Journal of Dynamics, Monitoring and Diagnostics, 2025, 4, 22-30
https://doi.org/10.37965/jdmd.2025.719

ISTP
RESEARCH ARTICLE

High-Quality Sound Recovery of Moving Sound Source Using Millimeter-Wave Radio

Yuyong Xiong,[1] Zhaoyu Liu,[1] Xiangyi Tang,[1] Haibin Meng,[1] and Zhike Peng[1,2]
[1]State Key Laboratory of Mechanical Systems and Vibration, Shanghai Jiao Tong University, Shanghai, China
[2]School of Mechanical Engineering, Ningxia University, Yinchuan, China

(Received 19 December 2024; Revised 21 January 2025; Accepted 12 February 2025; Published online 12 February 2025)

Abstract: The emerging millimeter-wave microphones have garnered considerable attention in recent years due to their potential for sound detection in various applications, particularly in situations where traditional microphones may be impractical. However, despite their promise, there is a notable lack of evidence demonstrating high-quality sound recovery of moving sources, which remains a significant challenge in the field. This paper addresses this critical gap by proposing a novel method for displacement alignment that improves the detection and recovery of sound signals from moving sources. The proposed method works by first aligning the displacement of the sound source over time, which ensures that the signals are synchronized and avoids interference from movement of sources. Subsequently, precise surface vibrations are extracted from the aligned signals, providing data for sound recovery. A finite impulse response (FIR) filter is applied to remove low-frequency motion, which often interferes with the clarity of the detected sound. Experimental results demonstrate the method's effectiveness in recovering high-quality sound from moving sources, offering a promising solution for advancing the emerging millimeter-wave microphone technology in real-world applications. This work could pave the way for more accurate and reliable sound detection systems, particularly in dynamic environments.

Keywords: displacement alignment; millimeter-wave microphone; moving sound source; sound recovery

I. INTRODUCTION

Sound detection is crucial in applications such as speech communication, human-machine interaction, vibration reduction, and fault diagnosis. Over the past few decades, numerous sound detection methods have been proposed, including capacitive microphones [1], laser microphones [2,3], visual microphones [4,5], and microwave microphones [6,7]. These methods have demonstrated significant effectiveness in measuring static sound sources. However, with the increasing demand for measuring moving sound sources, conventional methods suffer numerous challenges.

Capacitor microphones, which use diaphragms to detect air vibrations and recover sound signals, have been widely used for decades. However, their effectiveness is limited to close-range applications, hindering their use in wide-area monitoring. Different from capacitor, laser microphones, visual microphones, and microwave microphones capture surface vibrations and convert them into audio signals. Among them, laser microphones are well-developed with high bandwidth [2], making them suitable for multi-source, high-quality sound detection. However, they require a flat surface flatness and are less suitable for general environments. With advancements in computer vision and the widespread use of cameras, visual microphones have gained significant attention in recent years. Leveraging high-speed video cameras, auxiliary lighting, and advanced signal processing techniques [4,5], visual microphones can simultaneously identify multiple sound sources within a wide field of view, making them a promising candidate for various sound detection environments.

Corresponding author: Zhike Peng (e-mail: z.peng@sjtu.edu.cn).

Most of the mentioned studies have demonstrated significant effectiveness in measuring static sound sources. However, when sound sources are in motion, these methods face challenges such as Doppler effect and reduced measurement accuracy. To improve detection performance for moving sound sources, multimodal techniques have been proposed [8]. These combine audio and visual information to improve sound detection performance. Audio offers the frequency of the sound source, while visual information provides details such as the motion trajectory. By fusing multimodal information, more accurate detection of moving sound sources can be achieved. Beamforming techniques [9], which weight signals received by a microphone array to form directional beams, track the position and trajectory of moving sound sources. However, both techniques require strict synchronization and calibration, with complex post-processing steps.

With the development of integrated communication perception, microwave antennas have been fully equipped in thousands of households and explored as a possibility for perception [10,11]. These highly integrated and stable systems have fueled interest in sound detection, particularly through microwave microphones [6], which capture surface vibrations using electromagnetic waves and convert them into audio signals. Microwave microphones have demonstrated micron-level displacement measurement accuracy [12–15], enabling the detection of most vibrations [16–18] and sound recovery [19]. They primarily operate in two modes: single-frequency continuous wave [13] (CW) and frequency-modulated continuous wave [6,20–22] (FMCW). The displacement detection by CW radar is mainly realized by arctangent demodulation [23], parameterized demodulation [13] and modified differentiate and cross-multiply (MDACM) technique [24]. It is commonly used for vocal signal detection but lacks the capability for multi-source

detection. FMCW radar addresses this limitation by segmenting space into multiple range bins, allowing sound detection specific to each bin. Based on these researches, microwave microphone has been applied in voice user interfaces [25], demonstrating its versatility and feasibility across various applications. Moreover, several efforts have been made in static clutter elimination [20,26], accuracy self-evaluation [21], super resolution [27,28], Noise suppression [29,30] and weak signal enhancement [31,32] to improve accuracy of displacement and performance of recovered sound. Microwave microphones, which harness the power of electromagnetic waves, exhibit minimal susceptibility to Doppler effects. They boast compact device dimensions and facilitate synchronization, rendering them exceptionally promising for the detection of moving sound sources. Nonetheless, prior researches in this realm have not yet been explored, and the task of achieving high-quality sound recovery from moving sources remains a challenge.

To address the challenge, this work proposes and validates a displacement alignment method for high-quality sound recovery from moving sound sources. The method mitigates the effects of movement by aligning the sequential positions of the targets. In aligned signals, moving sound sources can be regarded as pseudo-static, thereby presenting high quality sound signals of the moving sound sources. This method is validated through experiments.

II. DISPLACEMENT MEASUREMENT USING MILLIMETER-WAVE RADIO

A. SYSTEM OVERVIEW

As the carrier of microwave microphone, radar recovers audio by sensing the vibration of the surface of the sound source, which is essentially a displacement measurement. A FMCW radar system is employed to realize the displacement measurement. Figure 1 illustrates the schematic of displacement with FMCW radar. The transmitted FMCW millimeter-wave signal is generated by voltage-controlled oscillator (VCO) and amplified by power amplifier (PA), emitted by transmitting antennas (TX). Signals reflected from the sound source surface are received by the receiving antennas (RX) and processed through a low-noise amplifier (LNA), mixers, low-pass filter (LPF) and analog-to-digital converter (ADC), the baseband signals are obtained to carry out the sound detection. The displacement of the surface of the sound source is obtained from the phase component of baseband signals.

B. DISPLACEMENT MEASUREMENT BASED ON PHASE ESTIMATION

As shown in Fig. 2, assume that frequency modulated continuous wave (FMCW) is:

$$s(t,q) = \exp\left(j2\pi\left(f_0 t + \frac{K}{2}t^2\right)\right) \tag{1}$$

where $j^2 = -1$, t denotes sample time in q-th frequency chirp. f_0 is carrier frequency of FMCW signal, K is the frequency slope of modulated signals. Suppose there is a moving target with slant range R at $q = 0$, then the target's range can be given as

$$r(q) = R + x(q) \tag{2}$$

where $x(q)$ is the target's displacement in q-th frame. If the transmitted FMCW is reflected by the sound source and

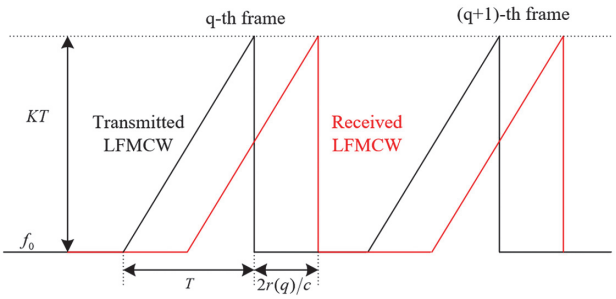

Fig. 2. Details of transmitted linear frequency modulated continuous wave.

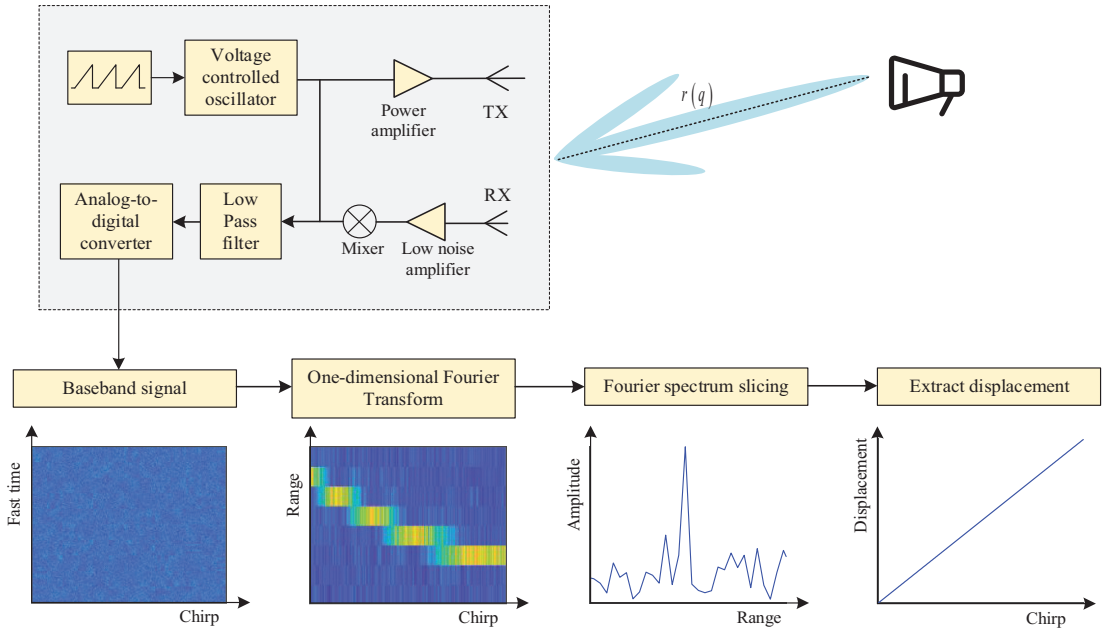

Fig. 1. Schematic of displacement detection with FMCW radar.

then received by RX, the baseband signal $s_B(t,q)$ can be represented as

$$s_B(t,q) = a \times \exp\left(j4\pi(Kt + f_0)\frac{r(q)}{c}\right) \tag{3}$$

where a is the amplitude of the received signal, c is the speed of the light.

After Fourier Transform, the heatmap within the q-th frame is

$$\begin{aligned} s_B(\widehat{R},q) &= \int_0^T s_B(t,n,q)\exp\left(-j4\pi K\frac{\widehat{R}}{c}t\right)dt \\ &\approx A(\widehat{R},q)\exp\left(j4\pi f_c\frac{r(q)}{c}\right) \end{aligned} \tag{4}$$

$$A(\widehat{R},q) = aT \times \exp\left(-j2\pi KT\frac{\widehat{R}}{c}\right)\mathrm{sinc}\left(2\pi KT\frac{r(q)-\widehat{R}}{c}\right) \tag{5}$$

where $f_c = f_0 + KT/2$, T is the maximum sampling time in a frame, $\widehat{R}$ denotes the estimated range of the target. Heatmap $s_B(\widehat{R},q)$ divides the range space into multiple range units to display the distribution of multiple targets. Boundary of the range units defined by

$$R - \widehat{R} + x(q) = \frac{\gamma c}{2KT},\quad \gamma = 1,2,\ldots \tag{6}$$

According to equation (5), the length represented by a range unit is $c/2KT$. If each range unit contains only one target and the target always stays in that range unit, then the estimated range $\widehat{R}$ can be regarded as constant. The estimated differential displacement $\widehat{x}(q) - \widehat{x}(q-1)$ can be obtained by cross-multiplying the $s_B(\widehat{R},q)$ and $s_B(\widehat{R},q-1)$, that is

$$\begin{aligned} \widehat{x}(q) - \widehat{x}(q-1) &= \frac{\lambda_c}{4\pi}\mathrm{angle}(s_B(\widehat{R},q)s_B^*(R,q-1)) \\ &= x(q) - x(q-1) \end{aligned} \tag{7}$$

For a sound source, such as a speaker, its outer surface vibrates when excited. Radar detects these vibrations by measuring the surface displacement, interpreting them as audio information. This approach could provide valuable insights for sound source detection in specific environments. However, if the target moves, the estimated range $\widehat{R}$ can no longer be considered constant, and phase changes in $A(\widehat{R},q)$ will distort displacement estimation. According to equation (5), when the total displacement exceeds half a range unit c/4KT, a phase shift in $A(\widehat{R},q)$ inevitably occurs. This situation is unavoidable in practical applications, necessitating a new method for detecting moving sound sources.

III. SOUND DETECTION AND RECOVERY OF MOVING SOUND SOURCE

A. DISPLACEMENT ALIGNMENT FOR MOVING TARGET

To cope with the effect of target movement on displacement measurements, it is necessary to remove time-varying variable $r(q)$ from $A(\widehat{R},q)$. Suppose $\widehat{x}(q)$ is a reference displacement. We define an operator $\Phi(t,\widehat{x}(q))$ as:

$$\Phi(t,\widehat{x}(q)) = \exp\left(-j4\pi(Kt + f_0)\frac{\widehat{x}(q)}{c}\right) \tag{8}$$

By multiplying $s(t,q)$ with $\Phi(t,\widehat{x}(q))$, we obtain a demodulated signal as

$$\begin{aligned} u(t,q) &= s_B(t,q)\Phi(t,\widehat{x}(q)) \\ &= a \times \exp\left(j4\pi(Kt + f_0)\frac{R + x(q) - \widehat{x}(q)}{c}\right) \end{aligned} \tag{9}$$

After applying the Fourier Transform, the demodulated signal $u(t,q)$ in range domain could be stated as

$$\begin{aligned} u(\widehat{R},q) &= aT \times \exp\left(j4\pi f_c\frac{R + x(q) - \widehat{R} - \widehat{x}(q)}{c}\right) \\ &\quad \times \mathrm{sinc}\left(2\pi KT\frac{R + x(q) - \widehat{R} - \widehat{x}(q)}{c}\right) \\ &= A(\widehat{R},q,\widehat{x}(q))\exp\left(j4\pi f_c\frac{x(q) - \widehat{x}(q)}{c}\right) \end{aligned} \tag{10}$$

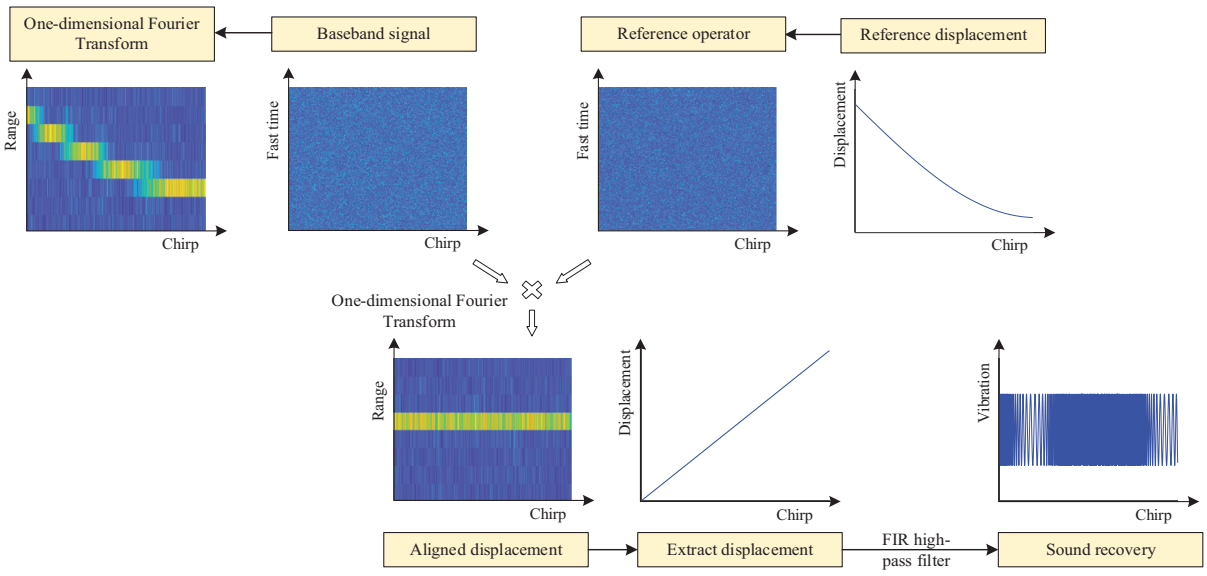

Fig. 3. Schematic of sound detection with millimeter-wave radar.

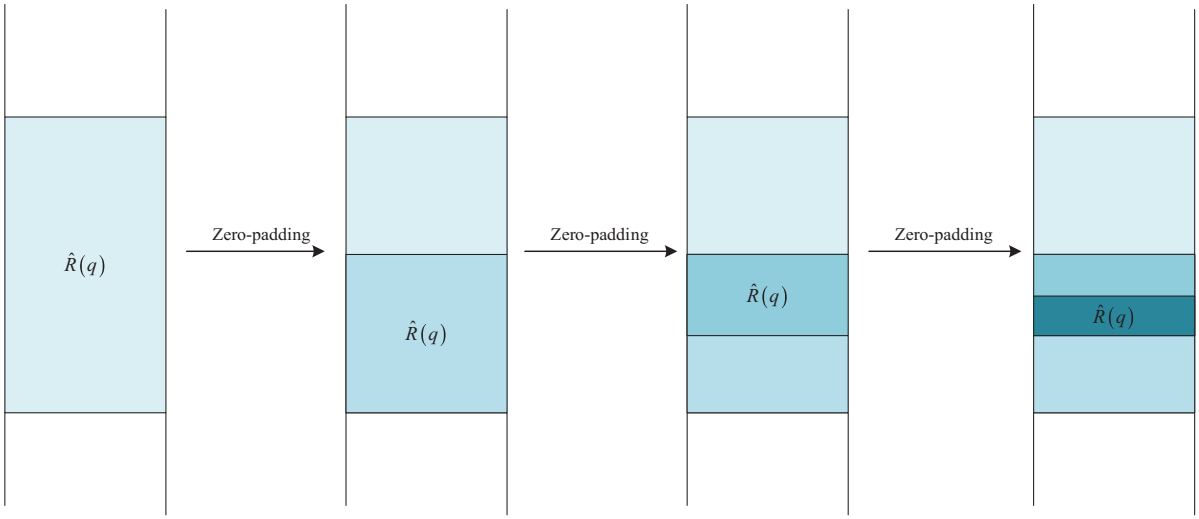

Fig.4. Schematic of the repeated zero-padding.

$$A(\widehat{R},q,\widehat{x}(q)) = aT \times \exp\left(j4\pi f_c \frac{R-\widehat{R}}{c}\right) \times \mathrm{sinc}\left(2\pi KT \frac{R+x(q)-\widehat{R}-\widehat{x}(q)}{c}\right) \quad (11)$$

If $\widehat{x}(q)$ and $x(q)$ have the same trend, then $|x(q)-\widehat{x}(q)| < |x(q)|$. It implies that the target displacement represented by signal $u(t,q)$ decreases. According to equation (5), smaller $|x(q)-\widehat{x}(q)|$ results in less time-varying characteristic in $A(\widehat{R},q)$, leading to higher accuracy in displacement $x(q)$. If $x(q)$ is approximated to $\widehat{x}(q)$, the time-varying characteristic in $A(\widehat{R},q)$ could be neglected. As shown in Fig. 3, after the baseband signal is multiplied by the operator constructed from the reference displacement, the target in the new signal no longer varies along the range dimension, hence the is regarded as displacement alignment. We could directly extract the residual displacement $x(q)-\widehat{x}(q)$ from $u(\widehat{R},q)$. Then, add $x(q)-\widehat{x}(q)$ and the reference displacement $\widehat{x}(q)$ to obtain the accurate displacement measurement result.

To ensure the reference displacement $\widehat{x}(q)$ closely approximates the actual displacement $x(q)$, zero-padding is applied in the Fourier transform. This enhances the resolution of Fourier transform, enabling localization accuracy far exceeding c/2KT. As shown in Fig. 4, each time a zero-padding expands the original signal to double its size, the maximum resolution error becomes half of its original one. By repeating the zero-padding, the maximum resolution error becomes smaller until $\widehat{R}(q)$ is sufficient to approximate the displacement $x(q)$. The resulting time-varying range $\widehat{R}(q)$ serves as a reference displacement $\widehat{x}(q)$ for obtaining $x(q)$. Note that, as referenced in [33], the accuracy of $\widehat{R}(q)$ is lower than that of the displacement $x(q)$ derived from phase. Therefore, $\widehat{R}(q)$ should only be used as a reference, not as a substitute for $x(q)$.

B. HIGH-QUALITY SOUND RECOVERY

The surface displacement of a stationary sound source directly corresponds to audio signal. However, for a moving sound source, the situation is more complex. Primarily, the motion of the target introduces frequency interference in the propagating electromagnetic waves, leading to the Doppler effect, which is a significant source of disturbance. Additionally, the surface displacement of a moving sound source consists of two components: a large amplitude macro-motion and relatively weak sound-induced vibrations, the latter of which are often masked by the former. Therefore, post-processing is crucial for extracting clear audio information.

To isolate the low-frequency macro-motion from the composite displacement signal, we employed a FIR high-pass filter, as illustrated in Fig. 3. The overall workflow of the proposed algorithm is presented in Algorithm I. For comparison, the conventional workflow of recovering sound is shown in Algorithm II. In summary, the displacement alignment method enables precise capture of the

Algorithm I. Proposed high-quality sound recovery method

1. Initialize: $g = 0, \varepsilon$
2. Obtain baseband signal $s_B(t,q)$
3. Apply Fourier Transform, get $s_B(R,q)$
4. Peak retrieval to obtain the $\widehat{R}^g(q)$ that maximizes $s_B(\widehat{R},q)$
5. While $|\widehat{R}^g(q) - \widehat{R}^{g-1}(q)| < \varepsilon$
6. Zero-padding for $s_B(t,q)$, then apply Fourier Transform to get finer $s_B(R,q)$
7. Peak retrieval to obtain the $\widehat{R}^{g+1}(q)$ that maximizes $s_B(\widehat{R},q)$, $\widehat{x}(q) = \widehat{R}^{g+1}(q)$
8. $g = g + 1$
9. Get demodulated signal $u(t,q) = s_B(t,q)\Phi(t,\widehat{x}(q))$
10. Apply Fourier Transform, get $u(R,q)$
11. Peak retrieval to obtain the $\widehat{R}$ that maximizes $u(\widehat{R},q)$
12. Cross-multiplying $u(\widehat{R},q)$ and $u(\widehat{R},q-1)$, get displacement $x(q)$
13. High-pass filtering to recover the sound signal

Algorithm II. Conventional sound recovery method

1. Obtain baseband signal $s_B(t,q)$
2. Apply Fourier Transform, get $s_B(R,q)$
3. Peak retrieval to obtain the $\widehat{R}(q)$ that maximizes $s_B(\widehat{R},q)$
4. Cross-multiplying $s_B(\widehat{R}(q),q)$ and $s_B(\widehat{R}(q),q-1)$, get displacement $x(q)$
5. High-pass filtering to recover the sound signal

sound source's surface vibration, while the high-pass filter effectively separates low-frequency motion interference, playing an essential role in the audio recovery process of moving targets.

C. DOPPLER EFFECT

While using electromagnetic waves to detect a moving sound source, the frequency of the signal received by the radar is different from the ideal frequency, which is known as Doppler effect. In this subsection, a discussion is presented to explore the impact of Doppler effect on sound recovery.

Suppose the radar emits a single-frequency signal with phase $\Phi_1(t)$

$$\Phi_1(t) = 2\pi f_1 t \tag{12}$$

where f_1 denotes frequency of the transmitted signal. The received signal phase $\Phi_2(t) = 2\pi f_1 \widehat{t}$ is supposed to be emitted by radar at time $\widehat{t}$, then $c(t-\widehat{t}) = r(q)$ is satisfied. If both radar and target maintain constant velocity v_1 and v_2, then

$$\widehat{t} = \frac{(c-v_2)t - r(q)}{c - v_1} \tag{13}$$

Substitute $\widehat{t}$ into received signal phase $\Phi_2(t)$, then

$$\begin{aligned}\Phi_2(t) &= f_1 \cdot \frac{c-v_2}{c-v_1} t - f_1 \cdot \frac{r(q)}{c-v_1} \\ &= f_1 \cdot \underbrace{\frac{c-v_2}{c-v_1}}_{frequency\ shift} \left(t - \underbrace{\frac{r(q)}{c}}_{time\ delay} \cdot \underbrace{\frac{c}{c-v_2}}_{time\ delay\ scaling} \right)\end{aligned} \tag{14}$$

The equation specifically notes the frequency shift and time delay scaling caused by the Doppler effect. Among them, time delay scaling means that the measured target position $r(q)$ and corresponding displacement $x(q)$ will be magnified to $\frac{c}{c-v_2}$ times the true ones. In the scenario discussed in this study, the target's movement is slow. Taking 1 m/s as an example, the total displacement scale of the target has to reach 1000 m to generate a maximum measurement error of 1 μm. Therefore, the scaling can be ignored. On the other hand, since the displacement measurement method proposed in this manuscript is entirely based on phase estimation, the impact of frequency offset on the results can also be ignored.

IV. EXPERIMENTAL VALIDATIONS

In this section, we apply the proposed method to analyze signals captured from a moving speaker, demonstrating its potential application. The experimental setup is shown in Fig. 5, involves a speaker mounted on a sliding platform facing the radar. When the speaker is excited, its surface vibrates, and these vibrations are captured by the radar as displacement. The radar operates within a frequency range of 77 GHz to 80 GHz, with a sampling interval of 0.5 ms, resulting in a sampling rate of 2000 Hz for the speaker's surface vibrations. According to the Nyquist sampling theorem, this setup is sufficient to capture audio frequencies below 1000 Hz.

During the experiment, the speaker played a piano audio recording while the platform moved away from the radar at a speed of 1 mm/s. The input audio signal and its

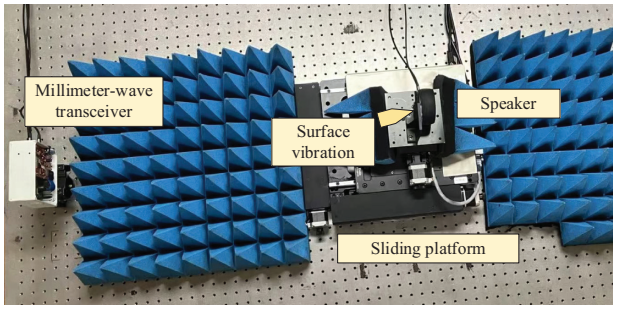

Fig. 5. Setup of the experiment.

corresponding speaker surface vibration are shown in Fig. 6. The input audio amplitude, representing relative volume, is normalized, while the speaker's surface vibration amplitude is around 10 to 20 mm. These vibrations can be effectively captured by the radar. To compare the performance of the proposed method, sound recovered by conventional method (Algorithm II) were also demonstrated, as shown in Fig. 6(c) and 6(f).

The corresponding time-frequency spectrum are shown on the right of Fig. 6. Compared to the time-frequency spectrum of the input signal, the spectrum obtained through the proposed method exhibits some degradation in the high-frequency range, yet it still captures the original trend. Additionally, vibration obtained via conventional method show significant jitter in specific intervals (5–6s, 10–12s), which obscures the corresponding region in time-frequency spectrum. To highlight the details, we provide zoomed-in spectra for the 50–300 Hz (Fig. 7) and 300–500 Hz (Fig. 8). In the 50–300 Hz range, the speaker's surface vibrations closely match the input audio signal. However, as the frequency approaches 500 Hz, the signal begins to degrade. Beyond 500 Hz, the spectrum becomes indistinguishable, as shown in Fig. 6. To investigate this performance gap, a frequency sweep experiment was conducted ranging from 0 to 1000Hz. The experimental results are shown in Fig. 9.

The amplitude of the input signal remains consistent across each frequency band, however, the vibration amplitude of the speaker surface is not uniform across all bands. Specifically, the vibration amplitude is approximately 0.01 mm in the extremely low-frequency band (below 200 Hz), 0.015 mm in the low-frequency band (200–500 Hz), and attenuates to around 2 mm in the high-frequency band (exceed 500 Hz). This explains the variations in sound recovery by the radar across different frequency bands.

A multiple objective experiment was also conducted. Among them, two sliding platform were adopted. One slider moved with a non-uniform velocity ranging from 10 mm/s to 2 mm/s, while the other moved at 5 mm/s. Each slider was rigidly connected to a speaker, designated as speaker A and speaker B, respectively. The radar-detected surface vibrations of the speakers are depicted in Fig. 10.

The motion of speaker A, which includes multi-stage constant-speed movements and acceleration/deceleration processes, can be considered nonlinear. In contrast, speaker B operates at a speed (5 mm/s) five times faster than that in the single-target experiment (1 mm/s). Using the proposed method, both speakers can recover sound frequencies ranging from 0 to 1000 Hz, showing no significant difference

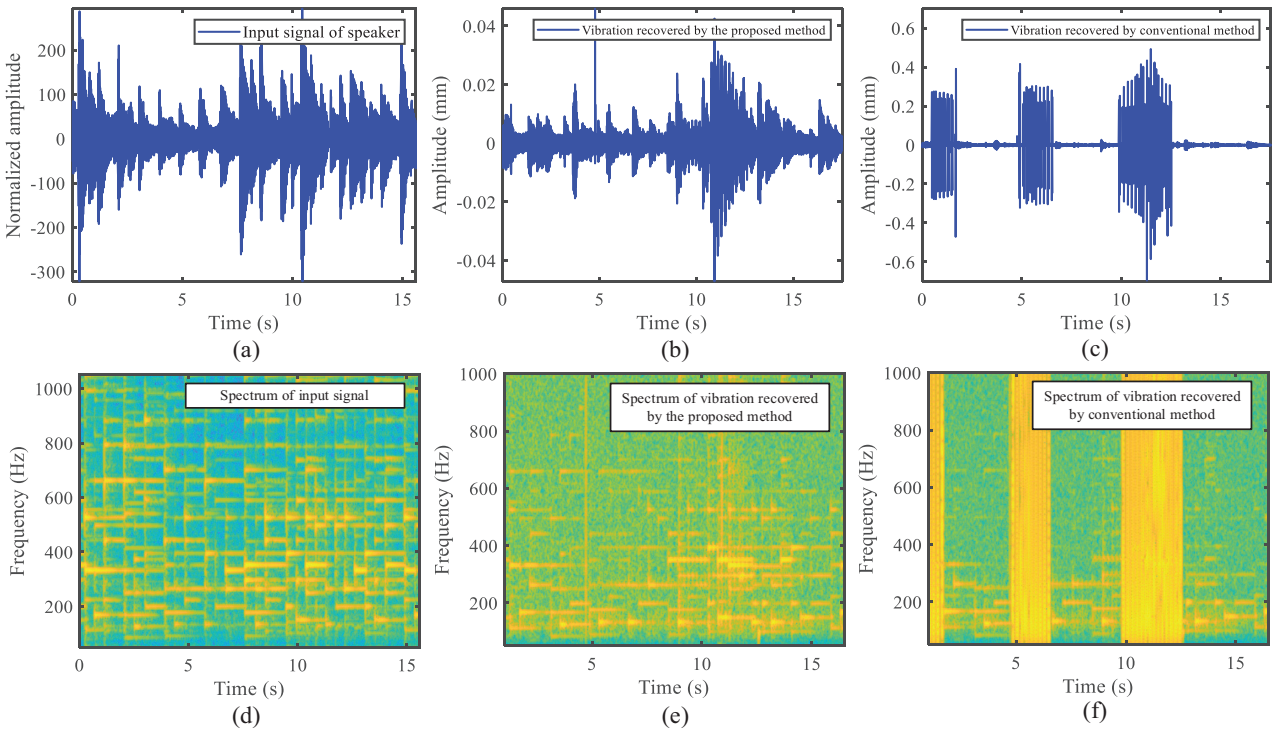

Fig. 6. Comparison between input signal and recovered sound from speaker. (a) Input signal of the speaker. (b) Vibration recovered by the proposed method. (c) Vibration recovered by conventional method. (d) Time-frequency spectrum of the input signal. (e) Time-frequency spectrum of vibration recovered by the proposed method. (f) Time-frequency spectrum of vibration recovered by conventional method.

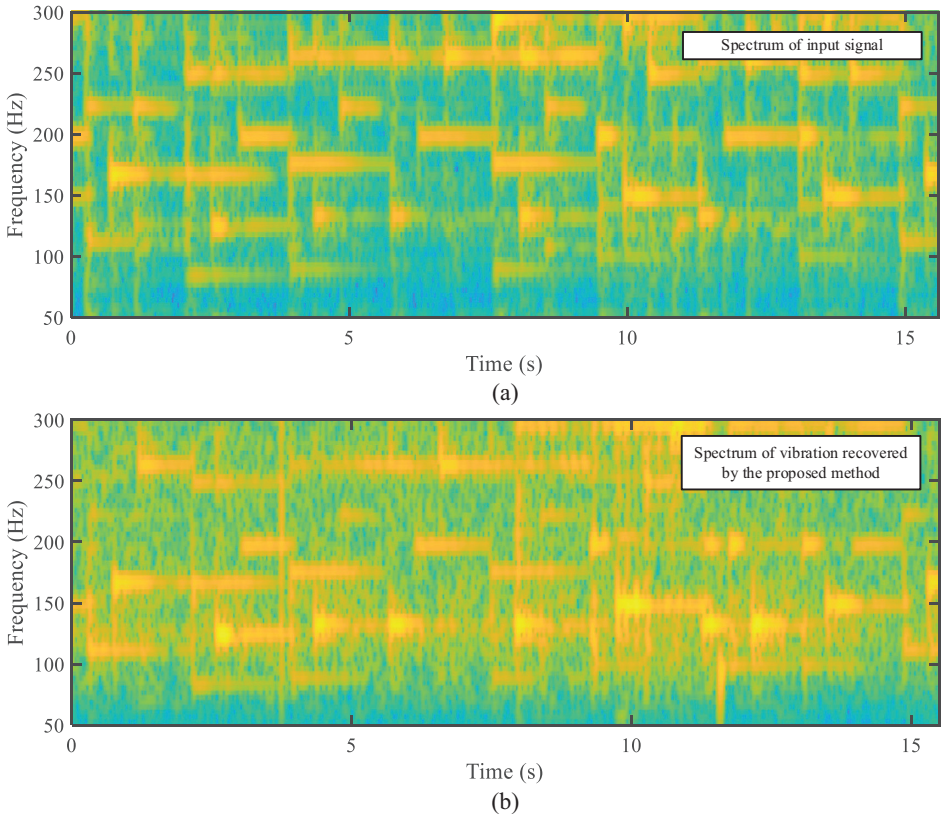

Fig. 7. Zoomed-in time-frequency spectrum for the 50–300 Hz. (a) Time-frequency spectrum of the input signal. (b) Time-frequency spectrum of the recovered sound.

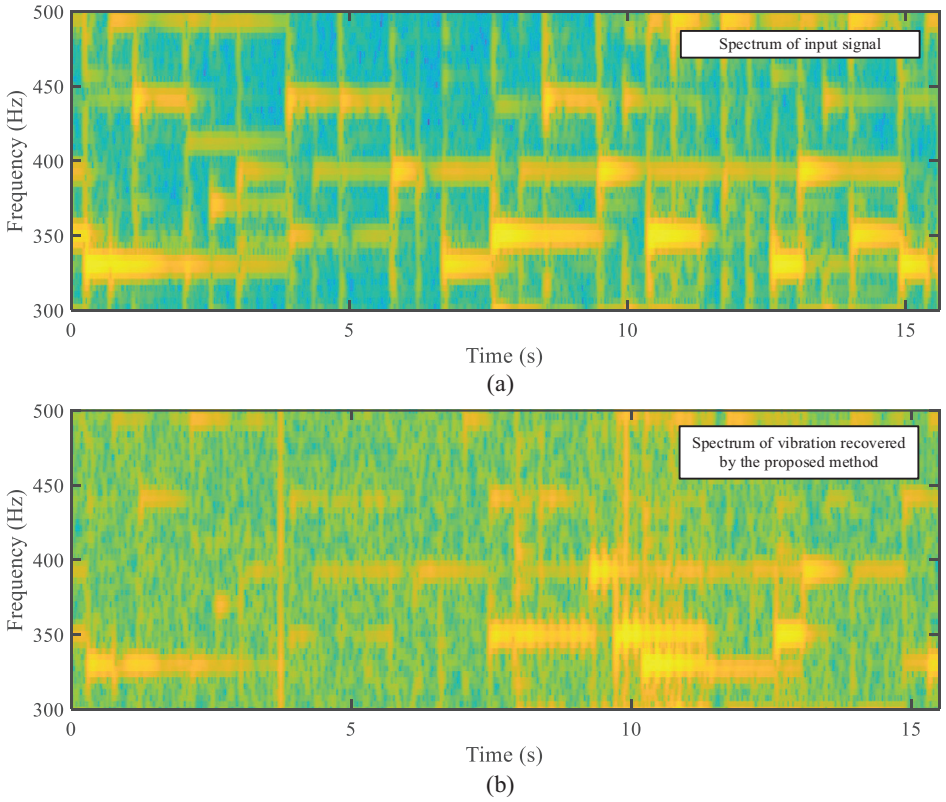

Fig. 8. Zoomed-in time-frequency spectrum for the 300–500 Hz. (a) Time-frequency spectrum of the input signal. (b) Time-frequency spectrum of the recovered sound.

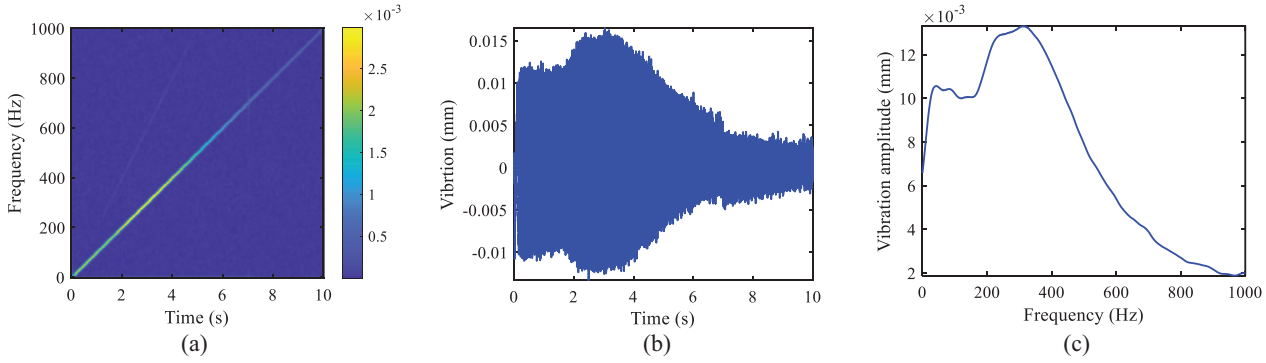

Fig. 9. (a) Time-frequency spectrum of the input signal. (b) Surface vibration detected from radar. (c) Surface frequency response of the applied speaker.

compared to the single-target experiment. It demonstrates that the proposed method is applicable to a range of velocities, multiple simultaneous sound sources, and even scenarios involving nonlinear movement patterns.

In summary, this experiment not only validates the effectiveness of the proposed method but also demonstrates its potential for processing signals from moving sound sources.

V. CONCLUSION

This paper introduces a novel displacement alignment method that effectively addresses the challenge of high-quality sound recovery from moving sources in millimeter-wave microphone sound detection. By aligning the displacement over time and utilizing FIR filtering to mitigate low-frequency interference, the proposed method demonstrates significant improvements in sound detection accuracy. Experimental results validate the effectiveness of this approach, offering a promising solution to enhance radar-based sound recovery, with potential applications in various fields requiring precise monitoring of dynamic sources.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 51905341 and the Natural Science Foundation of Shanghai under Grant 22ZR1433900.

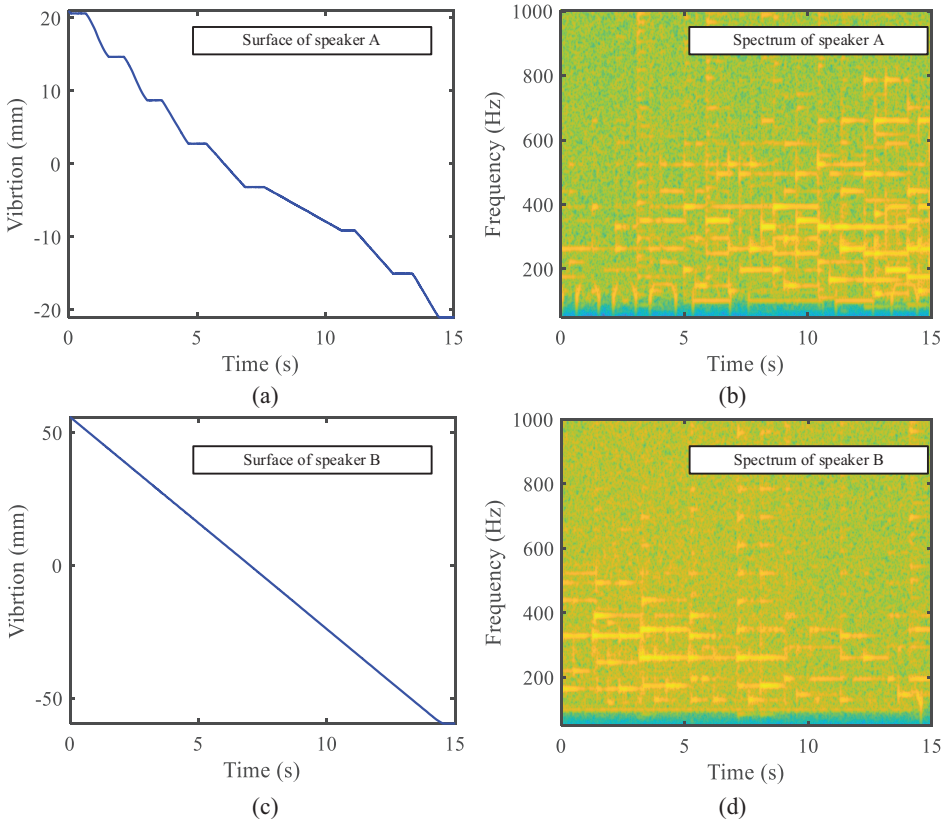

Fig. 10. Recovered sound from speaker A and speaker B. (a) surface displacement of speaker A. (b) Time-frequency spectrum of speaker A. (c) surface displacement of speaker A. (d) Time-frequency spectrum of speaker A.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

[2] Z. Xu et al., "Remote eavesdropping at 200 meters distance based on laser feedback interferometry with single-photon sensitivity," *Opt. Lasers Eng.*, vol. 141, p. 106562, 2021.

[3] J. Vanherzeele, S. Vanlanduit, and P. Guillaume, "Acoustic source identification using a scanning laser Doppler vibrometer," *Opt. Lasers Eng.*, vol. 45, no. 6, pp. 742–749, 2007.

[4] J. Ahn and D. Kim, "Simple and effective speech enhancement for visual microphone," in 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), Nanjing, China, Nov. 2017, pp. 694–699. doi: 10.1109/ACPR.2017.41.

[5] A. Davis et al., "The visual microphone: passive recovery of sound from video," *ACM Trans Graph*, vol. 33, no. 4, pp. 79:1–79:10, 2014.

[6] S. Li, Y. Xiong, P. Zhou, Z. Ren, and Z. Peng, "mmPhone: sound recovery using millimeter-wave radios with adaptive fusion enhanced vibration sensing," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 8, pp. 4045–4055, 2022.

[7] Y. Xiong, S. Li, C. Gu, G. Meng, and Z. Peng, "Millimeter-wave bat for mapping and quantifying micromotions in full field of view," *Research, vol.* 2021, pp. 1–13, 2021.

[8] J. Chen et al., "Multimodal fusion for indoor sound source localization," *Pattern Recognit.*, vol. 115, p. 107906, 2021.

[9] P. Chiariotti, M. Martarelli, and P. Castellini, "Acoustic beamforming for noise source localization – reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 120, pp. 422–448, 2019.

[10] J. A. Zhang et al., "Enabling joint communication and radar sensing in mobile networks—a survey," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 1, pp. 306–345, 2022.

[11] F. Liu et al., "Integrated sensing and communications: toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, 2022.

[12] B.-K. Park, O. Boric-Lubecke, and V. M. Lubecke, "Arctangent demodulation with DC offset compensation in quadrature doppler radar receiver systems," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 1073–1079, 2007.

[13] Y. Xiong, S. Chen, X. Dong, Z. Peng, and W. Zhang, "Accurate measurement in doppler radar vital sign detection based on parameterized demodulation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 11, pp. 4483–4492, 2017.

[14] L. Piotrowsky, T. Jaeschke, and N. Pohl, "Distance measurement using mmwave radar: micron accuracy at medium range," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 11, pp. 5259–5270, 2022.

[15] Y. Xiong, Z. Peng, G. Xing, W. Zhang, and G. Meng, "Accurate and robust displacement measurement for FMCW radar vibration monitoring," *IEEE Sens. J.*, vol. 18, no. 3, pp. 1131–1139, 2018.

[16] L. Piotrowsky, S. Kueppers, and N. Pohl, "Enabling high accuracy distance measurements with FMCW radar sensors,"

IEEE Trans. Microw. Theory Tech., vol. 67, no. 12, pp. 5360–5371, 2019.

[17] K. Han and S. Hong, "Vocal signal detection and speaking-human localization with MIMO FMCW radar," *IEEE Trans. Microw. Theory Tech.*, vol. 69, no. 11, pp. 4791–4802, 2021.

[18] Y. Xiong, Z. Liu, S. Li, G. Meng, and Z. Peng, "Scanning microwave vibrometer: full-field vibration measurement via microwave sensing with phase-encoded beam scanning," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[19] X. Shen, Y. Xiong, S. Li, and Z. Peng, "RFMic-phone: robust sound acquisition combining millimeter-wave radar and microphone," *IEEE Sens. Lett.*, vol. 6, no. 11, pp. 1–4, 2022.

[20] Y. Xiong, S. Chen, G. Xing, Z. Peng, and W. Zhang, "Static clutter elimination for frequency-modulated continuous-wave radar displacement measurement based on phasor offset compensation," *Electron. Lett.*, vol. 53, no. 22, pp. 1491–1493, 2017.

[21] Y. Xiong et al., "An effective accuracy evaluation method for LFMCW radar displacement monitoring with phasor statistical analysis," *IEEE Sens. J.*, vol. 19, no. 24, pp. 12224–12234, 2019.

[22] S. Scherr et al., "An efficient frequency and phase estimation algorithm with CRB performance for FMCW radar applications," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1868–1875, 2015.

[23] C. Li and J. Lin, "Complex signal demodulation and random body movement cancellation techniques for non-contact vital sign detection," In *IEEE MTT S Int Microwave Symp Dig*. Atlanta, GA, USA: IEEE, June 2008, pp. 567–570.

[24] W. Xu, Y. Li, C. Gu, and J.-F. Mao, "Large displacement motion interferometry with modified differentiate and cross-multiply technique," *IEEE Trans. Microw. Theory Tech.*, vol. 69, no. 11, pp. 4879–4890, 2021.

[25] X. Shen, Y. Xiong, S. Li, and Z. Peng, "Indoor human activity recognition using millimeter-wave radio signals," in *2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD)*, Harbin, China, Nov. 2022, pp. 1–6. doi: 10.1109/ICSMD57530.2022.10058412.

[26] J. Liu, J. Lu, Y. Li, C. Gu, and J. Mao, "Mitigation of leakage and stationary clutters in short-range FMCW radar with hybrid analog and digital compensation technique," *IEEE Trans. Microw. Theory Tech.*, vol. 70, no. 1, pp. 62–73, 2022.

[27] H. Qiao, S. Shahsavari, and P. Pal, "Super-resolution with noisy measurements: reconciling upper and lower bounds," in *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, pp. 9304–9308. doi: 10.1109/ICASSP40776.2020.9054159.

[28] Z. Liu, Y. Xiong, G. Wu, G. Meng, and Z. Peng, "Super-resolution and accurate full-field displacement measurement with millimeter-wave radars," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–11, 2023.

[29] F.-K. Wang et al., "Review of self-injection-locked radar systems for noncontact detection of vital Signs," *IEEE J. Electromagn. RF Microw. Med. Biol.*, vol. 4, no. 4, pp. 294–307, 2020.

[30] F.-K. Wang et al., "Single-antenna doppler radars using self and mutual injection locking for vital sign detection with random body movement cancellation," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 12, pp. 3577–3587, 2011.

[31] X. Li et al., "STGRFT for detection of maneuvering weak target with multiple motion models," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1902–1917, 2019.

[32] Z. Liu, Y. Xiong, W. Tian, G. Meng, and Z. Peng, "Displacement measurement of weak targets with imaging radar," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.

[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Hoboken, NJ, USA: Prentice-Hall, Inc., 1993.