, achieving vibration measurement at high precision [12], [13]. However, in these methods, tags are placed on the surface of the object to be detected, which imposes high requirements for sensor deployment and data collection. Millimeter-wave radar, on the other hand, analyzes the transmitted and received signals, extracts the phase of the signals, and uses phase variations to detect the changes in vibration amplitude. In [14], [15], multiple coherent observations of the same vibration signal are used to enhance geometric characteristics and improve vibration monitoring. Multi-Vib [16] proposed an innovative vibration marking design, enabling multi-point vibration monitoring with a single millimeter-wave radar, achieving micrometer-level vibration measurement. The unique wavelength characteristics of the millimeter wave radar enable micrometer-level vibration detection [1], [16], giving them a distinct advantage in wireless sensing solutions for vibration measurement. However, in many real-world industrial settings, the vibration amplitudes of various devices fall within the sub-millimeter to millimeter range [17], [18]. As a result, these applications do not require micrometer-level precision in vibration monitoring,

This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFB3900800 and in part by the Huzhou Science and Technology Plan Project under Grant 2024GZ68. (Corresponding author: Zhi Wang.)

Liu Yang, Guangyao Liu and Zhi Wang are with the College of Control Science and Engineering, Zhejiang University, Hangzhou 310000, China (e-mail: [liu.yang, 22232065, zjuwangzhi]@zju.edu.cn).

Xiaofei Li is with the School of Engineering, Westlake University, Hangzhou 310030, China, and also with the Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, China (e-mail: lixiaofei@westlake.edu.cn).

Wenwu Wang is with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: w.wang@surrey.ac.uk).

Xinheng Wang is with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China (e-mail: xinheng.wang@xjtlu.edu.cn).

Peisong Li is with the Huzhou Institute of Zhejiang University, Huzhou 313098, China, (e-mail: peisong.li20@alumni.xjtlu.edu.cn).

which would entail stringent environmental and deployment constraints. In addition, the high cost of the millimeter wave radar, along with the challenges of its deployment and data acquisition, makes it unsuitable for vibration measurements in scenarios requiring high mobility at any time.

Given the high cost of electromagnetic wave-based solutions in practical applications, some studies have adopted acoustic-based approaches for vibration detection [19]–[21]. For instance, the system presented in [22] achieves online non-intrusive detection of wood pellets in pneumatic conveying pipelines by capturing the vibrations and sounds generated by the collisions between particles and the pipeline. In [23], a fouling detection method is proposed for large-scale duct systems, using acoustic vibrational excitation for external, non-intrusive, and non-destructive fouling detection. Additionally, in [24], an ultrasonic signal with a single-frequency at 40 kHz is utilized to measure the vibrations of objects. However, current acoustic-based vibration detection methods primarily rely on ultrasonic signals, which are severely affected by air attenuation and face significant limitations in terms of vibration detection range. Moreover, such solutions often require customized equipment, making them unsuitable for vibration measurement needs in high-mobility scenarios.

Compared to customized acoustic solutions, smartphone-based acoustic approaches offer a low-cost and user-friendly alternative for vibration measurements. This approach enables vibration measurements to be performed anytime and anywhere, addressing the limitations of earlier methods in high-mobility, routine vibration monitoring scenarios, as well as overcoming the constraints on the detection range by traditional customized acoustic equipment.

Currently, there are several applications of acoustic sensing technology on commercial smartphones [25], [26]. Spirosonic [27] utilized smartphones to measure chest wall movements through phase analysis of single-frequency signals, thereby completing centimeter-level human pulmonary function tests. BlinkListener [28] quantitatively modeled the relationship between signal variations and subtle movements caused by blinking and interference, achieving millimeter-level human blink detection. These solutions analyze relatively slow-moving objects in noise-free environments, utilizing the characteristics of reflected signals to determine the features of the reflected objects, thereby achieving acoustic sensing.

However, these methods cannot be directly applied to vibration measurements in industrial environments, due to the following challenges. Firstly, in industrial environments, the vibration speed of objects is very high, far exceeding the frequency of movements such as respiration, while the requirement for vibration measurement accuracy is often at the sub-millimeter level. Secondly, there is significant multipath interference and low-frequency noise corruption in industrial environments. The received signals are often composed of reflected signals from vibrating objects and surrounding equipment. These signals, under the influence of low-frequency noise, make it difficult to extract the actual vibration signals. It is easy to misidentify noise as small variations in the vibration signal. Additionally, when using smartphones for vibration measurements in industrial settings, irrelevant

motions introduced by human hands can lead to errors in vibration measurements, affecting the actual results of the vibration measurements.

In order to address these challenges, we propose Mobile-Vib, a system for vibration measurement using a single smartphone. To this end, we first introduce the principles and methods of utilizing a single smartphone for vibration measurement, and the selection of custom acoustic signals. We then develop an innovative Sub-Frame Chirp (SFC) algorithm based on a ranging method, which resolves the issue of measuring vibrations when the object’s vibration frequency is much higher than the signal refresh rate. Unlike traditional solution [29] that rely on overlapping transmitted signals, our method avoids confusion between transmitted and received signals and ensures that the signal can meet high-precision ranging requirements in industrial scenarios. Thirdly, leveraging the dual microphones on the top and bottom of the smartphone, we propose a novel Top and Bottom Microphone (TBM) algorithm, combining the real-time computing power of smartphone. By analyzing the differences in the received signals through the physical propagation model of the dual-microphone channels, this algorithm effectively removes noise and multipath signals, achieving sub-millimeter detection accuracy. Lastly, we employ optimization algorithms to reduce vibration errors caused by hand movements and successfully achieve robust vibration tracking. To summarize, we have made the following novel contributions in this paper:

- We developed the Mobile-Vib system for vibration measurement, which, to our best knowledge, is the first such system based on acoustic signal using a single smartphone, achieving measurement accuracy at the sub-millimeter level.
- We designed SFC and TBM algorithms to address signal refresh rate issues and to mitigate multipath and noise effects in industrial environments. Using optimization methods, we corrected errors caused by hand motions during vibration measurement and subsequently implemented vibration monitoring based on the relationship between the variation of vibration and phase.
- We implemented Mobile-Vib on commercial smartphones and conducted extensive vibration measurement and monitoring experiments in both laboratory and real industrial environments, verifying the robustness of Mobile-Vib in complex industrial settings.

The remainder of this paper is organized as follows. We first introduce the principles of acoustic vibration measurement based on smartphones and the characteristics of the smartphone’s dual-microphone channels in Section II. Then, we discuss traditional methods for smartphone-based acoustic vibration measurement in ideal environments in Section III. In Section IV, we present a geometric fitting based signal processing method to remove interference in real-world environments, in conjunction with smartphone hardware. After this, we perform a systematic evaluation of commercial smartphone devices in Section V. In Section VI, we discuss the limitations of the system and its potential applications in industrial environments. Finally, we conclude the paper in

Section VII.

II. PROBLEM FORMULATION

The principle of acoustic vibration measurement based on smartphones involves using the smartphone's speaker to emit a custom signal $e(t)$, where t represents the time index. After being reflected by the vibrating object, the signal is received by the smartphone's microphone as the reflected signal $r(t)$. By calculating the time delay τ between transmitted and received signals, the distance between the smartphone and the vibrating object at a specific time stamp can be determined. Then, the smartphone calculates the distances $D(t)$ at two adjacent timestamps, denoted as $D(1)$ and $D(2)$, respectively. Subsequently, by taking the variation between these distances, the relative movement of the vibration $\Delta d = D(2) - D(1)$ is obtained. Finally, the vibration displacement is obtained by summing the relative movement of the vibration with the same sign over multiple consecutive timestamps. Given the relatively short duration of the reflected signal at each timestamp, it is assumed that the target object remains in a steady state while reflecting the acoustic signal. Fig. 1 illustrates the proposed smartphone-based acoustic vibration measurement system.

In real-world applications, the microphone in a smartphone receives the signals not only reflected by the vibrating object, but also directly propagated from its loudspeaker, as well as those reflected from other equipment in the industrial environment. The signals from different paths correspond to different time delays τ_i and different signal attenuation coefficients α_i , where i denotes the index of the path. According to the arrival times of the sound paths, the time delay τ_1 corresponding to the first path is considered as the direct path, which is the signal propagating directly from the smartphone's speaker to its microphone. The time delay τ_2 corresponding to the second path is taken as the reflection path from the vibrating object, which represents the actual reflected vibration signal (emitted from the smartphone's speaker, reflected off the vibrating object, and returning to the smartphone's microphone). Time delays τ_i , where $i > 2$, correspond to interfering reflection paths from other devices.

To effectively assess signal quality, the Signal-to-Noise Ratio (SNR) is commonly used for measurement. The signal refers to the valid signal corresponding to the reflection path, while the noise refers to the interference signal generated by mechanical equipment within the frequency range of the valid signal. A higher SNR indicates that the signal is stronger relative to the noise, resulting in a smaller impact of noise on

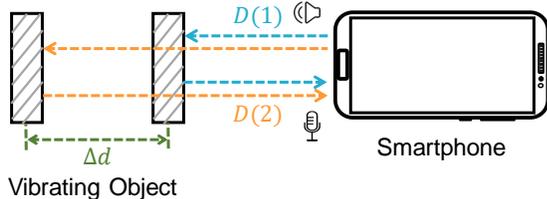


Fig. 1. The principle of estimating the vibration movement of a vibrating object using a single smartphone.

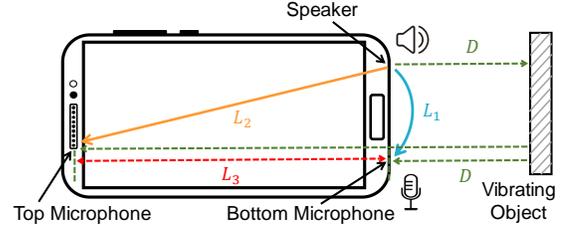


Fig. 2. The relationship between the actual distance from the smartphone to the vibrating object and the distances measured by the two microphones.

the signal in vibration monitoring, thereby enhancing system performance. In order to maximize the SNR in practice, efforts are made to keep the smartphone's screen plane perpendicular to the reflection plane of the vibrating object. This ensures that the sound waves hit the surface of the vibrating object perpendicularly and reflect back to the smartphone. This arrangement allows the shortest path method to be used during signal processing to distinguish direct signals, reflected signals, and interference signals from other devices. Although studies indicate that the size of the vibrating object affects the SNR [30], [31], in practical application scenarios targeted by the algorithm, we assume that the size of the vibrating object has a negligible impact on vibration monitoring.

However, there are other significant challenges that impact adversely on the accuracy of vibration measurements, including interference from multipath signals and ambient noise, as well as the Doppler effect on the reflected signals caused by the vibration of the object. Fig. 2 shows the layout of dual microphones on a smartphone. The smartphone features a dual-microphone design with the two microphones positioned at the bottom and top, respectively. The distances between the microphones and the speaker, corresponding to the direct path, are L_1 and L_2 , respectively. Meanwhile, the distance between the top microphone and the bottom microphone of the smartphone is L_3 , and the vertical distance from the surface of the vibrating object to the surface of the bottom microphone of the smartphone is D .

According to Fig. 2, when the acoustic signal is vertically reflected from the vibrating object back to the microphones, the distances $D^y(t)$ between the vibrating object and the bottom and top microphones at the same timestamp are defined to differentiate the measurements of the two microphones. Here, $y = 1$ or 2 , represents the index of the bottom and top microphones, respectively. The distance measured by the bottom microphone is given by $D^1(t) = (2D - L_1)/2$, while the distance measured by the top microphone is $D^2(t) = (2D + L_3 - L_2)/2$. A relationship between the actual distance from the smartphone to the vibrating object and the distances measured by the two microphones is established, which can be represented as follows:

$$D = \frac{2D^1(t) + L_1}{2} = \frac{2D^2(t) + L_2 - L_3}{2} \quad (1)$$

Additionally, at the same timestamp, there is a fixed difference between the distances measured by the two microphones, defined as $L_D = D^2(t) - D^1(t) = (L_3 + L_1 - L_2)/2$. By utilizing this fixed distance difference, the time difference between the two microphones is calculated, which helps

eliminate interference from multipath signals. At the same time, due to the symmetry of the target object's vibration in the same dimension, the signals from multiple timestamps are used for fitting, thereby mitigating the impact of the Doppler effect.

III. VIBRATION ESTIMATION BASED ON ACOUSTIC SIGNAL

In this section, the selection of custom signals and the vibration measurement methods based on smartphone acoustic signals are discussed, without considering the interference from multipath signals, noise, and other factors present in the actual environment.

A. Selection of custom signal

In terms of signal selection, acoustic chirp signals are chosen due to their strong correlation, which makes it easy to distinguish them from noise. An inaudible frequency range is also selected to reduce audible noise and improve measurement accuracy and resolution. The signal is defined as:

$$e(t) = \cos\left(2\pi f_{min}t + \frac{\pi Bt^2}{T}\right) \quad (2)$$

where f_{min} , B and T represent the initial frequency, bandwidth, and duration of the chirp signal, respectively.

In practical industrial scenarios, multipath effects are quite significant. The characteristic of the chirp signal, where the signal frequency changes linearly over time, gives it strong autocorrelation properties, making it robust against multipath interference.

Additionally, to enhance the robustness and effectiveness of the custom signal, a Hanning window $w(t) = 0.54 - 0.46 \cos(2\pi t/(T_w))$ is applied, where T_w is the duration of the Hanning window, which sharpens the correlation curve and accelerates sidelobe attenuation. Furthermore, considering the sampling rate limit of smartphones, which is normally 48 kHz, the signal frequency range of 17 kHz to 22 kHz is selected for acoustic vibration measurement, which does not affect users during operation, as these frequencies are inaudible to the human ear.

B. Vibration estimation based on a single smartphone

The distance $D(t)$ between the vibrating object and the smartphone can be divided into two parts, expressed as $D(t) = d_0 + d(t)$, where d_0 is the coarse distance and $d(t)$ is the fine distance. Coarse distance estimation is achieved through beat frequency estimation, and fine distance estimation is achieved through phase extraction. In this analysis, only the ideal case of a single vibrating object reflection is considered, with interference from multipath effects and noise being ignored.

1) *Beat frequency estimation*: The Frequency Modulated Continuous Wave (FMCW) method [29], [32] is used to achieve coarse distance estimation. For the custom chirp signal $e(t)$ emitted by the smartphone, its reflected signal was obtained after a time delay τ and a level attenuation α , as $r(t) = \alpha e(t - \tau)$, where $\tau = 2D(t)/c$, and c is the speed of sound. By using Trigonometric functions $\cos M \cos N =$

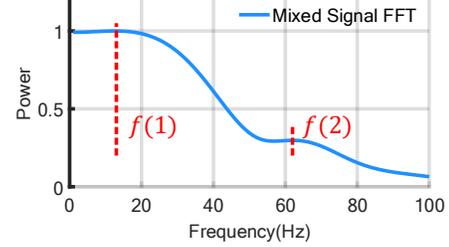


Fig. 3. Performing FFT on the mixed signals to obtain the frequencies of the direct and reflected signals, and estimating the beat frequency by calculating the frequency difference.

$(\cos(M - N) + \cos(M + N))/2$, and filtering out the irrelevant high frequency components, the resulting mixed signal $m(t)$ can be obtained by multiplying the emitted $e(t)$ and received signals $r(t)$ as:

$$m(t) = \frac{\alpha}{2} \cos\left(2\pi \frac{B}{T} \tau t + 2\pi f_{min} \tau - \frac{\pi B \tau^2}{T}\right) \quad (3)$$

The mixed signal is further simplified as:

$$m(t) = \frac{\alpha}{2} \cos(2\pi f t + \theta) \quad (4)$$

where $f = B\tau/T$ and $\theta = 2\pi f_{min}\tau - \pi B\tau^2/T$ denote the beat frequency and the initial phase of mixed signal $m(t)$, respectively.

The beat frequency is a function of time delay τ (and thus also a function of distance difference), therefore, it can provide the coarse distance d_0 of the reflecting object relative to the smartphone.

In practice, the Fast Fourier Transform (FFT) is utilized to extract the beat frequency by calculating the frequency difference between the direct signal corresponding to the time delay τ_1 and the reflected signal corresponding to the time delay τ_2 in the spectrum of the mixed signal. The frequency $f(1)$, corresponding to the maximum magnitude in the spectrum, represents the direct signal. The frequency of the actual reflected vibration signal, denoted as $f(2)$, normally corresponds to and is detected as the first peak following the maximum component as shown in Fig. 3.

Therefore, d_0 can be calculated by determining the frequency difference between the reflected signal and the direct signal:

$$d_0 = \frac{cB}{2T} \Delta f = \frac{cB}{2T} (f(2) - f(1)) \quad (5)$$

However, in practice, the distance between the smartphone and the vibrating device is relatively close and cannot be neglected compared to the distance between the microphone and speaker on the smartphone. Therefore, after calculating the absolute distance, it is necessary to compensate for the known distances L_1 , L_2 , and L_3 between the smartphone's microphone and speaker to obtain accurate coarse distance d_0 based on Eq. (1).

Specifically, the computational load of the smartphone is considered, and an FFT with a length of 1024×16 points is performed on the mixed signal $m(t)$, resulting in a spectrum with 8192 discrete frequency points. The frequencies $f(1)$

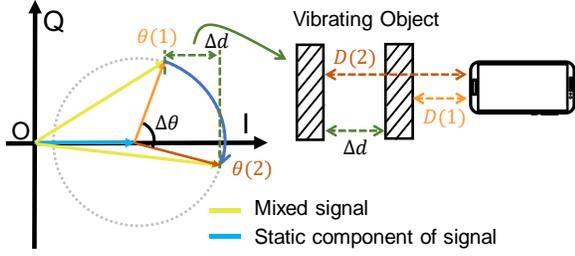


Fig. 4. The relationship between the angular variation of phase and the movement of the vibrating object.

and $f(2)$, corresponding to the maximum and the first peak following the maximum magnitudes, respectively, are then detected to calculate the frequency difference.

2) *Phase extraction*: The resolution in the coarse distance estimated with beat frequency is limited by the frequency bandwidth B of the chirp and can be calculated as $\frac{c}{2B}$. Therefore, when the coarse distance estimates at adjacent timestamps fall within the same range bin, meaning that the coarse distance estimates d_0 are the same at different times and within the precision resolution range, the vibration measurement Δd is only related to $d(t)$. Thus, vibration measurement can be achieved by solving the difference in fine distance estimates $d(t)$ through phase extraction at different times [27], [33]. However, vibration measurement with sub-millimeter precision is constrained by the vibration velocity of the target object, requiring that the object's vibration speed does not exceed the ratio of the coarse distance resolution to the time difference between adjacent timestamps.

According to Eq. (4), the specific expression for the initial phase $\theta = 2\pi f_{min}\tau - \pi B\tau^2/T$ can be obtained. Note that, since the quadratic term of τ is negligible, the initial phase is approximated as $\theta \approx 2\pi f_{min}\tau$. The relationship between the movement from distance $D(1)$ to $D(2)$ and the phase variation from the corresponding phase at distance $\theta(1)$ to $\theta(2)$ can be further obtained as follows:

$$\begin{aligned} \Delta d &= D(2) - D(1) = d(2) - d(1) \\ &= \frac{c(\tau(2) - \tau(1))}{2} = \frac{c}{2} \left(\frac{\theta(2) - \theta(1)}{2\pi f_{min}} \right) \end{aligned} \quad (6)$$

where $\tau(1)$ and $\tau(2)$ represent the time delay between the transmitted and received signal at time 1 and 2. Fig. 4 illustrates the relationship between the angular variation of mixed signal's phase and the movement of the vibrating object.

In practice, the beat frequency obtained from the coarse distance estimation is used to perform In-phase and Quadrature phase (IQ) decomposition on the mixed signal $m(t)$ [27], [34]. After performing the IQ decomposition, the phase values on the IQ coordinates can be expressed as:

$$\begin{aligned} I &= m(t) \times \cos(2\pi ft) = \frac{\alpha}{4} \cos(\theta) + \text{high frequency} \\ Q &= m(t) \times \sin(2\pi ft) = -\frac{\alpha}{4} \sin(\theta) + \text{high frequency} \end{aligned} \quad (7)$$

Using low-pass filtering, the actual initial phase can be determined as $\theta = \arctan(Q/I)$. Then, the initial phase $\theta(t)$ at different moments can be calculated based on the IQ decomposition, and the vibration movement is determined based

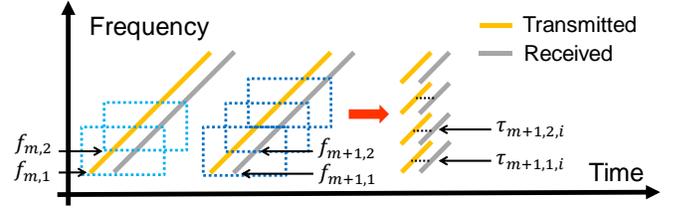


Fig. 5. Decomposing the chirp signal into sub-frame chirp signals.

on Eq. (6) by analyzing the initial phase variation between different moments. After that, the vibration displacement is obtained by summing the movement variations of the object with the same sign over consecutive timestamps.

IV. SYSTEM DESIGN AND IMPLEMENTATION

Our system implements vibration measurement based on smartphones. In practical industrial scenarios, issues like multipath signal effects and noise interference can arise. Therefore, an SFC algorithm is first designed based on smartphone audio signals, which helps address the signal refresh rate issue and reduces vibration measurement errors. Secondly, a TBM algorithm is established to mitigate multipath effects using the dual microphones at the top and bottom of the smartphone, thereby reducing noise through consistency constraint. Subsequently, leveraging the phase characteristics of the signal, the vibrational influence is corrected by quantifying geometric distortion. Ultimately, vibration tracking and estimation are conducted through the relationship between IQ samples and vibration movement.

A. Sub-Frame Chirp algorithm

In smartphone-based vibration measurement, as mentioned in Section III-B2, when the motion of a vibrating object fell within a specific range bin, the distance is obtained by calculating the phase variation between consecutive frames, i.e., between consecutive timestamp signals. However, this method is limited by the phase sampling rate. Assuming each frame of the signal lasts 50 ms, then 20 phase samples can be obtained per second. When the variation between adjacent phase samples exceeds π radians, indicating that the movement between two frames exceeds 5 mm, phase wrapping occurs due to the limitations of the phase sampling rate, leading to erroneous estimation of the phase variation [29], [35]. To avoid phase wrapping caused by signal refresh rate, an SFC algorithm is proposed to improve the refresh rate of vibration measurement.

The essence of our SFC lies in decomposing the chirp signal at the receiver end. Filters are utilized to break down each chirp signal into multiple sub-frame chirp signals, each characterized by different initial frequencies but with identical time duration and bandwidth, as shown in Fig. 5. Each of these sub-frame signals can be denoted as

$$e_{m,n}(t) = \cos(2\pi f_{m,n}t + \frac{\pi \tilde{B}t^2}{\tilde{T}}) \quad (8)$$

where m is the index of chirp signal, n is the index of sub-frame chirp signal, $f_{m,n}$ represents the initial frequency of

the n -th sub-frame chirp signal in the m -th chirp signal, \tilde{B} and \tilde{T} represent the constant bandwidth and time duration of sub-frame chirp signals, respectively. Each received sub-frame signal can be mixed with its corresponding transmitted sub-frame signal. Unlike the ideal mixed signal as presented in Eq. (3) that considers the entire chirp signal and for only one reflection path, now the mixed signal after applying the low-pass filter becomes

$$m_{m,n}(t) = \sum_{i=1}^N \frac{\alpha_{m,n,i}}{2} \cos\left(2\pi \frac{\tilde{B}}{\tilde{T}} \tau_{m,n,i} t + 2\pi f_{m,n} \tau_{m,n,i} - \frac{\pi \tilde{B} (\tau_{m,n,i})^2}{\tilde{T}}\right) \quad (9)$$

where $\alpha_{m,n,i}$ and $\tau_{m,n,i}$ represent the signal attenuation coefficient and time delay of the i -th path in the corresponding sub-frame signal, respectively. The mixed signal $m_{m,n}(t)$ can then be further simplified similarly to Eq. (4) as

$$m_{m,n}(t) = \sum_{i=1}^N \frac{\alpha_{m,n,i}}{2} \cos(2\pi f_{m,n,i} t + \theta_{m,n,i}) \quad (10)$$

where $f_{m,n,i} = \tilde{B} \tau_{m,n,i} / \tilde{T}$ and $\theta_{m,n,i} = 2\pi f_{m,n} \tau_{m,n,i} - \pi \tilde{B} \tau_{m,n,i}^2 / \tilde{T}$ represent the beat frequency and initial phase of the i -th path in the corresponding sub-frame signal, respectively.

The principle of identifying the peaks of direct and reflection signals, as shown in Fig. 3 is utilized to calculate the coarse distance for each sub-frame signal using the beat frequency, as described in Eq. (5). When the range bin of adjacent sub-frames are the same, the initial frequency of the signal does not affect the speed and delay of signal transmission, and the beat frequency of the mixed sub-frame chirp signal is the same as that of the mixed signal $m(t)$. The difference lies in the phase variation between them. In addition, given the brief duration of each sub-frame signal, it is assumed that the vibration pattern remains consistent across adjacent sub-frames. Using Eq. (6), the relationship between the phase variation of adjacent sub-frames and the vibration can be obtained as

$$\Delta d = \frac{c \Delta \theta}{2} = \frac{c}{2} \left(\frac{\theta_{m,n}}{2\pi f_{m,n}} - \frac{\theta_{m,n+1}}{2\pi f_{m,n+1}} \right) \quad (11)$$

where $\theta_{m,n}$ represents the initial phase of reflected signal from vibrating object in the corresponding sub-frame signal.

In the SFC algorithm, multiple phase samples are obtained within each frame of the chirp signal. The phase variation between two sub-frame chirp signals typically does not exceed the limit of π radians within 10 ms. Compared to traditional methods, this approach effectively improve the accuracy of fine distance estimation and increase the refresh rate for vibration measurement.

Specifically, each frame of the chirp signal is designed to have a duration of 50 ms, covering a 5 kHz frequency range from 17 kHz to 22 kHz. In the SFC algorithm, each chirp signal is decomposed into four (partially overlapped) sub-frame chirp signals with a bandwidth of 2 kHz and a duration of 20 ms. After performing the IQ decomposition

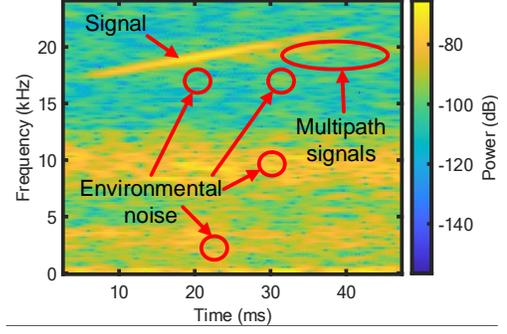


Fig. 6. Practical signal spectrum with direct path, multipath and noise.

on each mixed sub-frame signal, the phases of all mixed sub-frame signals are normalized based on their initial frequencies $f_{m,n}$, and the phase differences are taken based on the normalized results to obtain the actual phase variation, thereby achieving vibration measurement. It is important to note that our algorithm is typically used to measure vibrations below 50 Hz. However, for vibrations in the 50-100 Hz range, the parameters of the SFM algorithm need to be modified. The original sub-frame chirp signals (with a bandwidth of 2 kHz, a 1 kHz overlap, and a duration of 20 ms) are redesigned by setting the new signal bandwidth to 1 kHz, with a 500 Hz overlap and a duration of 10 ms, thereby improving the temporal resolution for measuring high-frequency signals and enabling vibration measurement at higher frequencies.

B. Top and Bottom Microphone algorithm

In industrial environments, multipath effects are very common, and there is also mid-to-low frequency environmental noise from vibrating objects and the surrounding environment. Fig. 6 shows the spectra of an acoustic signal recorded in real environment, from which multipath effect and noise can be observed. The multipath effects can be mitigated by putting the smartphone closer to the industrial equipment. However, in certain hazardous industrial environments, it is necessary to maintain a safe distance between the smartphone and the industrial equipment. Therefore, a TBM algorithm is proposed to reduce multipath interference and ensure the accuracy of the system.

High-precision vibration measurement relies on accurate phase extraction. From Eq. (11), it is evident that phase variation is influenced by the initial frequency f_{min} and the vibration movement $d(t)$. When the initial frequency changes while the vibration movement remains constant, the phase of the mixed signal changes. In the IQ domain, this results in the signal rotating around the origin of coordinates. Therefore, if the same vibration can be simultaneously measured with chirp groups having different initial frequencies, ideally, the mixed signal corresponding to each chirp rotates around a point in the coordinate system and forms a large arc. Thus, the TBM algorithm is established step by step as follows: 1) fitting of IQ phase sample circles for dual-microphone channel, 2) fitting circles for IQ phase samples of multiple sub-frame chirp

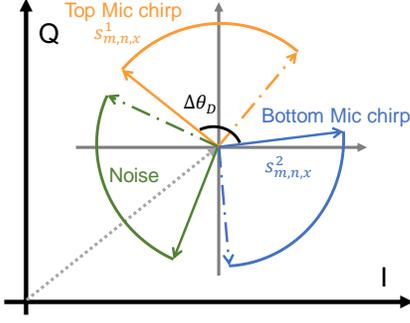


Fig. 7. Filtering out noise signals that do not meet the fixed phase variation between the signals received by the dual microphone channels.

signals from the same microphone, 3) merging fitted circles corrections to form a large arc.

1) *Fitting circles for the dual channel*: The proposed SFC algorithm is first separately applied to the signals received by two microphones. According to Eq. (7), any IQ sample in the IQ phase coordinates can be represented as $s_{m,n,x}^y$, where $y = 1, 2$ represent the bottom and top microphones, respectively, and x is the index of the number of sample points contained in each IQ sample, ranging from 0 to X . It is also worth noting that each IQ sample contains the same number of sample points, i.e. X . Subsequently, the IQ phase samples received simultaneously by the dual-microphone channels are subjected to circle fitting, by transforming the fitting into an optimization problem to obtain an optimized circle with center Z_c and radius R_c , where the summed geometric distance from each IQ sample point to the circle is minimized:

$$Z_c, R_c = \arg \min_{Z_c, R_c} \sum_{y=1}^Y \sum_{x=1}^X (\|s_{m,n,x}^y - Z_c\| - R_c)^2 \quad (12)$$

This optimization problem can be solved using the Gaussian-Newton algorithm.

In complex environment, IQ samples contain multipath effects resulting from reflections from both static and dynamic objects. Therefore, it is necessary to combine the characteristics of the dual microphones to improve the estimation robustness. The constant phase difference $\Delta\theta_D$ between the signals received by the bottom and top microphones is calculated in advance using the constant distance difference L_D between the smartphone's bottom and top microphones to the vibrating object, as shown in Fig. 2.

$$\Delta\theta_D = \text{unwrap}\left(\frac{4\pi f_{min} L_D}{c}\right) \quad (13)$$

where the unwrap function resolves the phase ambiguity. This known phase difference is used to filter out IQ phase samples that do not meet the requirements from the fitted circle, as illustrated in Fig. 7.

2) *Fitting circles for the same microphone*: In traditional FMCW method, only a single chirp signal is transmitted at a given time. However, through the SFC algorithm, multiple sub-frame chirp samples are obtained within a range bin. Since each sub-frame chirp is shorter compared to the whole chirp signal, and all the sub-frame chirps fall within the same range bin, the sub-frame chirp signals within a set of chirps can be

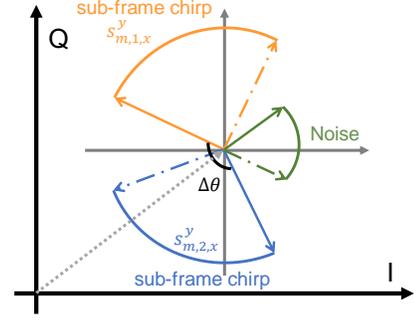


Fig. 8. Filtering out IQ samples from multipath signals that cannot fall on the concentric arc due to low SNR, where the impact of the Doppler effect is eliminated by fitting multiple sub-frame signals.

considered to occur simultaneously. This means that the ideal IQ samples fit into the arc that forms a concentric circle.

Our purpose is to utilize the symmetry of the object's motion in the same dimension to reduce the impact of the Doppler effect caused by reflected signals. By fitting multiple sub-frame signals to eliminate the Doppler effect, and filtering out IQ samples from multipath signals that cannot fall on the concentric arc due to low SNR, the IQ samples of each sub-frame signal are placed on this concentric circle. This principle is illustrated in Fig. 8.

In practice, the impact of the Doppler effect is first reduced by using the relationship between phase and IQ samples, applying Eq. (12) to obtain the fitted circle and solve the optimization problem. Then, a bias coefficient is established to measure the extent to which the IQ samples deviate from the fitted circle, which can be represented as:

$$K_{m,n} = \frac{1}{R_c} \frac{1}{X} \sum_{x=1}^X \|\|s_{m,n,x}^y - Z_c\| - R_c\| \quad (14)$$

Based on some preliminary experiments, if the arc distance from the fitted circle exceeds half of the fitted circle's radius, R_c , i.e. $K_{m,n} > 0.5$, the corresponding IQ sample is considered as noise, which will then be eliminated.

3) *Fitted circles correction*: As shown in Fig. 9, by combining the first and second steps, IQ samples identified as noise in both stages are filtered out. The multiple concentric circles are then unified, and the IQ samples of the sub-frame chirp signals from the two microphones are concatenated. Eq. (11) is

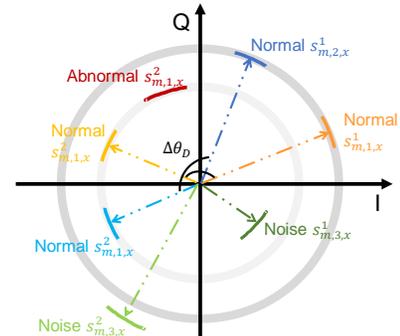


Fig. 9. Calculating the phase variation between the IQ phase samples corresponding to adjacent sub-frame signals after fitting all IQ samples into a circle.

then used to calculate the phase variation based on the initial frequency of each sub-frame chirp signal. Subsequently, the Quartile method [36] is employed to obtain the first three quartiles Q_1 , Q_2 , and Q_3 of the data set, respectively. Using these three quartiles, the upper and lower bounds of the data set are established, which can be represented as

$$\begin{cases} E_{max} = Q_3 + p(Q_3 - Q_1) \\ E_{min} = Q_1 - p(Q_3 - Q_1) \end{cases} \quad (15)$$

where p is a coefficient, typically taken as 1.5. The phase variation points outside the upper E_{max} and lower bounds E_{min} are considered as outliers and are excluded. This proposed TBM method is summarized in Algorithm1.

Algorithm 1 TBM Algorithm

Require: The mixed sub-frame chirp signal $m_{m,n}(t)$, the number of sub-frame chirp signals N in the m -th chirp signal

Ensure: Valid Phase variation J

- 1: Array $U = 0$, $W = 0$, $S_U = 0$, $S_W = 0$
 - 2: **for** $n = 2$ to N **do**
 - 3: Perform IQ decomposition for $m_{m,n}(t)$ to obtain IQ samples $s_{m,n,x}^y$, y is the index of microphones, ranging from 1 to 2.
 - 4: Fit circles based on Eq. (12) for IQ samples $s_{m,n,x}^y$ to obtain phase $\theta_{m,n}^y$ corresponding.
 - 5: **if** $(\theta_{m,n}^1 - \theta_{m,n}^2)$ satisfy known phase difference $\Delta\theta_D$ **then**
 - 6: $U \leftarrow s_{m,n,x}^1$, $W \leftarrow s_{m,n,x}^2$
 - 7: **end if**
 - 8: **end for**
 - 9: **for** $i = 1$ to $\text{length}(U)$ **do**
 - 10: Fit circles based on Eq. (12) for U and W
 - 11: Calculate bias coefficient K_U and K_W based on Eq. (14) for U and W
 - 12: **if** $K_U < 0.5$ && $K_W < 0.5$ **then**
 - 13: $S_U \leftarrow U(i)$, $S_W \leftarrow W(i)$
 - 14: **end if**
 - 15: **end for**
 - 16: Perform **Algorithm 2** on S_U and S_W to correct distortion and obtain valid IQ samples H_U and H_W
 - 17: Calculate $\Delta\theta_U$ and $\Delta\theta_W$ based on Eq. (11) for IQ samples in H_U and H_W
 - 18: Filter out outliers in $\Delta\theta_U$ and $\Delta\theta_W$ based on Eq. (15)
 - 19: $J \leftarrow \frac{\Delta\theta_U + \Delta\theta_W}{2}$
 - 20: **return** J
-

C. Distortion correction

In practice, the unrelated motion when holding the smartphone will affect the phase trajectory of the received signal, leading to irregular phase variations and causing errors in vibration measurements. Therefore, to address the phase deviation caused by unrelated motion, a distortion correction method is designed for the phase of sub-frame chirp signals to resolve the phase bias.

The normal IQ samples of the vibration signal should all fall on arcs of the same radius, as shown in Fig. 9. However,

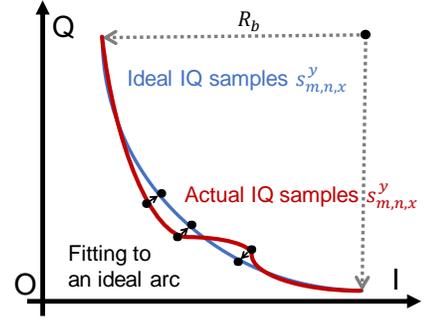


Fig. 10. Correcting the distortion of actual IQ samples to the ideal arc.

influenced by the unrelated motion of the hand, the actual IQ samples do not completely fall on the arcs of the same radius. This discrepancy can lead to erroneous estimation of the phase variation when normalized at the arc center, affecting the accuracy of vibration measurements.

The core idea of this method is to correct the IQ samples of each sub-frame chirp signal, aligning them to fall on arcs with the same radius. In our IQ coordinate system, each arc segment corresponding to an IQ sample is small, so the phase deviation caused by changes in arc radius due to unrelated motion has a weaker effect compared to the phase variation caused by vibration.

Therefore, this correction process is performed with a linear transformation, using the center Z_c and radius R_c obtained from Eq. (12) to correct each point in the IQ samples, ensuring that these IQ sample points fall on the arcs of the same radius. The Least Squares Method [37] is applied, with the weight matrix \mathbf{G} describing the deviation of each IQ sample point from the estimated center, to correct each IQ sample point. Specifically, the optimal center Z_b , denoted as (Z_{bx}, Z_{by}) , and radius R_b are estimated for each IQ sample, where the initial estimate of the radius R_b is R_c , and the initial estimate of the center Z_b is Z_c . The IQ sample is estimated as follows:

$$[\hat{Z}_{bx}, \hat{Z}_{by}, \hat{R}_c]^T = (\mathbf{A}^T \mathbf{G} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{G}^{-1} \mathbf{P} \quad (16)$$

where

$$\mathbf{A} = \begin{bmatrix} 2\|s_{m,n,1}^y|_X - Z_{bx}\| & 2\|s_{m,n,1}^y|_Y - Z_{by}\| & -1 \\ 2\|s_{m,n,2}^y|_X - Z_{bx}\| & 2\|s_{m,n,2}^y|_Y - Z_{by}\| & -1 \\ \vdots & \vdots & \vdots \\ 2\|s_{m,n,X}^y|_X - Z_{bx}\| & 2\|s_{m,n,X}^y|_Y - Z_{by}\| & -1 \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} \|s_{m,n,1}^y - Z_b\|^2 \\ \|s_{m,n,2}^y - Z_b\|^2 \\ \vdots \\ \|s_{m,n,X}^y - Z_b\|^2 \end{bmatrix},$$

$$\mathbf{G} = \frac{1}{X} \sum_{x=1}^X (\|s_{m,n,x}^y - Z_b\| - R_b)^2 \mathbf{I} \quad (17)$$

where \mathbf{I} is the identity matrix, $|s_{m,n,x}^y|_X$ and $|s_{m,n,x}^y|_Y$ respectively represent the value of the IQ sample $s_{m,n,x}^y$ on the horizontal and vertical axes.

Then, the result estimated by the Least Squares Method is used to map each sample individually to the arc, as shown in Fig. 10. In the presence of unrelated motion, this approach

ensures that the error in vibration measurement remains close to the error observed when the target is stationary, thus meeting the requirements for accurate vibration measurement. This proposed distortion correction method is summarized in Algorithm 2.

Algorithm 2 Distortion correction algorithm

Require: The center Z_b and radius R_b of the fitted circles, the valid IQ samples of two microphones B_U and B_W

Ensure: Valid IQ samples H_U and H_W

- 1: Array $H_U = 0, H_W = 0$
 - 2: **for** $i = 1$ to $\text{length}(B_U)$ **do**
 - 3: Calculate the center \hat{Z}_b and radius \hat{R}_c of estimated arc by Eq. (16) for $B_U(i)$ and $B_W(i)$, where \mathbf{A}, \mathbf{P} and \mathbf{G} are based on Eq. (17)
 - 4: Calculate the IQ samples $h_U(i)$ and $h_W(i)$ after correction from the center and radius of estimated arc
 - 5: $H_U \leftarrow h_U(i), H_W \leftarrow h_W(i)$
 - 6: **end for**
 - 7: **return** H_U, H_W
-

D. Vibration tracking

Based on the above circular fitting result, the relationship between the IQ samples $s_{m,n,x}^y$ and the phase $\Delta\theta_{m,n}$ is obtained, which can be represented as

$$\Delta\theta_{m,n} = \arctan\left(\frac{Q_{m,n+1} - Q_{m,n}}{I_{m,n+1} - I_{m,n}}\right) \quad (18)$$

where $I_{m,n}$ and $Q_{m,n}$ respectively represent the values of the IQ samples $s_{m,n,x}^y$ on the I and Q coordinates. Here, \arctan denotes the arctan function. When the result is positive, it indicates counterclockwise rotation of the IQ samples over time between the preceding and subsequent sub-frames.

According to Eq (6) and Eq. (18), it can be determined that when the IQ samples $s_{m,n,x}^y$ rotated clockwise over time, the vibrating object vibrates towards the direction of the smartphone, as shown in Fig. 11. Subsequently, by fitting the relationship between the variation of the circular arc over time, the relationship between phase variation and vibration movement can be derived, thereby enabling real-time vibration tracking. Then, by utilizing the relationship between phase variation and vibration direction, the phase difference between the IQ samples corresponding to the first and last sub-frame signals in the fitted circle is calculated during the period when

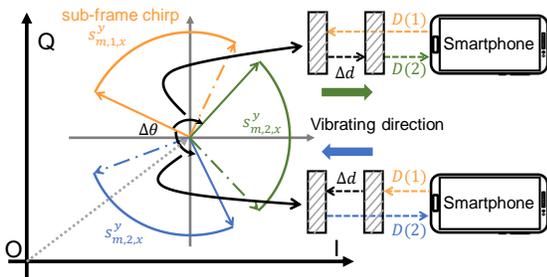


Fig. 11. The relationship between vibrating direction and IQ samples rotating direction.

the vibration direction remains unchanged, thus achieving vibration displacement measurement. Furthermore, this tracking can be used to calculate the vibration frequency of industrial equipment. Given the irregular nature of industrial equipment vibrations, the vibration frequency is defined as the number of directional changes in vibration movement within one second.

E. System capabilities

Mobile-Vib is a real-time vibration measurement and tracking system based on smartphones. The maximum measurable vibration displacement of the system is designed to satisfy the requirement of not exceeding half a phase cycle, that is, $\frac{c \times 180^\circ C}{4 \times 180^\circ C \times f_{\min}} = 5\text{mm}$, with the theoretical minimum value being 0. The maximum measurable frequency complies with the Nyquist sampling theorem, which is $\frac{1s}{2 \times 5\text{ms}} = 100\text{Hz}$, with the theoretical minimum value being 0. Furthermore, the system has a refresh rate of 100 Hz for measurement results. When the vibration speed of the object is within the speed range measurable by the Mobile-Vib system, and the SNR meets the vibration measurement requirements, the system can measure and track vibration displacements of 5 mm and vibration frequencies within 100 Hz.

However, the minimum measurable vibration displacement and frequency in practical use are limited by factors such as SNR and measuring distance. Under the requirements of SNR and object vibration speed, the system can measure a minimum vibration displacement of 0.482 mm and a minimum vibration frequency of 2.1 Hz at a distance of 0.3 meters from the vibrating object, with a relative error of no more than 50%, as shown in Section V-D.

V. EXPERIMENTAL EVALUATION

In this section, the implementation of Mobile-Vib and its experimental evaluation in our laboratory and the engine room of a small ship are introduced.

A. Implementation setting

1) *Implementation:* The proposed Mobile-Vib system utilizes the commercial smartphone Huawei P20 Pro and employs an ultrasonic frequency range of 17 kHz to 22 kHz for vibration measurement, which is beyond the range of human hearing. We have set the duration of the chirp signal to 50 ms. The sampling rate for both the speaker and the microphone is 48 kHz. The phone's volume is maintained at 80%, and all signal processing and analysis are conducted on the smartphone. In the ultrasonic measurement method, a power amplifier and a piezoelectric speaker are connected to the audio output of a laptop as the signal generator, while a custom high frequency response MEMS microphone serves as the signal receiver, as shown in Fig. 12(a). A 40 kHz ultrasonic signal is used for vibration measurement. The sampling rate for both the signal generator and the receiver is 96 kHz, and the laptop volume is maintained at 80%.

It is worth noting that temperature affects the speed of sound. Based on the relationship between temperature and sound speed [38], [39], we set the sound speed at 340 m/s,

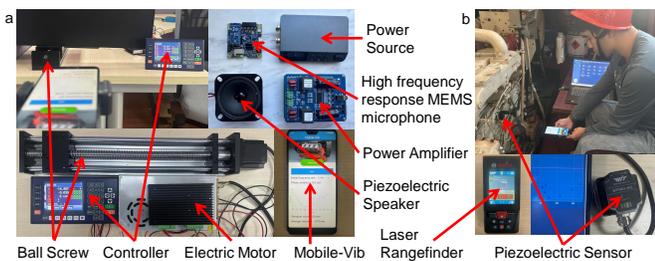


Fig. 12. Experimental setting. (a) In a laboratory environment, the vibration of an object is simulated by inputting displacement and frequency at the control end of the setup. The ball screw works with the motor to convert rotational motion into linear motion, thereby simulating the vibration of the object. A metal plate is mounted on the motion slider of the setup as the reflective surface. The Mobile-Vib system, or an ultrasonic-based signal generator with a high-frequency response MEMS microphone and a signal receiver consisting of a power amplifier and a piezoelectric speaker, is placed at a short distance from the reflective surface to measure the vibration. (b) In a case study of engine inspection on a real ship, the readings from the piezoelectric sensor are used as the ground truth for measuring the vibration of the object. The operator holds the Mobile-Vib system at a short distance to measure the vibration and uses a laser rangefinder to determine the distance between the operator and the vibrating object.

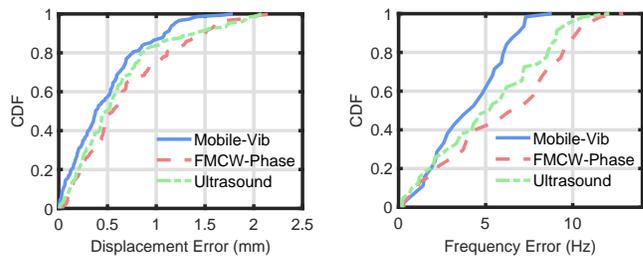
corresponding to 15°C , ensuring that the vibration displacement error caused by a maximum temperature difference of 15°C within the $0-30^{\circ}\text{C}$ range does not exceed 0.13 mm and remains within an acceptable range. For use under extreme temperature conditions, the speed of sound can be directly adjusted in the system according to the current temperature, ensuring that the vibration measurement error remains within this range.

2) *Experiment setting*: In a laboratory setting, we use an electric motor (Leadshine DM542) as the vibration source, with a metal plate positioned above it as shown in Fig. 12(a). A smartphone is placed on a tripod. We evaluate Mobile-Vib based on vibration distance and frequency. Additionally, we use Mobile-Vib to monitor real-world vibrations of the engine of a small ship, as shown in Fig. 12(b).

3) *Ground truth*: In the laboratory experiments, we use an electric motor to generate vibrations in the direction of the smartphone, with frequencies of 25 Hz , 50 Hz , and 100 Hz , and respective amplitudes of 1 mm and 2 mm . Simultaneously, we use a laser rangefinder (Bosch GLM 100-25C) to measure the distance between the smartphone and the steel plate on the engine. In the ship's engine room, we place a piezoelectric sensor (WITMOTION WTVB01-485) on the engine casing to obtain vibration data via a computer interface.

4) *Comparison*: We compare Mobile-Vib with the traditional acoustics-based phase method mentioned in Section III-B (referred to as FMCW-Phase [29]) and the ultrasonic measurement method (referred to as Ultrasound [24]), using the same experimental setup conditions and data pre-processing techniques for all three methods.

5) *Metrics*: We evaluate the performance of Mobile-Vib using displacement and vibration frequency estimation errors. The displacement error is obtained by comparing the ground truth value with the estimated displacement. The frequency error is derived from the variance in the vibration signal within a 1 s interval and the deviation from the ground truth value.



(a) CDF of displacement estimation (b) CDF of frequency estimation

Fig. 13. Overall performance of Mobile-Vib.

B. Benchmark experiment

In this section, we evaluate the performance of Mobile-Vib based on the estimation of distance and frequency.

1) *Overall performance*: We implemented the Mobile-Vib in the user space of the existing Huawei P20 Pro device, achieving an average processing time of about 200 ms for 1 s audio samples, which ensures real time processing of vibration detection. The runtime of the algorithm components, encompassing the SFM, TBM, correction, and vibration tracking algorithms is shown in Table I.

TABLE I
STATISTICS ACROSS 50 MEASUREMENTS RUNTIME OVER A 1 S SIGNAL.

Operation	Runtime (ms)
SFC algorithm	43.5 ± 5.7
TBM algorithm	120.1 ± 12.9
Distortion Correction	30.6 ± 3.1
Vibration Tracking	2.1 ± 0.6
Total	196.3 ± 22.3

Fig. 13 shows the performance of vibration measurement for the proposed Mobile-Vib method and the comparison method. The smartphone was placed at a distance of 0.5 m from the normal engine equipment. The results indicate that mean error of the Mobile-Vib in vibration displacement estimation is 0.484 mm , and the mean error in vibration frequency estimation is 4.4 Hz . Compared to the FMCW-Phase and Ultrasound methods, Mobile-Vib demonstrates superior noise suppression and robustness against multipath signal interference, effectively reducing displacement and frequency errors.

2) *Impact of different measuring distances*: In this experiment, we evaluate the vibration measurement performance of Mobile-Vib at different measuring distances in a laboratory setting. We conducted tests at distances of 0.3 m , 0.5 m , and 1 m from the vibrating object. Motor vibration is set to have 50

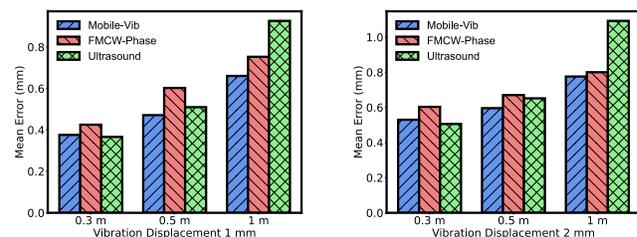


Fig. 14. Impact of measurement distance on displacement estimation.

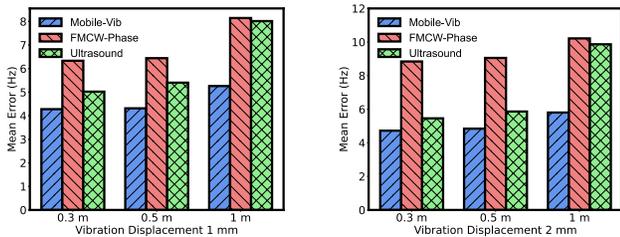


Fig. 15. Impact of measurement distance on frequency estimation.

Hz frequency, with 1 mm or 2 mm displacements, and towards the direction of the smartphone. Figs. 14 and 15 illustrate the vibration estimation mean errors of both the Mobile-Vib and FMCW-Phase systems at these three distances.

Under 1 mm vibration displacement, the mean error in vibration displacement estimation for Mobile-Vib at a distance of 0.5 m is 0.471 mm, while the mean error in vibration frequency estimation is 4.2 Hz. However, the error increases with increase in the measuring distance. In contrast, Ultrasound demonstrates better accuracy in both vibration displacement and frequency estimation at short distances. Nevertheless, as the distance increases, Ultrasound is significantly affected by air attenuation, leading to a substantial decrease in accuracy. At a distance of 1 m, the maximum relative error in vibration displacement exceeds 90%. FMCW-Phase, on the other hand, is more severely impacted by noise and multipath effects. While achieving comparable accuracy to Mobile-Vib in vibration displacement estimation, this method results in an error of 6.4 Hz in vibration frequency estimation. These results also suggest that the FMCW-Phase method struggles in extracting the correct vibration signal, and its performance can be significantly degraded by random environmental factors.

3) *Impact of different vibration frequencies:* We conducted experiments to evaluate the impact of vibration frequencies on the accuracy of vibration detection. Initially, we positioned the smartphone at a distance of 0.5 m from the vibrating

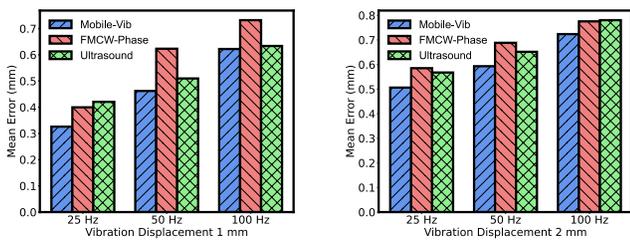


Fig. 16. Impact of vibration frequencies on displacement estimation.

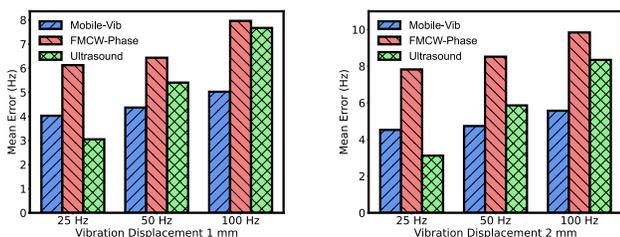


Fig. 17. Impact of vibration frequencies on frequency estimation.

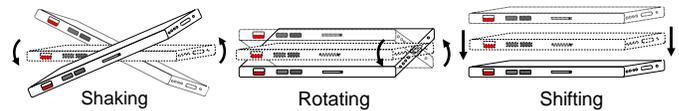


Fig. 18. Smartphone irrelevant motions.

object. Subsequently, we utilized a motor to induce vibrations at frequencies of 25 Hz, 50 Hz, and 100 Hz, with respective displacements of 1 mm and 2 mm. We then employed Mobile-Vib to detect the vibrations generated by the motor.

Figs. 16 and 17 display the vibration errors of Mobile-Vib and the FMCW-Phase scheme. For a 1 mm vibration displacement, the mean error in vibration displacement estimation at 50 Hz for Mobile-vib is 0.462 mm, and the mean error in vibration frequency estimation is 4.2 Hz. Ultrasound shows a smaller error in vibration frequency estimation for vibration at 25 Hz, but as the vibration frequency increases, its accuracy significantly drops below that of Mobile-Vib. Compared to the FMCW-Phase method, Mobile-Vib shows improvements in both vibration displacement and frequency estimation performance. In addition, it is observed that as the vibration frequency increases, the vibration error also increases. This is because, with an increase in frequency, the number of sampling points collected by the system within each vibration cycle decreases, leading to larger measurement errors between sampling points. However, it is worth noting that the errors generated by high-frequency vibration of the object do not cause significant impact on vibration detection.

C. Effect of practical factors

In this section, we evaluated the practical impact of real-world factors on Mobile-Vib. We placed the smartphone 0.5 m in front of the motor and induced 1 mm and 2 mm vibrations at a frequency of 50 Hz.

1) *Impact of irrelevant motions:* We evaluated the effectiveness of Mobile-Vib in removing irrelevant motions. During practical vibration measurements using Mobile-Vib, unavoidable slight movements of the hand can affect the vibration measurement results. We decomposed the slight movements of the hand into three types of motions: shaking, rotating, and shifting, as shown in Fig. 18. The angles for shaking and rotating range in between ± 20 degrees, while the vertical displacement for shifting is about 5 cm.

Subsequently, we placed the mobile phone at a distance of 0.5 m from the vibrating object, and set a motor vibration

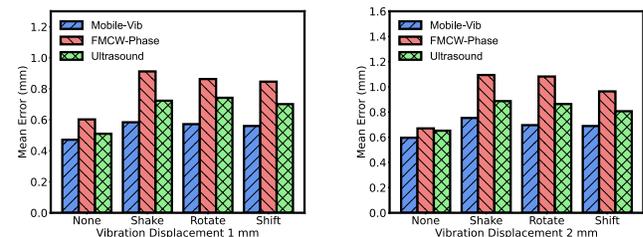


Fig. 19. Impact of irrelevant motions on displacement estimation.

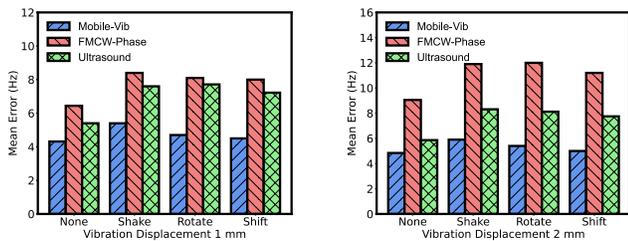


Fig. 20. Impact of irrelevant motions on frequency estimation.

frequency of 50 Hz. Figs. 19 and 20 illustrate the vibration displacement and frequency estimation results under these three types of motions, for Mobile-Vib and FMCW-Phase. The displacement error under 1 mm vibration in Mobile-Vib for shaking is 0.584 mm, with a frequency error of 5.4 Hz. For rotating, the displacement error is 0.572 mm, with a frequency error of 4.8 Hz, and for shifting, the displacement error is 0.561 mm, with a frequency error of 4.5 Hz.

From the results, it is evident that shaking produces the largest error. This is likely due to the fact that when the phone is shaking, the displacement variation between the phone's microphone and speaker relative to the vibrating object is large, whereas the displacement variations for shifting and rotating are relatively smaller. Compared to FMCW-Phase and Ultrasound, Mobile-Vib reduces the error by more than 20% when handling irrelevant motions. This result also indicates that, even when affected by hand motion, Mobile-Vib can ensure that the measurement errors remain within an acceptable range.

2) *Impact of different surroundings:* To study the robustness of the system in noisy environments, we tested Mobile-Vib under different environmental noise conditions. In the laboratory setting, we played the noise recorded from the ship's engine room and the water flow noise generated during the ship's actual operation. We define the SNR as the ratio of the signal strength received by Mobile-Vib in the 17-22 kHz effective signal frequency range, free from mechanical

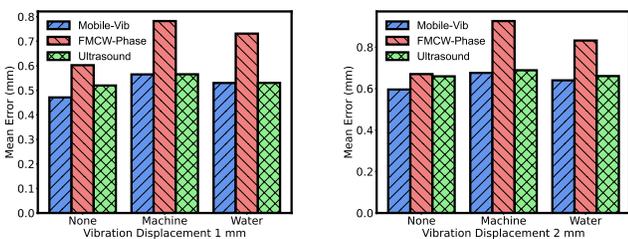


Fig. 21. Impact of surroundings on displacement estimation.

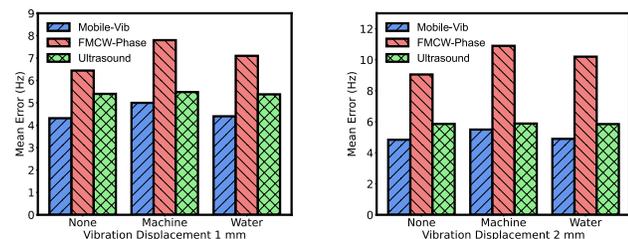
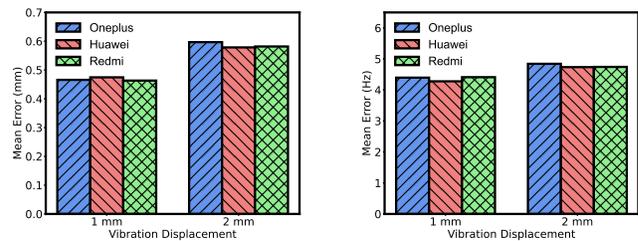


Fig. 22. Impact of surroundings on frequency estimation.



(a) Vibration displacement errors

(b) Vibration frequency errors

Fig. 23. Impact of smartphone models on vibration estimation.

equipment interference, to the noise strength generated by mechanical equipment. The SNR in the engine room noise background alone was 18.4 dB, and the SNR in the water flow noise background alone was 26.3 dB. Figs. 21 and 22 illustrate the vibration measurement errors of Mobile-Vib and FMCW-Phase under these two types of noise signals. It was observed that the water flow noise signal has a smaller impact on Mobile-Vib compared to the ship's engine room noise. The mean error in vibration 1 mm displacement estimation under the water flow noise is 0.531 mm, while the mean error in vibration frequency estimation is 4.4 Hz. This is due to the fact that the SNR of the noise from the ship's engine room is lower than that of the water flow noise, and its frequencies, similar to those used in Mobile-Vib vibration detection, interfere with signal transmission, causing larger measurement errors.

Compared to FMCW-Phase, Mobile-Vib can reduce the displacement estimation error by 30% and the frequency estimation error by 50%. Ultrasound, due to its higher frequency, is less affected by environmental noise. However, Mobile-Vib still manages to reduce the vibration displacement and frequency estimation errors by about 10% compared to Ultrasound. Considering the overall error results, Mobile-Vib is able to constrain the errors caused by these two types of noise signals within an acceptable range, thus not significantly affecting the vibration detection results.

3) *Impact of different smartphone models:* In practical applications, operators could use Mobile-Vib to conduct vibration detection on various models of smartphones. Therefore, to validate the applicability of Mobile-Vib, we conducted a comprehensive assessment using multiple models of smartphones in our experiments. In this study, we evaluated three models of Android smartphones: OnePlus 8 Pro, Redmi, and Huawei P20 Pro. As depicted in Fig. 23, we observed that the mean error in vibration displacement estimation and vibration frequency estimation using Mobile-Vib on the three devices was similar, with relative errors within a 3% range. This indicates that Mobile-Vib is capable of meeting the requirements of various smartphones, demonstrating its universality.

D. Case study of engine inspection of real ship

We conducted practical experiments by using a Huawei P20 Pro smartphone equipped with Mobile-Vib to conduct engine inspections within the engine room of a small ship. Fig. 24 depicts our actual experimental setup, where we mounted a piezoelectric sensor on the engine casing. The data acquired from this sensor was transmitted via a wired connection to a host computer, serving as the ground truth for the vibration measurements.

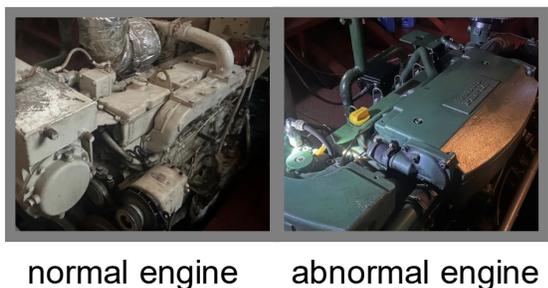


Fig. 24. Actual test engine with two status.

In the actual experiments, we conducted tests on known normal engines and engines with slight anomalies. It is important to note that the anomalies in the engines were due to minor loosening of their screws, and the extent of these anomalies was not sufficient to pose a risk to normal operation or to the personnel conducting the inspections. Therefore, we, equipped with protective gear, manually held smartphones and ensured that the engines operated at minimal power. We conducted vibration measurements on the normal engine at distances of 0.3 m, 0.5 m, and 1 m from the engine. For safety reasons, we only performed vibration tests on the anomalous engine at a distance of 1 m. Fig. 25 shows the measurement error of vibration displacement and frequency, and Fig. 26 illustrates an example clip of movement measurement on normal and abnormal engines. These results demonstrate the system capabilities of Mobile-Vib, showing that under real-world conditions, the vibration displacement estimation error for the normal engine at a distance of 1 m is 0.629 mm, and the vibration frequency estimation error is 5.6 Hz. For the anomalous engine at a distance of 1 m, the displacement error is 0.663 mm, and the frequency error is 5.1 Hz. This indicates that Mobile-Vib is capable of achieving sub-millimeter-level vibration measurements in practical applications.

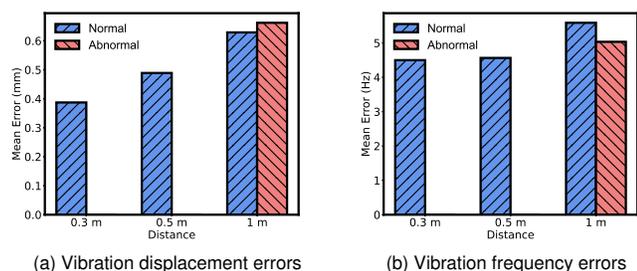


Fig. 25. The performance of Mobile-Vib in real-world environment.

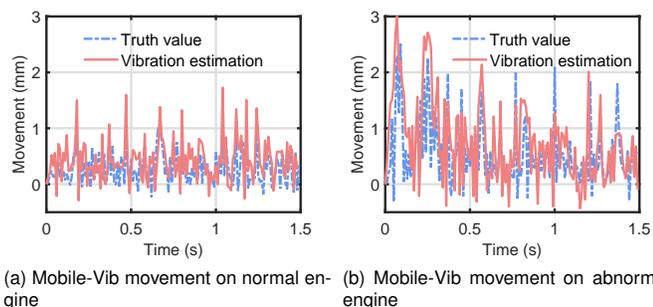


Fig. 26. The result of Mobile-Vib movement.

VI. DISCUSSION

In this section, we discuss the limitations of the system and its potential applications in industrial environments.

Mobile-Vib is subject to certain limitations in terms of its algorithm. Although our algorithm leverages the symmetry of vibrations within the same dimension and uses a fitting approach to mitigate the effects of the Doppler shift, it is still constrained by velocity. Specifically, Mobile-Vib requires the object's motion speed to remain below the coarse distance resolution between two consecutive timestamps, which is limited to $\frac{340\text{m/s}}{2 \times 5\text{kHz} \times 5\text{ms}} = 6.8\text{m/s}$. Apart from this limitation, the system also faces performance limitations. In practical applications, the minimum measurable displacement and frequency are constrained by factors such as signal sampling rate, SNR, and measuring distance, making it difficult to achieve theoretical limits. This was verified in the case study conducted during the experiments. Under conditions where the relative error of vibration remains below 50%, and the distance between Mobile-Vib and the vibrating object is 0.3 m, the minimum detectable displacement was measured at 0.482 mm, and the minimum measurable frequency was 2.1 Hz. These results deviate from the theoretical values of 0 mm and 0 Hz.

In practical applications, we will take advantage of the fast, low-cost, and portable characteristics of Mobile-Vib to apply it in industrial scenarios where noise intensity is moderate, mechanical operations do not pose risks to human safety, and proximity to equipment is feasible during operation in low-temperature environments. Furthermore, we can extend the vibration measurement of the system to three dimensions. Although a single measurement can only detect vibrations on the plane perpendicular to the smartphone screen, the portability of the system allows for manual adjustments to the measurement orientation. By ensuring that the vibration dimensions of multiple measurements are non-parallel, we can utilize orthogonal decomposition to project the measured vibration values onto a three-dimensional coordinate system, thereby achieving three-dimensional vibration measurement of the target object.

VII. CONCLUSION

In this paper, we have proposed Mobile-Vib for sub-millimeter vibration measurements based on smartphones. In the Mobile-Vib system, a SFC algorithm was designed to improve signal-to-noise ratio and resolve phase wrapping issues, thereby reducing vibration errors. Additionally, we developed a TBM algorithm, which emulates multiple antennas using the dual microphones of a smartphone to separate the vibration signal of the vibrating object from multipath noise. Finally, we corrected phase distortion in the signal, effectively addressing phase deviations caused by irrelevant motions from human movements.

Our evaluation demonstrates that Mobile-Vib performs well in both laboratory and field environments, achieving accurate vibration measurements. For vibrations with a displacement of 1 mm, the mean displacement error is 0.462 mm and frequency error is 4.2 Hz.

REFERENCES

- [1] J. Guo, Y. He, C. Jiang, M. Jin, S. Li, J. Zhang, R. Xi, and Y. Liu, "Measuring micrometer-level vibrations with mmwave radar," *IEEE Transactions on Mobile Computing*, vol. 22, no. 4, pp. 2248–2261, 2021.
- [2] S. W. Doebbling, C. R. Farrar, M. B. Prime, and D. W. Shevitz, "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review," 1996.
- [3] T. Sakorikar, H. J. Pandya *et al.*, "An mems-based force sensor: Packaging and proprioceptive force recognition through vibro-haptic feedback for catheters," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.
- [4] E. P. Carden and P. Fanning, "Vibration based condition monitoring: a review," *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, 2004.
- [5] J. Liu, C. Wang, Y. Chen, and N. Saxena, "Vibwrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 73–87.
- [6] A. Sabato, C. Niezrecki, and G. Fortino, "Wireless mems-based accelerometer sensor boards for structural vibration monitoring: A review," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 226–235, 2016.
- [7] S. Kon, K. Oldham, and R. Horowitz, "Piezoresistive and piezoelectric mems strain sensors for vibration detection," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2007*, vol. 6529. SPIE, 2007, pp. 898–908.
- [8] A. Di Donato, L. Scalise, and L. Zappelli, "Noncontact speckle-based velocity sensor [laser doppler vibrometer]," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 1, pp. 51–57, 2004.
- [9] Y. Zhang, G. Laput, and C. Harrison, "Vibrosight: Long-range vibrometry for smart environment sensing," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 2018, pp. 225–236.
- [10] J. Zhou, H. Li, L. Zhang, X. Wang, and Y. Li, "Vibration measurement with video processing based on alternating optimization of frequency and phase shifts," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [11] D. Zhan, D. Jing, M. Wu, D. Zhang, L. Yu, and T. Chen, "An accurate and efficient vision measurement approach for railway catenary geometry parameters," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 12, pp. 2841–2853, 2018.
- [12] L. Yang, Y. Li, Q. Lin, H. Jia, X.-Y. Li, and Y. Liu, "Tagbeat: Sensing mechanical vibration period with cots rfid systems," *IEEE/ACM Transactions on Networking*, vol. 25, no. 6, pp. 3823–3835, 2017.
- [13] B. Xie, J. Xiong, X. Chen, and D. Fang, "Exploring commodity rfid for contactless sub-millimeter vibration sensing," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 15–27.
- [14] S. Bi, X. Gao, V. M. Lubecke, O. Boric-Lubecke, D. Matthews, and X. L. Liu, "A multi-arc method for improving doppler radarmotion measurement accuracy," in *2018 IEEE/MTT-S International Microwave Symposium-IMS*. IEEE, 2018, pp. 244–247.
- [15] D. K. Kim and Y. Kim, "Quadrature frequency-group radar and its center estimation algorithms for small vibrational displacement," *Scientific Reports*, vol. 9, no. 1, p. 6763, 2019.
- [16] Y. Yang, H. Xu, Q. Chen, J. Cao, and Y. Wang, "Multi-vib: Precise multi-point vibration monitoring using mmwave radar," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–26, 2023.
- [17] L. Zhao, X. Huang, Y. Zhao, and W. Si, "Design of a wireless vibration metre for conductor vibration monitoring," *Structural Control and Health Monitoring*, vol. 25, no. 4, p. e2143, 2018.
- [18] H. Czichos, *Handbook of Technical Diagnostics: Fundamentals and Application to Structures and Systems*. Springer Science & Business Media, 2013.
- [19] N. Zhou, S. Zhong, J. Lin, M. Luo, W. Nsengiyumva, Z. Peng, and Y. Yu, "Acoustic-excitation optical coherence vibrometer for real-time microstructure vibration measurement and modal analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 7209–7217, 2020.
- [20] F. Figueroa and E. Barbieri, "An ultrasonic ranging system for structural vibration measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, no. 4, pp. 764–769, 1991.
- [21] R. Morello, C. De Capua, and A. Meduri, "A wireless measurement system for estimation of human exposure to vibration during the use of handheld percussion machines," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2513–2521, 2010.
- [22] D. Sun, Y. Yan, R. M. Carter, L. Gao, G. Lu, G. Riley, and M. Wood, "On-line nonintrusive detection of wood pellets in pneumatic conveying pipelines using vibration and acoustic sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 993–1001, 2013.
- [23] J. J. da Silva, A. M. N. Lima, F. H. Neff, and J. S. da Rocha Neto, "Non-invasive fast detection of internal fouling layers in tubes and ducts by acoustic vibration analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 108–114, 2008.
- [24] M. Young and Y. Li, "A high precision ultrasonic system for vibration measurements," *Review of Scientific Instruments*, vol. 63, no. 11, pp. 5435–5441, 1992.
- [25] H. Wan, S. Shi, W. Cao, W. Wang, and G. Chen, "Respracker: Multi-user room-scale respiration tracking with commercial acoustic devices," in *IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.
- [26] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao, "Strata: Fine-grained acoustic-based device-free tracking," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 15–28.
- [27] X. Song, B. Yang, G. Yang, R. Chen, E. Forno, W. Chen, and W. Gao, "Spirosonic: monitoring human lung function via acoustic sensing on commodity smartphones," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–14.
- [28] J. Liu, D. Li, L. Wang, and J. Xiong, "Blinklistener: "listen" to your eye blink using your smartphone," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–27, 2021.
- [29] J. Liu, D. Li, L. Wang, F. Zhang, and J. Xiong, "Enabling contact-free acoustic sensing under device motion," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, pp. 1–27, 2022.
- [30] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. Princeton University Press, 1986.
- [31] V. Twersky, "On scattering and reflection of sound by rough surfaces," *The Journal of the Acoustical Society of America*, vol. 29, no. 2, pp. 209–225, 1957.
- [32] W. Mao, J. He, and L. Qiu, "Cat: High-precision acoustic motion tracking," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 69–81.
- [33] A. Wang, J. E. Sunshine, and S. Gollakota, "Contactless infant monitoring using white noise," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.
- [34] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 82–94.
- [35] W. Mao, M. Wang, W. Sun, L. Qiu, S. Pradhan, and Y.-C. Chen, "Rnn-based room scale hand motion tracking," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–16.
- [36] D. Luo, X. Wan, J. Liu, and T. Tong, "Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range," *Statistical Methods in Medical Research*, vol. 27, no. 6, pp. 1785–1805, 2018.
- [37] Å. Björck, "Least squares methods," *Handbook of Numerical Analysis*, vol. 1, pp. 465–652, 1990.
- [38] B. N. Postma and B. F. Katz, "Correction method for averaging slowly time-variant room impulse response measurements," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. EL38–EL43, 2016.
- [39] A. Nowoświat, M. Olechowska, and M. Marchacz, "The effect of acoustical remedies changing the reverberation time for different frequencies in a dome used for worship: A case study," *Applied Acoustics*, vol. 160, p. 107143, 2020.



Liu Yang received the B.S. degree from Zhejiang University of Technology, Hangzhou, China, in 2023. He is currently pursuing the Ph.D. degree in degree in control science and engineering with Zhejiang University, Hangzhou, China. His research interests include acoustic signal processing, mobile systems and industrial equipment vibration measurement.



Xiaofei Li is an assistant Professor at Westlake University, China. Before, he worked at INRIA Grenoble Rhone-Alpes, Grenoble, France, as a post-doctoral researcher from Feb. 2014 to Jan. 2016, and as a starting research scientist from Feb. 2016 to Dec. 2019. He obtained his PhD degree from Peking University, China, in Jul. 2013. His research interests lie in the field of acoustic, audio and speech signal processing, including the topics of speech de-noising, dereverberation, separation and localization; sound/speech semi-supervised learning and unsupervised pre-training; sound field reproduction and personal sound zone.



Wenwu Wang (Senior Member, IEEE) was born in Anhui, China. He received the B.Sc., M.E., and the Ph.D. degrees from Harbin Engineering University, China, in 1997, 2000, and 2002, respectively. He then worked with King's College London, Cardiff University, Tao Group Ltd. (now Antix Labs Ltd., Reading, U.K.), and Creative Labs, Staines, U.K., before joining University of Surrey, U.K., in May 2007, where he is currently a Professor in Signal Processing and Machine Learning and Associate Head in External Engagement, School of Computer

Science and Electronic Engineering, University of Surrey, UK. He is also an AI Fellow at the Surrey Institute for People Centred Artificial Intelligence. His current research interests include signal processing, machine learning and perception, artificial intelligence, machine audition (listening), and statistical anomaly detection. He has (co-)authored over 300 papers in these areas. He has been recognized as a (co-)author or (co-)recipient of more than 15 awards, including the 2022 IEEE Signal Processing Society Young Author Best Paper Award, ICAUS 2021 Best Paper Award, DCASE 2020 and 2023 Judge's Award, DCASE 2019 and 2020 Reproducible System Award, and LVA/ICA 2018 Best Student Paper Award. He is an Associate Editor (2020-2025) for IEEE/ACM Transactions on Audio Speech and Language Processing and Associate Editor (2024-2026) for IEEE Transactions on Multimedia. He was a Senior Area Editor (2019-2023) and Associate Editor (2014-2018) for IEEE Transactions on Signal Processing. He is the elected Chair (2023-2024) of IEEE Signal Processing Society (SPS) Machine Learning for Signal Processing Technical Committee, a Board Member (2023-2024) of IEEE SPS Technical Directions Board, the elected Chair (2025-2027) and Vice Chair (2022-2024) of the EURASIP Technical Area Committee on Acoustic Speech and Music Signal Processing, an elected Member (2021-2026) of the IEEE SPS Signal Processing Theory and Methods Technical Committee. He was a Satellite Workshop Co-Chair for IEEE ICASSP 2024 and INTERSPEECH 2022, a Publication Co-Chair for IEEE ICASSP 2019, Local Arrangement Co-Chair of IEEE MLSP 2013, and Publicity Co-Chair of IEEE SSP 2009. He is a Special Session Co-Chair of IEEE MLSP 2024, and Technical Program Co-Chair of IEEE MLSP 2025.



Xinheng Wang (Senior Member, IEEE) received the B.E. and M.Sc. degrees in electrical engineering from Xian Jiaotong University, Xi'an, China, in 1991 and 1994, respectively, and the Ph.D. degree in electrical engineering from Brunel University, Uxbridge, U.K., in 2001. He was the founding Head of Department of Mechatronics and Robotics, Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou 215123, China. Prior to joining XJTLU, he was a professor in computing and Computer Networks with different universities in the UK. He has been an Investigator or Co-Investigator of 30+ research projects sponsored from EU, U.K., EPSRC, Innovate U.K., China NSFC, and industry. The total research funding that he has received is over \$6.8 million (50 million CNY). He has authored or coauthored 250+ referred papers. He holds 26 granted patents, including 1 US, 1 Japan, 4 South Korea and 20 Chinese patents. He was one of the key developers of the world's first smart trolley to provide intelligent services to passengers at airports. He is currently leading the XJTLU-uGo Robotics Research Centre with multi-million Yuan sponsorship from industry to develop robots for airport services. His current research interests include intelligent and connected systems; robotic SLAM; path planning; acoustic communications, localization, and sensing; and mobile edge computing.



Peisong Li (Student Member, IEEE) is currently with the Hangzhou Institute of Technology, Xidian University, Hangzhou 311200, China. He received a B.S. degree from the Guilin University of Electronic Technology in 2017, a M.S. degree from the Shanghai Maritime University, China, in 2020, and a Ph.D. degree from Xi'an Jiaotong-Liverpool University in 2024. His research interests include edge computing, vehicular edge computing, edge AI, and deep reinforcement learning.



Guangyao Liu received the B.S. degree from Central South University, China, in 2022. He is currently pursuing the Ph.D degree in Control Engineering, with Zhejiang University, China. His research interests include robot navigation, machine learning, acoustic positioning.



Zhi Wang (Member, IEEE) received the B.E. degree in mechanism from Shenyang Jianzhu University, Liaoning, China, in 1991, the M.Sc. degree in mechanical engineering from Southeast University, Nanjing, Jiangsu, China, in 1997, and the Ph.D. degree in automated control from the Shenyang Institute of Automation, China Academy of Science, Liaoning, in 2000. His current research interests include indoor positioning, satellite navigation, and location-based services. He has co-authored over 100 publications in international journals (such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS) and conferences.