

ActiveVital: Geometry-Aware Embodied Vital Signs Monitoring for Home Healthcare Robots

Yuxuan Hu^{1,2}, Shihao Li¹, Yang Xiao¹, Gen Li¹, Feng Xu², and Jianfei Yang^{1*}

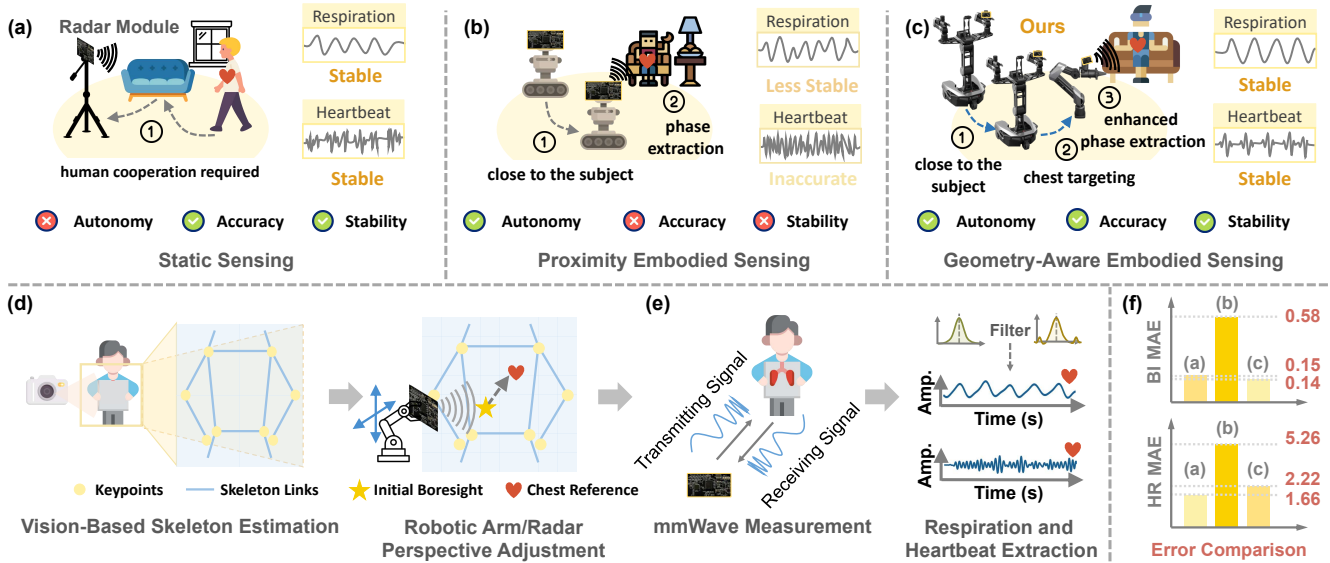


Fig. 1: **Geometry-Aware Embodied mmWave Vital Signs Sensing (ActiveVital).** (a) Static Sensing relies on user alignment to maintain favorable observation geometry. (b) Proximity Embodied Sensing improves signal strength by reducing distance but does not regulate anatomical alignment. (c) Geometry-Aware Embodied Sensing actively controls sensor pose to achieve chest-centric geometric alignment and stable measurement. (d) Vision-based chest localization and geometric reference estimation. (e) Closed-loop radar alignment and mmWave-based respiration and heartbeat extraction. (f) Quantitative error comparison across sensing paradigms, demonstrating the benefit of active geometry regulation.

Abstract—Home robots require reliable vital signs monitoring to support long-term companionship and safety in daily environments, yet obtaining respiration and heart rate without physical contact remains challenging in unconstrained home settings. Millimeter-wave (mmWave) radar offers a promising solution due to its phase sensitivity to sub-millimeter motions. However, mmWave measurements are fundamentally constrained by observation geometry, since only the radial component of motion is observable. Consequently, arbitrary robot-human orientations often introduce angular misalignment that destabilizes vital signs estimation. To address this limitation, we reformulate vital signs monitoring from passive signal recovery to active geometric regulation. We propose ActiveVital, a vision-guided sensing framework that treats sensing geometry as an explicit control variable for robots. It localizes the chest anchor via visual keypoints and converts alignment errors into control commands. This steers the robot-mounted radar toward near-normal incidence to the thoracic surface, maximizing radial observability within a perception-action loop. A differential phase enhancement module further stabilizes signal

extraction under motion. Experiments show that ActiveVital reduces respiration interval error from 0.87 s to 0.14 s and heart rate error from 13.59 bpm to 2.22 bpm, achieving accuracy comparable to controlled static sensing while remaining robust under unconstrained robot-human configurations. Project page: <https://yuxuanhu9.github.io/ActiveVital/>

I. INTRODUCTION

Home robots are increasingly deployed for long-term in-home service and companionship. Beyond task execution, these robots are expected to support user well-being through health monitoring during daily interaction. Reliable respiration and heart rate monitoring is therefore critical for ensuring user safety and enabling health-aware interaction. However, obtaining these vital signs without physical contact remains challenging in unconstrained home environments. Millimeter-wave (mmWave) radar has been proven to be a promising solution, as it captures thoracic micro-motion contactlessly through phase variations [1], [2]. Beyond vital signs, wireless sensing has also advanced broader human perception tasks such as pose estimation, mesh reconstruction, and action recognition [3], [4], [5].

This work is jointly supported by MOE Singapore Tier 1 Grant RG83/25, RS36/24 and a Start-up Grant from Nanyang Technological University.

¹MARS Lab, Nanyang Technological University, Singapore

²Fudan University, Shanghai, China

*Corresponding author: Jianfei Yang, jianfei.yang@ntu.edu.sg.

Nevertheless, mmWave radar measures displacement only along the line of sight, meaning that only the radial component of respiration and heartbeat motion is observable. Consequently, geometric misalignment directly attenuates the measurable signal, particularly for heartbeat components with extremely small amplitudes. This geometry-induced observability constraint indicates that measurement quality is governed primarily by sensor–human relative pose rather than signal processing alone.

Most existing systems adopt a static sensing paradigm [6], as illustrated in Fig. 1(a), requiring users to remain aligned with the radar to maintain favorable observation geometry. Some studies deploy embodied sensing on mobile platforms [7], [8], as shown in Fig. 1(b), but they regulate only spatial proximity and improve measurement quality mainly through distance reduction. Consequently, current methods lack a closed-loop pose regulation mechanism that directly optimizes vital sign observability.

In domestic environments, where users may be sleeping, seated, or moving freely, pose and interaction behaviors are inherently uncertain, making sustained near-axial observation difficult. Consequently, observation geometry must be actively regulated by the robot rather than passively relying on user alignment. To address this challenge, we propose ActiveVital, a chest-centric embodied sensing framework (Fig. 1(c)). Visual keypoints are used to estimate a chest reference, enabling the robot to regulate radar pose and achieve near-axial alignment with the thoracic surface within a closed-loop control framework. To ensure stable signal extraction under residual pose variations, a position-robust phase enhancement module is integrated for reliable respiration and heartbeat separation. The overall perception-to-alignment pipeline and quantitative comparison are illustrated in Fig. 1(d)–(f).

The system is implemented on a mobile robot equipped with a 60 GHz frequency-modulated continuous-wave (FMCW) radar and evaluated under three sensing modes: Static Sensing, Proximity Embodied Sensing, and Geometry-Aware Embodied Sensing. Experimental results show that closed-loop chest alignment significantly improves estimation accuracy, enabling stable non-contact vital sign monitoring in real domestic environments.

The main contributions are summarized as follows:

- **Geometry-Regulated Embodied Sensing:** An embodied vital sign sensing paradigm that actively regulates radar observation geometry through feedback-driven pose alignment, enabling reliable sensing under unconstrained robot–human configurations.
- **Vision-Guided Pose Regulation with Robust Signal Extraction:** A vision-guided alignment pipeline steers the robot-mounted radar toward near-axial observation, while phase-based signal enhancement stabilizes respiration and heartbeat extraction under pose variations.
- **Cross-Paradigm and Cross-Baseline Validation:** Evaluation under three sensing paradigms (Static Sensing, Proximity Embodied Sensing, and Geometry-Regulated Embodied Sensing) with three baseline meth-

ods demonstrates consistent accuracy improvements over existing approaches.

II. RELATED WORK

This section reviews prior studies on non-contact vital signs monitoring from the perspective of perception–control coupling in robotic systems, highlighting the limitations of passive static observation and the need for geometry-aware embodied sensing as a controllable observation paradigm.

A. Static Vital Signs Monitoring

Non-contact vital signs monitoring has been explored using visible-light imaging, thermal sensing, acoustics, Wi-Fi, and radar [9], [10], [11], [12], [13]. Among these modalities, FMCW mmWave radar is widely adopted due to its sensitivity to thoracic micro-motion along the radar line of sight [14]. Most mmWave systems follow a static sensing paradigm in which the sensor is fixed and subjects are assumed to remain within favorable observation geometry. Under this assumption, research primarily concentrates on backend signal processing, including differential phase enhancement [15], model-based separation [16], and high-resolution spectral analysis [17]. These methods improve extraction performance when geometry is stable, but they treat observation geometry as externally imposed rather than actively controlled. Recent multimodal approaches incorporate vision or thermal sensing to assist localization and motion compensation [18], [19], [20], [21], [22]. However, these systems still rely on passively configured sensing axes. Prior studies further indicate that misalignment between the radar probing vector and the effective thoracic surface reduces the radial projection of cardiac-induced micro-motion [16], which degrades heartbeat-related phase observability.

B. Proximity Embodied Vital Signs Sensing

To alleviate fixed-view limitations, several studies embed mmWave radar into mobile robotic platforms, leveraging mobility to achieve spatial proximity and enhance echo amplitude [8], [23], [24]. However, existing approaches primarily emphasize distance reduction rather than structure-aware geometric alignment. The coupling between chest anatomy and radar observation geometry is rarely modeled explicitly, and observation pose is seldom treated as a controllable variable. Pose-aware systems that introduce directional adjustment mechanisms [7], [25] typically rely on heuristic rules without explicitly incorporating anatomical constraints. In summary, existing work either optimizes signal extraction under fixed geometry or exploits mobility for proximity without achieving explicit regulation of radar line-of-sight alignment with the thoracic surface. A sensing paradigm that integrates anatomical priors with mmWave geometric constraints and treats observation geometry as a controllable variable remains largely unexplored. To address this limitation, we propose ActiveVital, which actively constructs near-axial alignment through vision-guided embodied alignment and combines it with a geometry-robust phase refinement module for stable respiration and heartbeat estimation under posture variation.

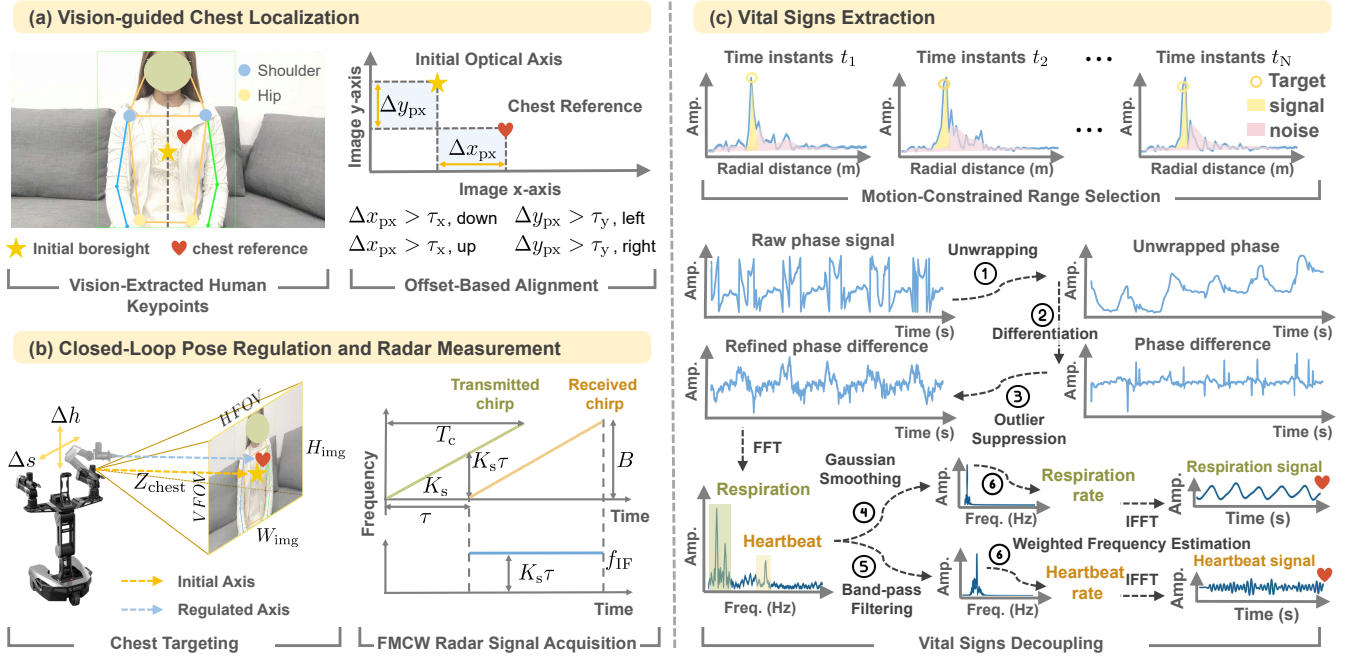


Fig. 2: **Processing pipeline of the proposed Geometry-Aware Embodied Sensing framework.** (a) Vision-guided chest localization using human skeletal keypoints and image-space offset alignment to define a geometric reference. (b) Closed-loop pose regulation and FMCW radar signal acquisition, where the robotic arm actively adjusts sensor orientation to achieve chest-centric alignment. (c) Motion-constrained vital signs extraction, including range selection, phase refinement, spectral decoupling, and respiration/heartbeat rate estimation.

III. METHOD

ActiveVital resolves the geometry-induced observability constraint through a three-stage pipeline integrating vision-guided pose regulation and phase-domain enhancement (Fig. 2). First, visual keypoints define a chest-centered reference in the image domain. Second, this reference drives closed-loop pose control to align the radar line-of-sight with the thoracic surface normal. Third, differential phase enhancement stabilizes signal extraction and separates respiration and heartbeat components.

A. Chest Localization

As illustrated in Fig. 2(a), posture variations in non-cooperative scenarios can misalign the radar line-of-sight with the chest surface, reducing cardiopulmonary motion observability. To enable active pose regulation under such conditions, a chest-centered reference is defined in the image domain. We employ ViTPose [26] to detect four torso keypoints:

$$P_s, P_{rs}, P_{lh}, P_{rh} \in \mathbb{R}^2, \quad (1)$$

where each keypoint is written as $P_i = [P_{x,i}, P_{y,i}]^\top$ in an image coordinate system with the x -axis pointing rightward and the y -axis upward. The shoulder and hip midpoints are computed as

$$P_{sc} = \frac{P_s + P_{rs}}{2}, \quad P_{hc} = \frac{P_{lh} + P_{rh}}{2}. \quad (2)$$

The vertical coordinate of the chest reference is interpolated along the torso axis as

$$P_{y,ref} = \alpha P_{y,sc} + (1 - \alpha) P_{y,hc}, \quad (3)$$

where $\alpha \in (0, 1)$. We set $\alpha = 0.75$ to bias the sensing anchor toward the upper chest region. To compensate for lateral anatomical asymmetry, a horizontal offset proportional to shoulder width is introduced. Define

$$W_{shoulder} = |P_{x,ls} - P_{x,rs}|, \quad (4)$$

and the horizontal coordinate is defined as

$$P_{x,ref} = P_{x,sc} + \beta W_{shoulder}, \quad (5)$$

where $\beta = 0.1$ shifts the reference toward the cardiac-adjacent region. The final chest sensing anchor is given by

$$P_{ref} = [P_{x,ref}, P_{y,ref}]^\top. \quad (6)$$

B. Chest-Centric Observation Construction

1) *Closed-Loop Error-Driven Pose Regulation:* As illustrated in Fig. 2(b), the robot regulates its sensor pose so that the radar probing axis intersects the chest reference point P_{ref} . Pixel deviations are converted into motion commands within a closed-loop control scheme. Let the image dimensions be W_{img} and H_{img} . The image center is defined as

$$L_x = \frac{W_{img}}{2}, \quad L_y = \frac{H_{img}}{2}. \quad (7)$$

Let $P_{\text{ref}} = [x, y]^T$. The pixel error vector is

$$\mathbf{E} = [e_x, e_y]^T, \quad e_x = x - L_x, \quad e_y = y - L_y. \quad (8)$$

To suppress oscillatory motion caused by estimation noise, tolerance thresholds $\delta_x, \delta_y \in \mathbb{R}_+$ are introduced. When $|e_x| \leq \delta_x$ and $|e_y| \leq \delta_y$, no motion is executed. The control output is defined as $\mathbf{A} = [a_v, a_h]^T$, where

$$a_v = \begin{cases} 1, & e_y > \delta_y, \\ -1, & e_y < -\delta_y, \\ 0, & \text{otherwise,} \end{cases} \quad a_h = \begin{cases} 1, & e_x > \delta_x, \\ -1, & e_x < -\delta_x, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Here, $a_h > 0$ and $a_v > 0$ correspond to rightward and upward sensor motion in the robot coordinate frame, respectively. The binary control vector \mathbf{A} determines motion direction along the horizontal and vertical axes, while the motion magnitude is scaled according to the projected metric displacement derived from \mathbf{E} . Alignment is established before each measurement, after which the seated subject stays still over a short window so the radar holds a fixed chest-centric pose.

2) *Image-to-Robot Spatial Mapping*: Pixel-level motion commands are converted into metric displacements using depth information and camera intrinsics. Let D denote the stereo depth map. The target distance is estimated within a neighborhood $\mathcal{N}(P_{\text{ref}})$ as

$$D_{\text{target}} = \text{median}\{D(u, v) \mid (u, v) \in \mathcal{N}(P_{\text{ref}})\}. \quad (10)$$

Using the pinhole camera model with focal lengths f_x and f_y , the spatial compensation corresponding to \mathbf{E} is approximated by

$$\Delta H = a_v \frac{|e_y| D_{\text{target}}}{f_y}, \quad \Delta L = a_h \frac{|e_x| D_{\text{target}}}{f_x}. \quad (11)$$

Depth is regulated by $\Delta D = D_{\text{target}} - D_{\text{ref}}$, where D_{ref} is the desired sensing distance. The resulting command $\mathbf{u} = [\Delta L, \Delta H, \Delta D]$ drives radar-chest alignment.

C. Geometry-Constrained Phase Processing

1) *Phase Modeling*: Under chest-centric alignment, radar phase variation is governed by radial cardiopulmonary displacement, modeled as

$$R(t) = R_0 + \Delta R(t), \quad \Delta R(t) = d(t) \cos \theta, \quad (12)$$

where $d(t)$ is the true chest displacement and θ is the angle between the radar line-of-sight and the chest normal. Active alignment drives $\theta \rightarrow 0$, maximizing radial observability.

$$S_T(t) = A_T \exp(j(2\pi f_0 t + \pi K_s t^2)), \quad (13)$$

where f_0 is the carrier frequency and K_s is the chirp slope. The received signal reflected from the chest is

$$S_R(t) = A_R \exp(j[2\pi f_0(t - \tau) + \pi K_s(t - \tau)^2 + 2\pi f_d(t - \tau)]), \quad (14)$$

where $\tau(t) = 2R(t)/c$ and f_d capture range modulation and micro-Doppler induced by chest motion. After mixing and low-pass filtering, the intermediate-frequency (IF) signal is obtained, whose phase encodes radial displacement.

Under the Geometrical Theory of Diffraction (GTD) interpretation, the chest echo can be approximated by a finite set of dominant scattering centers. For the i -th scatterer with distance $R_i(t)$ and delay $\tau_i(t) = 2R_i(t)/c$, the corresponding IF phase term is approximated by

$$\phi_{\text{if}}^{(i)}(t) \approx 2\pi f_0 \tau_i + 2\pi K_s \tau_i t - 2\pi f_d^{(i)} t. \quad (15)$$

A range FFT is applied to obtain the range profile, and the chest bin j_{trgt} is selected by amplitude-based detection with temporal continuity constraint. The slow-time phase sequence is then constructed as

$$\phi(m) = \arg(S_{\text{IF}}(j_{\text{trgt}}, m)), \quad m = 1, \dots, M. \quad (16)$$

2) *Differential Phase Enhancement*: The raw phase sequence $\phi(m)$ is first unwrapped to obtain a continuous trajectory $\phi_{\text{unwrap}}(m)$.

To suppress low-frequency drift and enhance periodic micro-motion, first-order temporal differencing is performed:

$$\Delta\phi_{\text{unwrap}}(m) = \phi_{\text{unwrap}}(m) - \phi_{\text{unwrap}}(m-1). \quad (17)$$

To mitigate impulsive noise, a Hampel filter is applied within the window

$$\Omega_m = \{m-r, \dots, m+r\}. \quad (18)$$

Within this window, the local median and deviation are computed as

$$\text{med}(m) = \text{median}\{\Delta\phi_{\text{unwrap}}(i)\}, \quad (19)$$

$$\text{MAD}(m) = \text{median}|\Delta\phi_{\text{unwrap}}(i) - \text{med}(m)|. \quad (20)$$

The robust scale estimate is

$$\sigma(m) = \kappa \text{MAD}(m), \quad (21)$$

where $\kappa = 1.4826$.

Samples satisfying

$$|\Delta\phi_{\text{unwrap}}(m) - \text{med}(m)| > \tau\sigma(m) \quad (22)$$

are replaced by $\text{med}(m)$ to suppress outliers.

The refined phase sequence is denoted as $\tilde{\phi}(m)$.

3) *Respiration and Heartbeat Rate Estimation*: The magnitude spectrum of the refined phase sequence is computed as

$$P(f) = \left| \text{FFT}\{\tilde{\phi}(m)\} \right|. \quad (23)$$

Respiration and heartbeat components are separated via Gaussian smoothing and FIR band-pass filtering, yielding $P_{\text{resp}}(f)$ and $P_{\text{heart}}(f)$.

The heart rate is estimated via energy-weighted averaging within the cardiac band:

$$f_{\text{heart}} = \frac{\sum_k f_k P_{\text{heart}}(f_k)}{\sum_k P_{\text{heart}}(f_k)}. \quad (24)$$

The final heart rate is $\text{HR} = 60f_{\text{heart}}$.

For respiration, if K peaks are detected between t_1^{resp} and t_K^{resp} , the breathing rate is

$$\text{BR} = \frac{K-1}{t_K^{\text{resp}} - t_1^{\text{resp}}} \times 60. \quad (25)$$

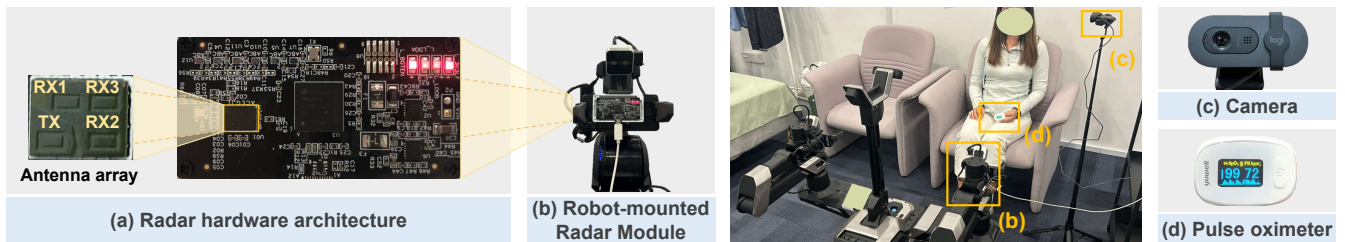


Fig. 3: **Experimental hardware configuration and scene setup.** (a) 60 GHz FMCW radar hardware, including antenna array and radar board. (b) Robot-mounted radar with onboard RGB-D for chest localization. (c) External RGB camera for the respiration-rate reference. (d) Fingertip pulse oximeter providing reference HR. The robot-mounted radar actively regulates its pose toward the seated participant for embodied vital signs sensing.

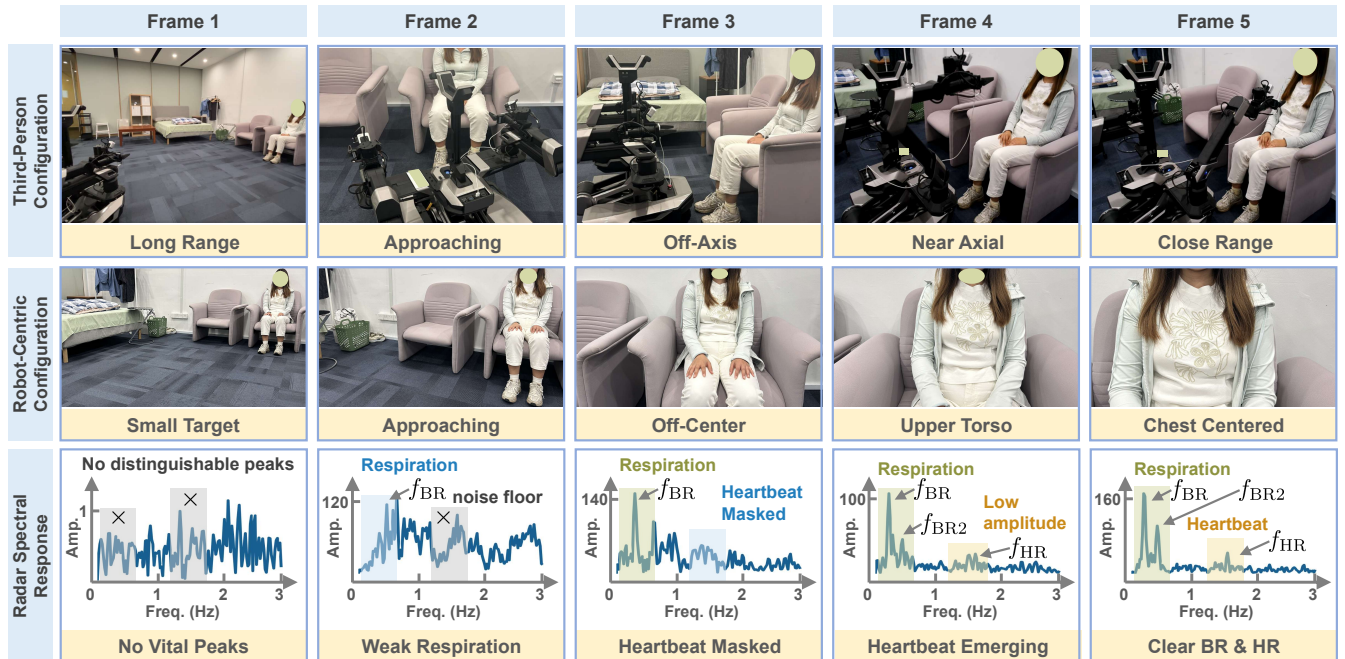


Fig. 4: **Progressive evolution of observation geometry and corresponding spectral responses.** Upper panels show representative physical configurations from long-range and off-axis conditions to chest-centered alignment. Lower panels present the associated phase-derived spectra, illustrating the transition from indistinguishable signals to clearly separable respiration and heartbeat components.

IV. EXPERIMENT

A. Experimental Setup

The proposed ActiveVital system is implemented on a Galaxea R1 Lite mobile robotic platform integrating millimeter-wave radar and RGB-D vision modules, as illustrated in Fig. 3. The radar module is rigidly mounted on the robot arm to enable active pose regulation, while an RGB camera provides visual keypoints for chest localization.

Specifically, the system incorporates a 60 GHz FMCW radar with a 1T3R antenna configuration (Fig. 3(a)(b)). The radar operates over 57.5–63.5 GHz with a chirp slope of 91.5 MHz/ μ s and a sampling rate of 2 MHz. Each chirp lasts 71.8 μ s with a frame period of 50 ms, yielding a range resolution of approximately 0.0228 m. Raw ADC data are streamed to a host computer for signal processing.

An onboard RGB-D camera (Fig. 3(b)) extracts human

skeletal keypoints for geometric reference construction. Pixel deviations between the radar optical axis and the chest reference are converted into metric displacements through the image-to-robot mapping described in Section III, enabling closed-loop regulation of the sensing pose.

All experiments are conducted in a 6.6 m \times 6.3 m \times 3.5 m indoor environment under controlled lighting conditions. Participants remain seated in a natural posture without explicit alignment instructions. Heart rate ground truth is synchronously recorded using a fingertip pulse oximeter (Fig. 3(d)), while respiration rate is derived from chest motion measured by the RGB camera (Fig. 3(c)) as an external reference for evaluation.

At the beginning of each trial, the robot is positioned approximately 1 m in front of the subject, with the chest located within the field of view of the arm-mounted camera.

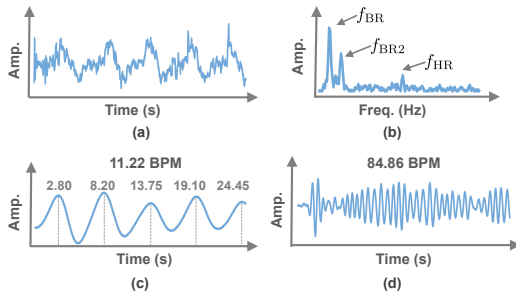


Fig. 5: Radar-measured phase signal representations under ActiveVital. (a) Slow-time unwrapped phase after phase refinement. (b) Corresponding amplitude spectrum showing distinct respiration fundamental (f_{BR}), respiration harmonic (f_{BR2}), and heartbeat (f_{HR}) components. (c) Extracted respiration waveform and estimated breathing rate. (d) Extracted heartbeat waveform and estimated heart rate.

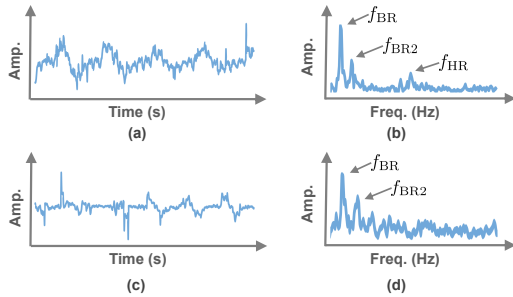


Fig. 6: Radar-measured phase signal representations under Static and Proximity Embodied Sensing. (a)(c) Slow-time phase sequences. (b)(d) Corresponding amplitude spectra. The first row corresponds to Static Sensing, and the second to Proximity Embodied Sensing.

Chest localization is then performed using ViTPose [26], and chest-centric alignment is achieved within a few seconds.

To illustrate the impact of observation geometry, Fig. 4 presents representative configurations from long-range and off-axis conditions to chest-centered alignment. The corresponding spectral responses demonstrate the progressive emergence of respiration and heartbeat components as geometric alignment improves.

B. Evaluation Across Sensing Paradigms

1) *Geometry-Dependent Observability*: We compare three sensing paradigms under identical hardware configuration and signal processing pipeline: Static Sensing, Proximity Embodied Sensing, and the proposed Geometry-Aware Embodied Sensing. They differ only in geometry. In Static sensing, the radar is fixed and the seated subject faces it while manually keeping near-axial alignment. In Proximity sensing, the robot uses the same keypoints to approach and shorten the distance but does not align the radar to the chest [7], [8]. In the Geometry-aware setting (ActiveVital), the robot aligns the radar with the chest to reach near-normal incidence. Fig. 5 shows the refined slow-time phase sequence measured from the radar signal and its corresponding spectrum under

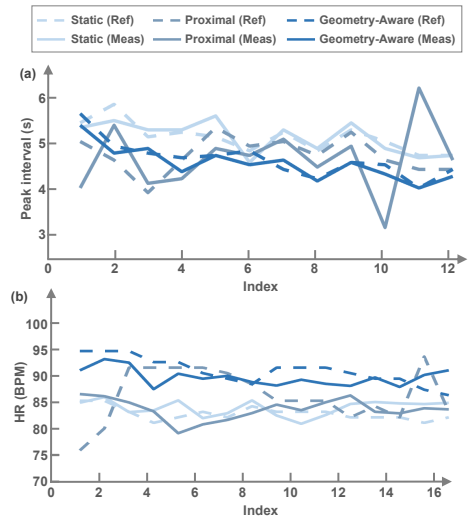


Fig. 7: Estimation results under three sensing paradigms using the same signal processing method. (a) BI estimation across 12 peak intervals under Static Sensing, Proximity Embodied Sensing, and ActiveVital, compared with reference measurements. (b) HR estimation across 16 sliding-window segments under the same paradigms, with corresponding reference values.

ActiveVital. The respiration fundamental, harmonic, and heartbeat peaks are clearly separated, enabling stable vital signs extraction. Fig. 6 presents the corresponding results under Static and Proximity Embodied Sensing. Static Sensing preserves distinguishable respiration and heartbeat peaks because subjects are deliberately positioned to maintain near-axial alignment with the radar line of sight. In contrast, under Proximity Embodied Sensing, misalignment reduces the radial projection of cardiac micro-motion. The attenuated heartbeat component falls close to the noise floor, making the spectral peak indistinguishable from background noise and leading to unstable estimation. These results indicate that deviation from optimal geometric alignment reduces the radial projection amplitude of cardiac micro-motion, thereby degrading its spectral observability.

2) *Geometry-Dependent Estimation Accuracy*: Measurement results of breath interval (BI) and heart rate (HR) under the three sensing paradigms are shown in Fig. 7, where estimated values are compared against reference measurements. Quantitative accuracy is further evaluated using mean absolute error (MAE) and root mean square error (RMSE), as summarized in Table I. BI is computed from 75 s of data, yielding 12 peak-to-peak intervals. HR is estimated from 50 s of data using a sliding window (20 s initialization, 2 s update), resulting in 16 HR estimates. For BI, Static and Geometry-aware Sensing achieve comparable accuracy (MAE: 0.15 s vs. 0.14 s; RMSE: 0.19 s vs. 0.17 s), while Proximity Sensing exhibits significantly larger errors (MAE: 0.58 s; RMSE: 0.78 s). For HR, Static yields the lowest error (MAE: 1.66 bpm; RMSE: 1.96 bpm). Geometry-aware Sensing achieves comparable performance (MAE: 2.22 bpm;

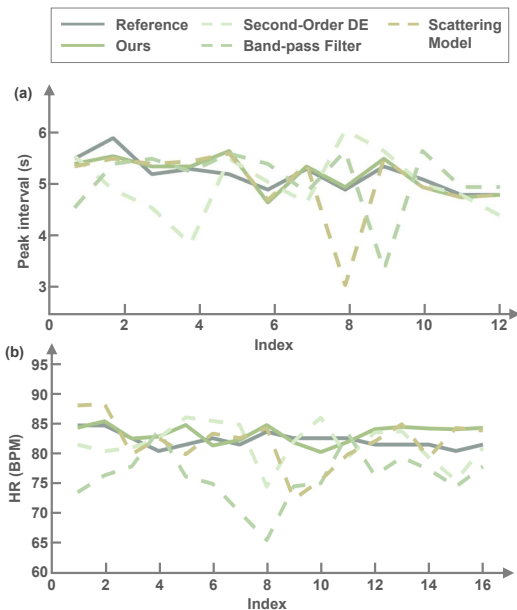


Fig. 8: Algorithmic performance under Static Sensing. (a) Breath interval estimation across 12 peak intervals. (b) Heart rate estimation across 16 time segments. The proposed sequential phase refinement method is compared with three baseline approaches against reference measurements.

TABLE I: Quantitative comparison of BI and HR under different sensing paradigms using the proposed method.

Paradigm	BI (s)		HR (bpm)	
	MAE	RMSE	MAE	RMSE
Static	0.15	0.19	1.66	1.96
Proximity	0.58	0.78	5.26	6.59
ActiveVital	0.14	0.17	2.22	2.59

RMSE: 2.59 bpm), substantially outperforming Proximity Sensing (MAE: 5.26 bpm; RMSE: 6.59 bpm).

Although Static Sensing achieves the lowest HR error under manually enforced near-axial alignment, this configuration requires explicit subject cooperation to maintain favorable geometry. In contrast, ActiveVital autonomously regulates the radar–thorax alignment, constructing comparable observation geometry without requiring posture constraints. These results demonstrate that preserving near-axial alignment maintains the radial projection amplitude of cardiac micro-motion, thereby stabilizing heartbeat estimation under realistic posture variation.

C. Algorithm Performance Across Sensing Paradigms

To analyze how different signal processing algorithms perform under varying sensing paradigms, we compare four methods across Static Sensing, Proximity Embodied Sensing, and ActiveVital. While the previous subsection examined the effect of geometry on observability and estimation accuracy, this subsection investigates the interaction between sensing geometry and algorithm design. Fig. 8 presents temporal comparisons under *Static Sensing*. Aggregated MAE and

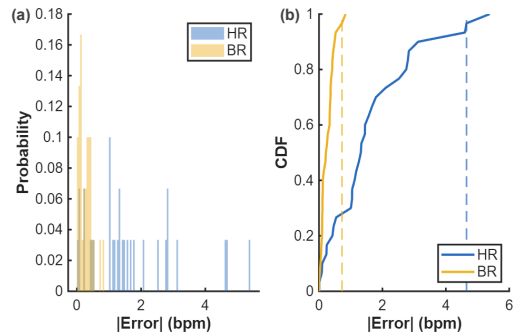


Fig. 9: Distribution of absolute estimation errors for HR and BR. (a) Histogram of absolute errors. (b) Cumulative distribution function (CDF). Dashed vertical lines indicate the 95th percentile error thresholds.

RMSE results across all three sensing paradigms are summarized in Table II.

Under *Static Sensing*, the proposed ActiveVital method achieves the lowest MAE for both BI (0.15 s) and HR (1.65 bpm), outperforming all baseline methods. When switching to *Proximity Embodied Sensing*, errors increase for all algorithms due to degraded observation geometry. However, the proposed method exhibits smaller performance degradation, particularly in HR estimation. Under *Geometry-aware*, BI estimation remains competitive across methods, while the proposed approach achieves the lowest HR error (2.22 bpm), indicating stronger preservation of weak cardiac-induced phase components.

Across sensing paradigms, all methods capture overall rhythmic trends. However, baseline approaches exhibit more pronounced error amplification as geometric conditions deteriorate. For example, the HR MAE of Baseline2 increases from 6.29 bpm under *Static* to 13.59 bpm under *Proximity*, whereas the proposed method increases from 1.65 bpm to 5.26 bpm. This cross-paradigm comparison demonstrates that the proposed algorithm maintains superior stability under geometry-induced observability variation.

D. Statistical Evaluation of BR and HR Estimation

To assess consistency, we analyzed 30 independent 25 s recordings, each yielding one breathing rate (BR) and HR estimate against the reference. Fig. 9 shows the absolute-error distribution (histogram and CDF) and Table III summarizes MAE, RMSE, and P95. BR errors are tightly concentrated (MAE 0.26 bpm, P95 0.72 bpm), while HR shows broader but bounded dispersion (MAE 1.67 bpm, P95 4.65 bpm) due to the weaker, lower-SNR cardiac micro-motion.

V. CONCLUSION

This work introduced ActiveVital, an embodied sensing framework that regulates radar–chest alignment through vision-guided pose control and phase-domain enhancement to maximize radial cardiopulmonary observability. Experiments across three sensing paradigms show that cardiac observability is geometry-dependent, and ActiveVital attains

TABLE II: Cross-algorithm comparison under three sensing paradigms (MAE and RMSE for BI and HR).

Algorithm	Static				Proximal				Geometry-aware			
	BI (s)		HR (bpm)		BI (s)		HR (bpm)		BI (s)		HR (bpm)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Second-Order DE [15]	0.51	0.68	2.93	3.44	0.65	0.75	9.27	12.12	0.93	1.40	7.60	9.48
Band-pass Filter [8], [17]	0.56	0.75	6.29	7.36	0.87	1.18	13.59	15.07	0.30	0.37	14.48	15.88
Scattering Model [16]	0.31	0.57	2.81	3.60	0.60	0.89	11.82	13.67	0.12	0.16	6.78	8.55
ActiveVital (Ours)	0.15	0.19	1.65	1.96	0.58	0.78	5.26	6.59	0.14	0.17	2.22	2.59

TABLE III: Statistical error metrics for HR and BR estimation over 30 independent recordings (25 s each).

Metric	MAE (bpm)	RMSE (bpm)	P95 (bpm)
BR	0.26	0.33	0.72
HR	1.67	2.16	4.65

accuracy comparable to static sensing without active subject cooperation while clearly surpassing proximity-based sensing, with the largest gains in heart rate where cardiac signals are weak. These results establish sensing pose as a controllable variable for robust non-contact physiological monitoring. As an opportunistic function of a multipurpose home robot, ActiveVital provides contactless monitoring at a comfortable standoff without wearables or pre-installed infrastructure.

REFERENCES

- [1] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electronics*, vol. 2, no. 6, p. 252–262, June 2019.
- [2] T. Yuan, Y. Kuo, A. Ren, R. Huo, Q. Sun, and A. Zahid, "State-of-the-art technologies and emerging trends in radar-based vital signs sensing: A review," *IEEE Sensors Journal*, vol. 26, no. 3, pp. 3446–3459, 2026.
- [3] J. Yang, H. Huang, Y. Zhou, X. Chen, Y. Xu, S. Yuan, H. Zou, C. X. Lu, and L. Xie, "Mm-fi: multi-modal non-intrusive 4d human dataset for versatile wireless sensing," in *Proceedings of the 37th International Conference on Neural Information Processing Systems*, ser. NIPS '23. Red Hook, NY, USA: Curran Associates Inc., 2023.
- [4] J. Fan, Y. Zhou, Y. Yang, X. Cui, J. Zhang, L. Xie, J. Yang, C. X. Lu, and F. Ding, "M4human: A large-scale multimodal mmwave radar benchmark for human mesh reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026, pp. 42 836–42 846.
- [5] X. Chen and J. Yang, "X-fi: A modality-invariant foundation model for multimodal human sensing," *arXiv preprint arXiv:2410.10167*, 2025.
- [6] H. Li, A. G. B. Ayuknsio, N. Yang, Z. Wei, M. Zhi, Q. Hua, and B. Zhao, "Radar-based vital signal extraction via morphological component analysis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 73, no. 11, pp. 9263–9278, 2025.
- [7] R. Cittadini, L. R. Buonocore, M. Di Castro, and L. Zollo, "Contactless respiration rate monitoring and human body pose detection for search and rescue robots," in *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, 2024, pp. 1119–1125.
- [8] C. A. Schroth, C. Eckrich, I. Kakouche, S. Fabian, O. von Stryk, A. M. Zoubir, and M. Muma, "Emergency response person localization and vital sign estimation using a semi-autonomous robot mounted sfw radar," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 6, pp. 1756–1769, 2024.
- [9] G. Scabbia, G. Da Poian, and W. Karlen, "Multispectral video fusion for non-contact monitoring of respiratory rate and apnea," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 350–359, 2021.
- [10] C. Hessler, M. Abouelenien, and M. Burzo, "A non-contact method for extracting heart and respiration rates," in *2020 17th Conference on Computer and Robot Vision (CRV)*, 2020, pp. 1–8.
- [11] A. Wang, D. Nguyen, A. R. Sridhar, and S. Gollakota, "Using smart speakers to contactlessly monitor heart rhythms," *Communications Biology*, vol. 4, no. 1, p. 319, Mar. 2021.
- [12] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, and J. Yang, "Monitoring vital signs and postures during sleep using wifi signals," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2071–2084, 2018.
- [13] I.-S. Lee and J.-R. Yang, "Signal preprocessing for heartbeat detection using continuous-wave doppler radar," *IEEE Microwave and Wireless Technology Letters*, vol. 33, no. 4, pp. 479–482, 2023.
- [14] V. P. Tran, A. A. Al-Jumaily, and S. M. S. Islam, "Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: A comprehensive review," *Big Data and Cognitive Computing*, vol. 3, no. 1, 2019.
- [15] Y. Xiong, Z. Peng, C. Gu, S. Li, D. Wang, and W. Zhang, "Differential enhancement method for robust and accurate heart rate monitoring via microwave vital sign sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 7108–7118, 2020.
- [16] Y. Hu, Z. Xia, and F. Xu, "Investigating respiration-heartbeat separation through a multipoint scattering chest wall motion model: 60-ghz fmcw radar assessment," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024.
- [17] S. F. Mirhosseini, M. Alae-Kerahroodi, A. E. Oik, U. Schroeder, and B. S. Mysore R, "In-car life detection and vital signs monitoring using mmwave radar sensors," *IEEE Transactions on Radar Systems*, vol. 4, pp. 207–231, 2026.
- [18] C. Yang, B. Bruce, X. Liu, B. Gholami, and N. Tavassolian, "A hybrid radar-camera respiratory monitoring system based on an impulse-radio ultrawideband radar," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 2646–2649.
- [19] D.-M. Chian, C.-K. Wen, C.-J. Wang, M.-H. Hsu, and F.-K. Wang, "Vital signs identification system with doppler radars and thermal camera," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 1, pp. 153–167, 2022.
- [20] A. Shokouhmand, S. Eckstrom, B. Gholami, and N. Tavassolian, "Camera-augmented non-contact vital sign monitoring in real time," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11 965–11 978, 2022.
- [21] Y. Wang, Z. Wang, J. A. Zhang, H. Zhang, and M. Xu, "Vital sign monitoring in dynamic environment via mmwave radar and camera fusion," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 4163–4180, 2024.
- [22] Y. Gou, Y. Xiong, W. Lin, S. Hong, and Z. Peng, "Vision-assisted chest wall localization for microwave vital sign monitoring," *IEEE Sensors Journal*, vol. 26, no. 2, pp. 2709–2723, 2026.
- [23] Y. Hu, K. Zuo, B. Ma, S. Li, Z. Xia, F. Xu, and J. Yang, "Wave-man: mmwave-based room-scale human interaction perception for humanoid robots," *arXiv preprint arXiv:2601.07454*, 2026.
- [24] Y. Lu, Y. Hu, S. Yuan, X. Zhou, K. Zuo, B. Lyu, X. Yuan, and J. Yang, "Em-fall: Embodied mmwave sensing for day-and-night fall detection on humanoid robots," *arXiv preprint arXiv:2606.11109*, 2026.
- [25] H.-W. Huang, J. Chen, P. R. Chai, C. Ehmke, P. Rupp, F. Z. Dadabhoj, A. Feng, C. Li, A. J. Thomas, M. da Silva, E. W. Boyer, and G. Traverso, "Mobile robotic platform for contactless vital sign monitoring," *Cyborg and Bionic Systems*, vol. 2022, p. 9780497, 2022.
- [26] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, "Vitpose: simple vision transformer baselines for human pose estimation," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, ser. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022.