I-Measure-Ur-♡! A Comprehensive Evaluation of Multi-Position Wearable SCG

Arina Borzistaia[1,2], Ruichao Cheng[1,3], and Denys Matthies[1,2]

[1] Technical University of Applied Sciences Lübeck, Lübeck, Germany
[2] Fraunhofer IMTE, Lübeck, Germany
[3] East China University of Science and Technology, Shanghai, China

Abstract. Seismocardiography (SCG) using inertial measurement units (IMUs) offers a promising alternative for continuous heart rate (HR) monitoring. However, its accuracy is highly dependent on both signal processing techniques and anatomical sensor placement. This study systematically evaluates the impact of filtering strategies, HR detection methods, and 20 distinct body positions to optimize IMU-based wearable cardiac monitoring. Tri-axial acceleration data were recorded at 200 Hz across 20 body locations on five healthy subjects, generating a dataset of 75,830 analysis windows. We cross-evaluated two filtering approaches (Direct Bandpass vs. Hilbert Envelope) and two detection methods (FFT vs. adaptive Peak detection). Our pipeline performance was evaluated against a synchronized PPG reference using statistical means. The Hilbert Envelope pipeline combined with adaptive peak detection achieved the highest accuracy, reducing the Mean Absolute Error by (MAE) by 14.69 BPM ($d = 0.66$) compared to standard bandpass filtering. Not surprisingly, the body position exerted a major effect on signal quality: trunk locations ($MAE = 3.13 BPM$) significantly outperformed limbs ($MAE = 7.21 BPM$), with the lower sternum being the most accurate ($MAE = 1.65 BPM$). While the standard wrist placement proved highly unreliable ($MAE = 9.73 BPM$), the triceps emerged as the best limb-based alternative ($MAE = 3.63 BPM$). In this paper, we evidenced that utilizing a Hilbert envelope with adaptive peak detection enables highly accurate SCG-based tracking. In terms of sensor placements, one should prioritize trunk or triceps positions, as conventional wrist-worn placements present severe limitations for reliable vital sign extraction.

Keywords: Seismocardiography (SCG) · Inertial Measurement Unit (IMU) · Signal Processing · Heart Rate Estimation · Sensor Placement.

1 Introduction

Cardiovascular disease remains the leading cause of death worldwide, accounting for approximately 32% of all global deaths [1, 2]. This underscores the need for continuous, easy-to-use physiological monitoring systems that support early detection of adverse cardiac events. Heart rate (HR) is among the most fundamental cardiovascular indicators, and wearable devices have emerged as a practical solution for long-term HR tracking outside clinical settings [9].

Among wearable sensor modalities, inertial measurement units (IMUs) offer advantages over photoplethysmography (PPG) and electrocardiography (ECG): no optical or electrical skin contact is required, a single sensor can simultaneously provide activity recognition, respiration monitoring, and HR estimation [4, 6], and the approach is robust to skin tone variation that affects PPG accuracy [11]. IMU-based seismocardiography (SCG) captures cardiac mechanical vibrations propagated through body tissue [7, 8], enabling HR estimation at moderate sampling rates (200 Hz) suitable for low-power wearable deployment.

However, the accuracy of IMU-based HR estimation depends strongly on the signal processing pipeline and sensor placement. Prior studies have explored individual aspects (Haescher et al. [4, 5]) demonstrated wrist-based IMU HR monitoring using FFT methods, and Kontaxis et al. [9] achieved MAE = 1.32 BPM at the waist during sleep using adaptive filtering. Taebi et al. [8] note that SCG amplitude varies by more than 30% with just 1 cm of displacement, yet existing studies primarily focus on single locations (sternum or wrist) without systematic cross-position comparison. No prior study has evaluated which body positions, filtering strategies, and detection methods support reliable HR estimation using a unified statistical framework with effect size analysis.

This paper addresses that gap by proposing a systematic pipeline evaluating two filtering strategies, two HR detection methods, five candidate signals, and two polarities across 20 body positions on 5 subjects. We address four research questions:

RQ1: Which filtering strategy yields the best HR estimation accuracy?
RQ2: Which HR detection method (FFT or peak-based) performs best?
RQ3: Which accelerometer axis provides the most reliable HR estimate?
RQ4: Which body position achieves the highest HR detection accuracy?

By answering these questions through ANOVA and effect size analysis (Cohen's d, η^2) [12], distinguishing statistical from practical significance, this work provides quantitative guidance for IMU-based wearable HR system design.

2 Methodology

2.1 Data Collection

Hardware Setup The measurement system consisted of an ESP32-S3 QT Py microcontroller [16] and an LSM6DSOX IMU sensor [17] recording tri-axial acceleration at 200 Hz. A Wellue FS20F Bluetooth finger-clip pulse oximeter [20] served as the heart rate reference, providing PPG-based beat-averaged readings at approximately 1 Hz with ±1–2 BPM accuracy under resting conditions [9, 4]. All data were transmitted to a laptop via serial connection for offline processing.

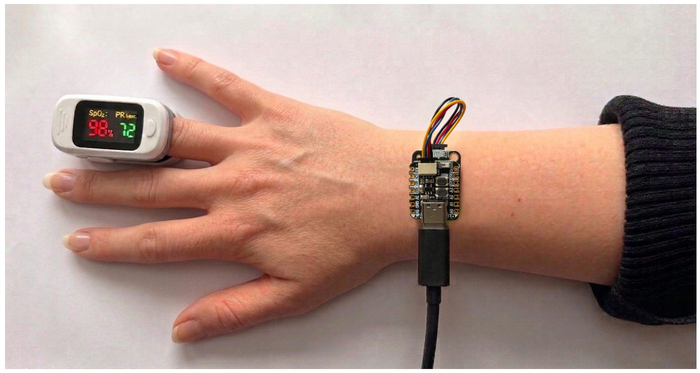


Fig. 1: Showing the apparatus (ESP32-S3 QT Py microcontroller) mounted on top of the LSM6DSOX IMU sensor breakoutboard, which is double-sided-taped to the user's wrist. A pulse oximeter provides the ground truth heart rate.

Participants and Protocol Five male participants (age: 22, BMI: 18.6–27.8, resting HR: 74–95 BPM; Table 1) were seated at rest. Data were recorded for two minutes per position ($5 \times 20 \times 120 = 12{,}000$ s total). Depending on position, measurements were performed either directly on skin or through a single layer of clothing.

Table 1: Participant characteristics and resting heart rate (mean ± SD across all 20 positions, reference pulse oximeter).

No.	Sex	Age	Height (m)	Weight (kg)	BMI	HR (BPM)
1	Male	22	1.74	65	21.47	88.2 ± 4.5
2	Male	22	1.83	80	23.89	74.1 ± 4.1
3	Male	22	1.75	85	27.76	94.5 ± 6.3
4	Male	22	1.72	55	18.59	81.7 ± 4.0
5	Male	22	1.75	70	22.86	88.4 ± 3.1

Measurement Positions Twenty body positions were selected on the left side of the anterior and posterior body surface (Fig. 2, Table 2), covering limbs, chest, abdomen, neck, and back. Highly mobile joint regions were excluded; common wearable locations (wrist, ankle) were retained for practical relevance.

2.2 Signal Processing Pipeline

The pipeline evaluates 40 processing configurations per recording: 5 candidate signals × 2 polarities × 2 filter strategies × 2 HR detection methods.

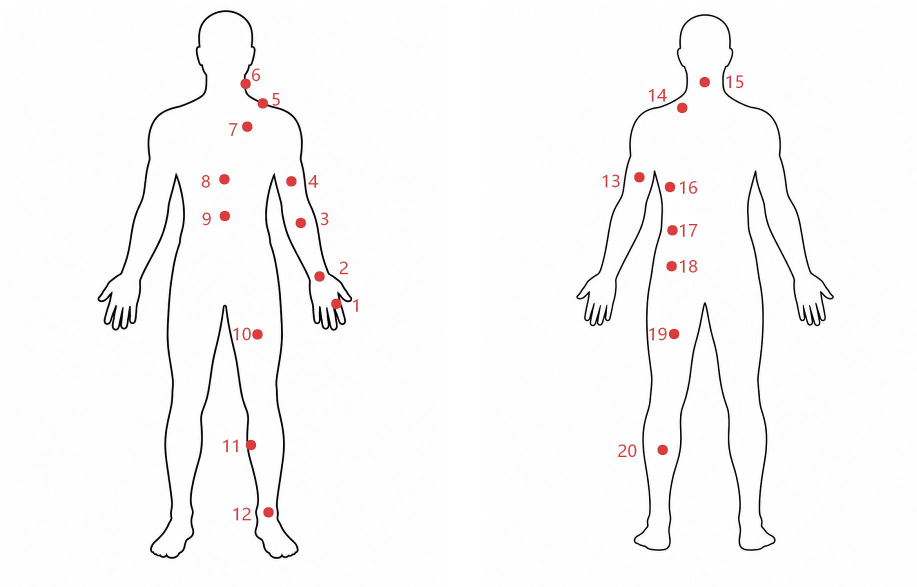


Fig. 2: Measurement positions on the anterior (left) and posterior (right) body surface.

Table 2: Measurement positions used in this study.

No.	Position	Side
1	Index finger	Anterior
2	Wrist	Anterior
3	Forearm	Anterior
4	Upper arm	Anterior
5	Upper trapezius	Anterior
6	Lateral neck	Anterior
7	Upper left chest	Anterior
8	Lower sternum	Anterior
9	Upper abdomen	Anterior
10	Anterior mid-thigh	Anterior
11	Medial lower leg	Anterior
12	Ankle	Anterior
13	Triceps region	Posterior
14	Posterior trapezius	Posterior
15	Posterior neck	Posterior
16	Below scapula	Posterior
17	Lower back	Posterior
18	Post. superior iliac spine	Posterior
19	Posterior mid-thigh	Posterior
20	Posterior mid-calf	Posterior

Steps 1–3: Loading, Candidates, and Preprocessing Raw IMU data are cleaned and resampled to a uniform 200 Hz grid using PCHIP interpolation [14], then synchronized with the pulse oximeter. Five candidate signals are generated: the three acceleration axes (a_x, a_y, a_z), the vector magnitude ($|\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$), and the first principal component (PCA1) [5]. Each candidate undergoes NaN interpolation, outlier clipping at ±4 SD, and linear detrending.

Step 4: Filtering Two strategies are applied in parallel:

Direct Bandpass: A fourth-order Butterworth bandpass filter isolates 0.75–3.0 Hz (45–180 BPM) [4] directly from the acceleration signal. Initial filter transients are discarded where RMS exceeds three times the median energy. The output is smoothed with a Savitzky–Golay filter (order 3, window 0.15 s) [19].

Hilbert Envelope: The signal is bandpass-filtered at 4–35 Hz to isolate SCG cardiac vibrations [8]. The Hilbert transform extracts the amplitude envelope [15], smoothed with a 50 ms moving average. The envelope is then bandpass-filtered at 0.75–3.0 Hz and smoothed with a Savitzky–Golay filter (order 3, window 0.15 s).

Fig. 3 illustrates both strategies for a representative segment. The Direct Bandpass output (b) remains contaminated by non-cardiac spectral content, while the Hilbert Envelope (c) shows a clean periodic signal with identifiable cardiac peaks.

Step 5: Segmentation Filtered signals are divided into 20-second windows with 5-second step (75% overlap). This window length ensures at least 15 cardiac cycles at 45 BPM, providing sufficient basis for FFT frequency resolution ($\Delta f = 0.05$ Hz ≈ 3 BPM) and robust median-based interval estimation. Longer windows would improve resolution but reduce the number of windows per recording, limiting statistical robustness. Each window is paired with the corresponding pulse oximeter HR value, yielding approximately 19 windows per recording.

Step 6: Heart Rate Detection Both signal polarities (+1, −1) are tested because cardiac-induced acceleration direction depends on sensor orientation. This doubles the filtered signals to 20 per recording (5 candidates × 2 strategies × 2 polarities). Two HR detection methods are applied to each window:

FFT Method: A Hann window is applied, followed by zero-padded FFT (4× signal length for 0.75 BPM peak localization precision). The dominant spectral peak within 0.75–3.0 Hz is identified and converted to BPM ($\text{HR} = f_{\text{peak}} \times 60$). Quality is scored from spectral SNR (60% weight) and dominance ratio (40% weight).

Peak Method: Peaks are detected using adaptive thresholds: minimum distance 0.4 s (≤150 BPM), minimum prominence 0.3 × SD, minimum height mean + 0.2 × SD. Inter-beat intervals (IBI) are filtered by physiological bounds (0.33–1.33 s, i.e., 45–180 BPM) and outlier rejection (>30% deviation from median). Heart rate is estimated as HR = 60/median(IBI). Quality is scored from

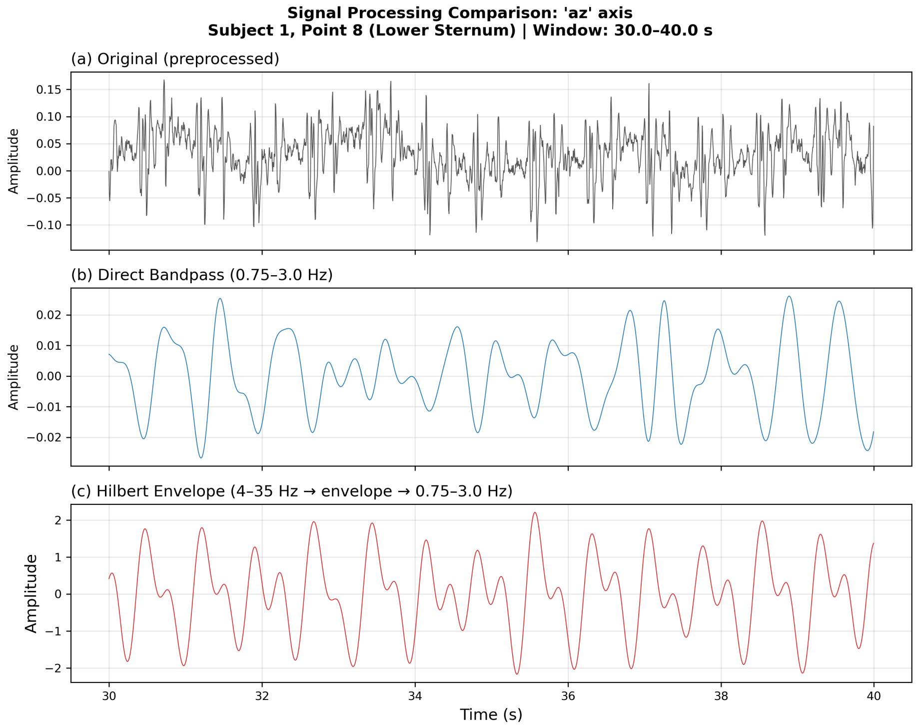


Fig. 3: Signal processing comparison for the a_z axis (Subject 1, Point 8, 10-second segment). (a) Preprocessed signal, (b) Direct Bandpass, (c) Hilbert Envelope.

interval regularity (50%), temporal coverage (25%), and peak count adequacy (25%).

Fig. 4 shows both methods on the same window.

Windows with quality below 0.3 or HR outside 40–200 BPM are excluded (<0.2% of windows), yielding a final dataset of 75,830 per-window observations.

2.3 Evaluation

Metrics For each configuration, three per-window error metrics are computed: Mean Error (ME $= HR_{est} - HR_{ref}$, systematic bias), Mean Absolute Error (MAE $= |HR_{est} - HR_{ref}|$), and Mean Absolute Percentage Error (MAPE $= |HR_{est} - HR_{ref}|/HR_{ref} \times 100\%$).

Statistical Analysis Given the large sample size ($\sim$75,830 observations), even trivial differences reach statistical significance ($p \leq 10^{-5}$). We therefore employ effect size measures to distinguish statistical from practical significance [12].

All research questions are evaluated using one-way ANOVA supplemented by:

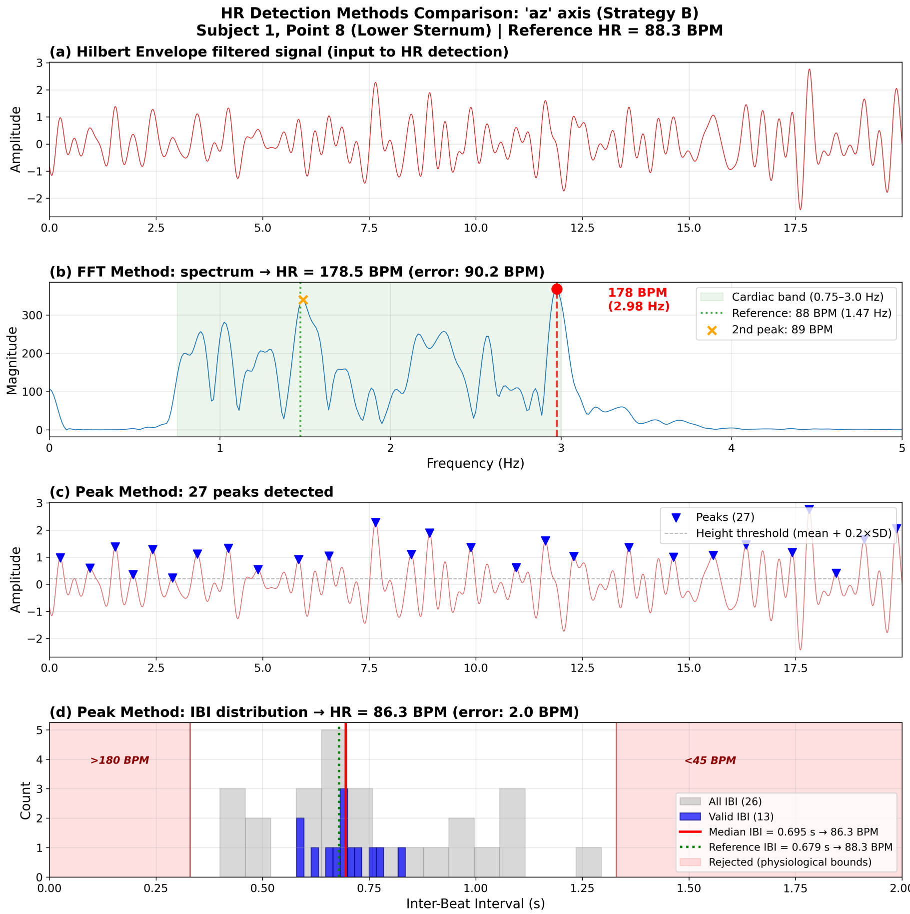


Fig. 4: FFT vs. Peak detection on a single analysis window (a_z, Hilbert Envelope, Point 8). Red zones in (d) indicate rejected physiological bounds.

Eta-squared (η^2): proportion of variance explained by the grouping factor. Thresholds [12]: small (0.01), medium (0.06), large (0.14).

Cohen's d: standardized mean difference between groups. Thresholds: negligible (<0.2), small (0.2–0.5), medium (0.5–0.8), large (>0.8). Positive d indicates higher MAE (worse performance).

For multi-group comparisons (RQ3, RQ4), a pairwise Cohen's d matrix is computed and groups are ranked by average d (negative = better). RQ3 and RQ4 use only the optimal pipeline (Hilbert Envelope + Peak) to avoid inflating within-group variance with suboptimal configurations.

Table 3: Statistical analysis framework for each research question.

Research Question	**Groups**	**Omnibus**	**Effect Size**
RQ1: Filter strategy	2	ANOVA + η^2	Cohen's d
RQ2: HR detection	2	ANOVA + η^2	Cohen's d
RQ3: Candidate axis	5	ANOVA + η^2	Pairwise d
RQ4: Body position	20	ANOVA + η^2	Pairwise d

3 Results

3.1 RQ1: Which Filter Strategy is Best?

Table 4 summarizes the comparison between the Direct Bandpass and Hilbert Envelope strategies across all 20 measurement points, 5 candidate signals, both polarities, and both HR detection methods (75,830 per-window observations).

Sign convention. Cohen's d is computed as $(\mu_{\text{Direct}} - \mu_{\text{Envelope}})/\sigma_{\text{pooled}}$: a positive value indicates Direct Bandpass has higher MAE, i.e. the Hilbert Envelope is more accurate; a negative value indicates the opposite.

Table 4: Comparison of filtering strategies across all positions.

Metric	Direct Bandpass	Hilbert Envelope	Cohen's d
Mean ME (BPM)	+14.56	+2.60	+0.48 (small)
Median ME (BPM)	+1.45	+0.11	—
Mean MAE (BPM)	21.26	6.56	+0.66 (medium)
Median MAE (BPM)	6.93	2.03	—
Mean MAPE (%)	25.72	8.17	+0.64 (medium)
Median MAPE (%)	8.09	2.35	—
Points won	0/20	20/20	—

One-way ANOVA confirmed highly significant differences for all three error metrics (ME: $F = 4326.94$; MAE: $F = 8366.51$; MAPE: $F = 7819.30$; all $p < 10^{-5}$). The Hilbert Envelope reduced mean MAE by 14.69 BPM (Cohen's $d = +0.66$, medium effect) and MAPE by 17.55 percentage points ($d = +0.64$, medium effect). The signed error revealed that Direct Bandpass exhibits a systematic positive bias of +14.56 BPM (median: +1.45 BPM), indicating consistent overestimation caused by spectral leakage of non-cardiac components into the 0.75–3.0 Hz cardiac band. The Hilbert Envelope achieved near-zero bias (ME = +2.60 BPM, median = +0.11 BPM, $d = +0.48$), confirming effective isolation of cardiac content.

At the per-point level, the Hilbert Envelope was significantly better at all 20 positions for MAE and MAPE. Per-point effect sizes ranged from $d = +0.35$ (small, Point 12) to $d = +1.00$ (large, Point 16); six positions showed large effects ($d > 0.8$) and the remaining 14 medium effects ($d > 0.5$).

3.2 RQ2: Which HR Detection Method is Best?

To isolate the effect of detection method from filtering strategy, this analysis uses only the Hilbert Envelope pipeline (19,040 FFT + 19,026 Peak windows = 38,066 observations). Table 5 summarizes the comparison.

Sign convention. Cohen's d is computed as $(\mu_{\text{FFT}} - \mu_{\text{Peak}})/\sigma_{\text{pooled}}$: a positive value indicates FFT has higher MAE, i.e. Peak detection is more accurate; a negative value indicates the opposite.

Table 5: Comparison of HR detection methods (Hilbert Envelope only).

Metric	FFT	Peak Detection	Cohen's d
Mean ME (BPM)	+3.11	+2.10	+0.07 (negligible)
Median ME (BPM)	+0.20	+0.00	—
Mean MAE (BPM)	7.95	5.17	+0.21 (small)
Median MAE (BPM)	1.70	2.40	—
Mean MAPE (%)	9.88	6.46	+0.20 (small)
Median MAPE (%)	1.96	2.80	—
Points won (MAE)	0/20	20/20	—

One-way ANOVA confirmed statistically significant differences for all three metrics (MAE: $F = 434.46$; MAPE: $F = 391.18$; ME: $F = 46.93$; all $p < 10^{-5}$). The practical magnitude was, however, small: peak detection reduced mean MAE by 2.78 BPM ($d = +0.21$) and MAPE by 3.42 percentage points ($d = +0.20$), while the bias difference was negligible ($d = +0.07$). Peak detection won at all 20 positions for MAE, but only 16/20 reached statistical significance; per-point effect sizes ranged from $d = +0.03$ (negligible, Point 16) to $d = +0.46$ (small, Point 12), with no position reaching a medium effect ($d \geq 0.5$). The six positions with the largest effects (Points 12, 14, 6, 3, 1, 8; $d = 0.31$–0.46) spanned both trunk and limb locations without a clear anatomical pattern.

A notable nuance emerged in the median errors: FFT achieved a lower median MAE (1.70 vs. 2.40 BPM) despite a substantially higher mean (7.95 vs. 5.17 BPM). Per-point inspection explains this reversal: FFT produces accurate estimates for the majority of windows but generates occasional large errors due to harmonic misidentification, which inflate the mean disproportionately. Peak detection avoids these outliers at the cost of a slightly higher baseline error on well-conditioned windows, making it the more *reliable* choice in practice — particularly at high-motion positions where harmonic confusion is most likely.

Methodological note. When both filtering strategies are pooled (75,830 observations), the FFT–Peak difference appears substantially larger ($d = +0.69$), because FFT is disproportionately degraded by Direct Bandpass signals. The Envelope-only comparison reported here controls for this confound, revealing a smaller but consistent advantage for peak detection.

3.3 RQ3: Which Candidate Axis is Best?

Restricted to the optimal pipeline (Hilbert Envelope + Peak, 19,026 observations), candidate axis choice had no detectable effect on accuracy. Only MAE reached borderline significance ($F = 3.02$, $p = 0.017$, $\eta^2 = 0.0006$), while ME ($F = 0.98$, $p = 0.42$) and MAPE ($F = 2.18$, $p = 0.068$) were *not* significant. All pairwise Cohen's $d < 0.06$. At the per-point level, the best candidate changed unpredictably across positions and the mean per-point η^2 was 0.023 (small), with only 2/20 positions reaching a medium effect ($\eta^2 \geq 0.06$). The top-ranked candidates (pca1 and a_z) differed by less than 0.01 BPM in both mean and median MAE (Table 6). Per-position details for the seven locations with the largest effects are provided in Table 12 in the Appendix. Given the negligible overall effect, the five candidates are pooled for the remainder of the analysis.

Table 6: Candidate ranking using only the optimal pipeline (Hilbert Envelope + Peak detection, 19,026 observations).

Rank	Cand.	Mean MAE	Med. MAE	Mean ME	Med. ME	Mean MAPE	Med. MAPE	Avg d (MAE)	Avg Rank
1	pca1	4.92	2.24	+1.94	+0.05	6.18%	2.58%	−0.038	2.75
2	a_z	4.93	2.23	+1.93	−0.04	6.17%	2.58%	−0.036	2.50
3	mag	5.25	2.54	+2.20	−0.05	6.56%	2.95%	+0.012	3.20
4	a_y	5.33	2.53	+2.15	+0.00	6.60%	2.94%	+0.025	3.30
5	a_x	5.43	2.57	+2.27	+0.02	6.77%	2.97%	+0.038	3.25

3.4 RQ4: Which Body Position is Best?

Using only the optimal pipeline (Hilbert Envelope + Peak detection, 19,026 per-window observations), Table 7 presents the ranking of all 20 measurement points. All 5 candidates in both polarities are pooled per position, as the per-point candidate analysis (RQ3) showed only small effects (mean $\eta^2 = 0.023$, only 2/20 reaching medium) with no consistent best axis.

One-way ANOVA was highly significant for all three metrics (ME: $F = 54.49$, $\eta^2 = 0.0517$; MAE: $F(19, 19006) = 102.02$, $\eta^2 = 0.0925$; MAPE: $F = 91.36$, $\eta^2 = 0.0837$; all $p < 10^{-5}$), confirming that body position explains 5–9% of total variance (medium effect for MAE and MAPE, small-to-medium for ME). Fig. 5 visualizes the full error distribution.

Point 8 (lower sternum) achieved the best performance: MAE = 1.65 BPM (median: 1.13), near-zero bias (ME = +0.19 BPM, median = −0.09 BPM), and lowest relative error (MAPE = 2.06%, median = 1.36%). It was significantly better than all 19 other positions, with pairwise Cohen's d ranging from +0.22 (vs. Point 7) to +1.17 (vs. Point 12). This result aligns with the anatomical proximity of the xiphoid process to the heart, minimizing vibration attenuation through direct bone conduction via the sternum [7, 8].

Table 7: Complete measurement point ranking (Hilbert Envelope + Peak, all candidates pooled, 19,026 observations).

Rank	Location	Mean MAE	Med. MAE	Mean ME	Med. ME	Mean MAPE	Med. MAPE	Avg d (MAE)
1	Lower sternum (P8)	1.65	1.13	+0.19	−0.09	2.06%	1.36%	−0.583
2	Below scapula (P16)	2.34	2.01	−1.54	−1.55	2.73%	2.38%	−0.432
3	Post. trapezius (P14)	2.42	1.61	−0.08	+0.03	2.85%	1.84%	−0.405
4	Lateral neck (P6)	2.72	1.52	+0.31	+0.44	3.36%	1.81%	−0.318
5	Lower back (P17)	2.81	1.94	−0.42	−0.77	3.32%	2.20%	−0.313
6	Upper left chest (P7)	2.82	1.42	+0.95	−0.07	3.68%	1.59%	−0.276
7	Upper abdomen (P9)	3.44	2.23	+0.99	+0.18	4.27%	2.65%	−0.193
8	Post. neck (P15)	3.52	2.14	+0.51	−0.51	4.24%	2.49%	−0.178
9	Triceps region (P13)	3.63	2.16	+0.35	−0.50	4.51%	2.49%	−0.161
10	Upper trapezius (P5)	4.37	1.74	+2.58	+0.04	5.86%	1.90%	−0.055
11	Post. sup. iliac (P18)	5.18	3.07	+1.09	−0.44	6.07%	3.55%	+0.059
12	Ant. mid-thigh (P10)	5.45	2.61	+2.58	+0.13	7.00%	3.00%	+0.096
13	Post. mid-calf (P20)	5.92	3.05	+2.33	+0.08	7.12%	3.55%	+0.151
14	Index finger (P1)	6.18	3.15	+2.38	+0.37	7.21%	3.47%	+0.180
15	Forearm (P3)	7.29	3.49	+4.12	+0.49	9.14%	4.17%	+0.286
16	Upper arm (P4)	7.66	2.97	+5.07	+0.75	10.24%	3.60%	+0.288
17	Med. lower leg (P11)	7.64	3.26	+4.54	+0.28	9.98%	3.84%	+0.307
18	Post. mid-thigh (P19)	8.09	4.20	+3.62	−0.12	9.85%	4.76%	+0.405
19	Wrist (P2)	9.73	4.88	+6.43	+2.05	12.46%	5.55%	+0.486
20	Ankle (P12)	10.56	7.10	+5.96	+2.96	13.18%	8.15%	+0.655

To complement the statistical boxplot analysis, Fig. 6 visualizes the spatial distribution of HR estimation accuracy across the human body. Points are color-coded by mean MAE values according to a four-tier gradient: green indicates superior performance (median MAE $<$ 3 BPM), blue represents good accuracy (3 median MAE $<$ 5 BPM), yellow denotes moderate accuracy (5 median MAE $<$ 8 BPM), and red indicates suboptimal performance (median MAE 8 BPM).

Region-based comparison. Positions were grouped into trunk (Points 5–9, 14–18; $N = 9{,}510$) and limb (Points 1–4, 10–13, 19–20; $N = 9{,}516$). Table 8 summarizes the comparison.

Sign convention. Cohen's d for the trunk/limb comparison is computed as $(\mu_{\text{Limbs}} - \mu_{\text{Trunk}})/\sigma_{\text{pooled}}$: a positive value indicates limb positions have higher MAE, i.e. trunk positions are more accurate.

One-way ANOVA confirmed highly significant differences (ME: $F = 567.47$, $\eta^2 = 0.029$, $d = +0.35$; MAE: $F = 1195.84$, $\eta^2 = 0.059$, $d = +0.50$; MAPE: $F = 1064.56$, $\eta^2 = 0.053$, $d = +0.47$; all $p < 10^{-5}$). Trunk positions achieved nearly unbiased estimation (mean ME = +0.46 BPM, median = −0.24 BPM, 46%/54% over-/underestimation), while limb positions exhibited systematic positive bias (mean ME = +3.74 BPM, 53.8% overestimation).

A sub-group analysis separated limbs into upper (Points 1–4, 13) and lower (Points 10–12, 19–20). Both performed worse than trunk (upper: $d = +0.50$;

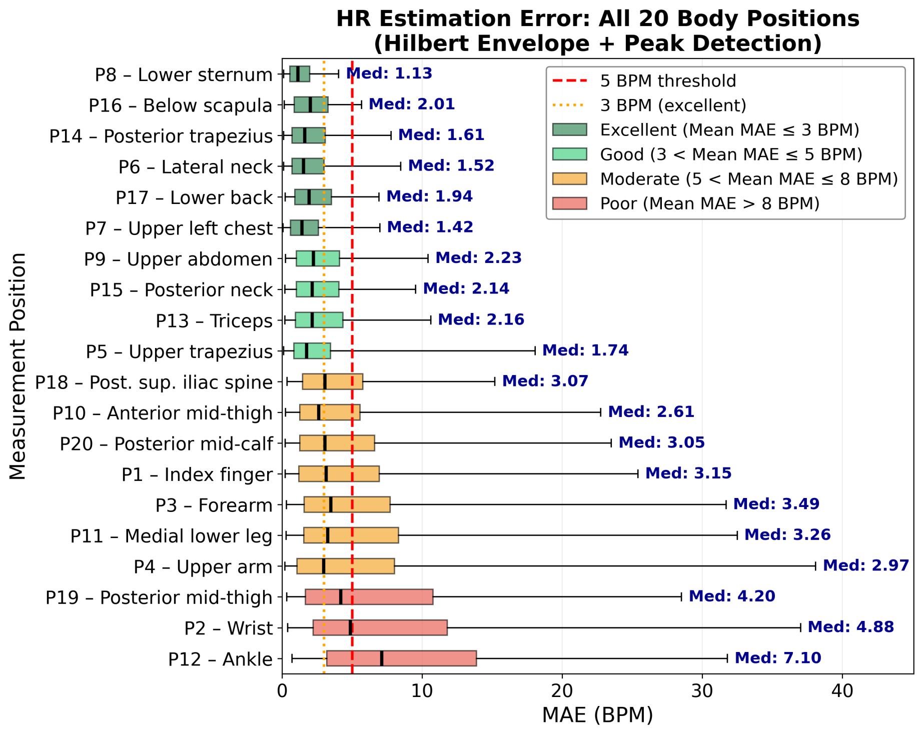


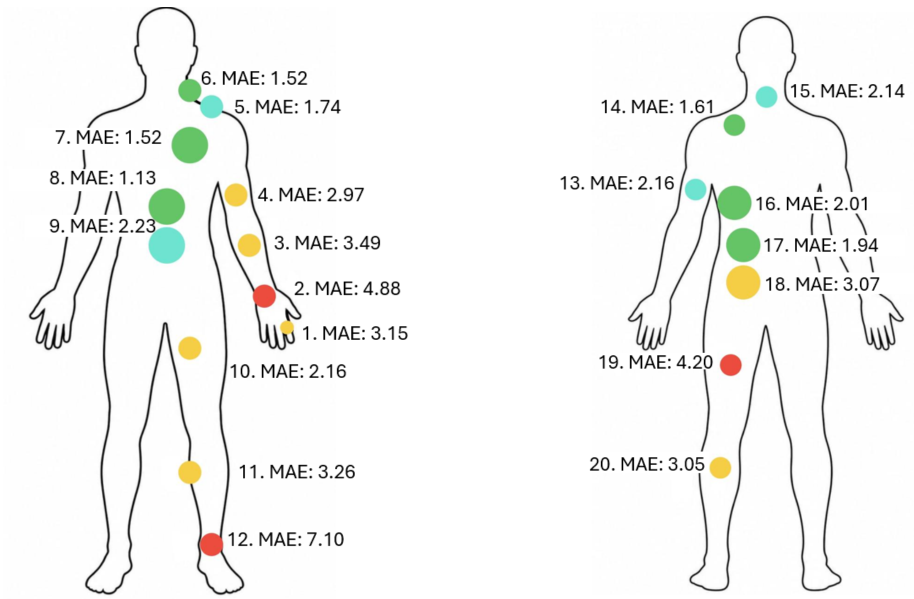
Fig. 5: HR estimation error distribution across all 20 body positions (Hilbert Envelope + Peak detection, all candidates pooled). Positions are sorted by mean MAE. Dashed lines indicate 3 BPM and 5 BPM thresholds.

Table 8: Trunk vs. limbs comparison (Hilbert Envelope + Peak detection).

Metric	Trunk	Limbs	Difference
Mean ME (BPM)	+0.46	+3.74	+3.28
Median ME (BPM)	−0.24	+0.39	+0.63
Mean MAE (BPM)	3.13	7.21	+4.09
Median MAE (BPM)	1.81	3.44	+1.63
Mean MAPE (%)	3.84	9.07	+5.22
Median MAPE (%)	2.10	3.98	+1.88
Q95 MAE (BPM)	9.29	29.85	+20.56
% windows ≤5 BPM	86.0%	62.2%	−23.8%
% windows ≤3 BPM	70.1%	45.3%	−24.8%

lower: $d = +0.61$), but the difference between upper and lower limbs was negligible ($d = +0.06$), indicating that attenuation occurs primarily at the trunk-to-limb transition.

Best limb position and practical alternative to trunk placement. Among all limb positions, Point 13 (triceps region) stands out as a clinically rel-



(a) Anterior view

(b) Posterior view

Fig. 6: HR estimation accuracy across body positions visualized on anatomical maps. The written values around circles correspond the number of measure points and their median MAE. The color-coded by mean MAE values according to a four-tier gradient.

evant alternative to trunk-mounted sensors. It achieved mean MAE = 3.63 BPM (median: 2.16 BPM), MAPE = 4.51%, and 80.6% of windows within 5 BPM , performance that approaches the lower trunk positions (cf. Points 9, 15, 18 in Table 7). Crucially, it was the only limb location with near-zero and symmetric bias (mean ME = +0.35 BPM, median = −0.50 BPM), whereas all other limb positions exhibited systematic overestimation (mean ME ranging from +2.33 to +6.43 BPM). This behaviour is attributable to the mechanical coupling of the upper arm to the thorax via the shoulder joint, which preserves cardiac vibration transmission better than more distal segments.

From a practical standpoint, this makes the triceps a recommended fallback when trunk sensor placement is not feasible, for example, in subjects who find chest or back-mounted devices uncomfortable, in post-operative patients with thoracic wounds, or in wearable applications requiring unobtrusive placement. By contrast, the wrist (Point 2), the most common consumer wearable location, achieved MAE = 9.73 BPM with substantial positive bias (ME = +6.43 BPM, MAPE = 12.46%) and failed to meet the 5 BPM threshold under any pipeline configuration, confirming it as the least reliable position tested.

3.5 Inter-Subject Variability

This analysis uses only the optimal pipeline (Hilbert Envelope + Peak detection) with all five candidates pooled (19,026 observations across 5 subjects, 20 positions, ≈190 windows per subject–position pair).

Overall inter-subject performance Table 9 summarises per-subject global error. Four of five subjects showed consistent performance (MAE: 3.35–4.34 BPM, ME: −0.57 to +1.05 BPM), while Subject 2 exhibited substantially elevated error (MAE = 11.31 BPM) with strong systematic positive bias (ME = +10.63 BPM), indicating persistent HR overestimation across nearly all positions.

Table 9: Per-subject global error (Hilbert Envelope + Peak, all positions and candidates pooled).

Subject	BMI	Resting HR (BPM)	Mean MAE (BPM)	Mean ME (BPM)	Mean MAPE (%)
1	21.47	88.2	3.45	−0.29	4.11
2	23.89	74.1	11.31	+10.63	14.67
3	27.76	94.5	3.35	−0.30	3.92
4	18.59	81.7	4.34	+1.05	5.35
5	22.86	88.4	3.43	−0.57	4.17

Subject 2: systematic bias and likely mechanism Subject 2 showed severe overestimation at 15 of 20 positions (MAE > 5 BPM), with a near-pure positive bias pattern: 84.2% of windows were overestimations (mean ME when overestimating: +13.03 BPM), compared to 41.4% for all other subjects combined.

The most likely mechanism is related to Subject 2's substantially lower resting HR (74.1 BPM vs. 81.7–94.5 BPM for others), corresponding to longer interbeat intervals (≈0.81 s vs. 0.63–0.73 s). Longer IBIs increase the probability that spurious noise peaks between true cardiac events are accepted by the peak detector, shortening the estimated median IBI and producing systematic HR overestimation.

Position-dependent resilience to subject variability To quantify how sensor placement affects robustness to inter-subject differences, Table **??** compares three representative positions: Point 8 (lower sternum, best overall trunk position from RQ4), Point 13 (triceps, best limb position from RQ4), and Point 2 (wrist, standard consumer wearable reference).

Excluding Subject 2 improved absolute error at all three positions: MAE by 23.6% at Point 8, 27.8% at Point 13, and 35.1% at Point 2; MAPE by 28.2%, 31.5%, and 42.1% respectively. However, ME at Points 8 and 13 moved *farther* from zero after exclusion, revealing that Subject 2's strong positive bias had been

Table 10: Group error at three key positions, reported including and excluding Subject 2.

Position	Group	MAE (BPM)	ME (BPM)	MAPE (%)
P8 — Lower sternum	All 5 (mean)	1.65	+0.39	2.06
	Excl. Sub 2 (mean)	1.26	−0.54	1.48
	All 5 / Excl. Sub 2	1.31	−0.72	1.39
P13 — Triceps	All 5 (mean)	3.63	+0.38	4.51
	Excl. Sub 2 (mean)	2.62	−1.30	3.09
	All 5 / Excl. Sub 2	1.39	−0.29	1.46
P2 — Wrist	All 5 (mean)	9.73	+6.44	12.46
	Excl. Sub 2 (mean)	6.31	+2.20	7.21
	All 5 / Excl. Sub 2	1.54	2.93	1.73

partially cancelling the slight negative bias of the remaining subjects. Notably, Point 2 remained unacceptable even without Subject 2 (MAE = 6.31 BPM, MAPE = 7.21%), confirming that wrist placement is inherently unreliable regardless of individual variability.

Position ranking consistency across subjects Kendall's coefficient of concordance across all five subjects was $W = 0.667$ (moderate agreement), rising to $W = 0.758$ (strong agreement) when Subject 2 was excluded. In both cases the trunk-superior, limb-inferior ordering was preserved: Point 8 ranked first or second for all subjects, and Points 2 and 12 ranked last or second-to-last for all subjects. This confirms that the position ranking reported in RQ4 is robust across individuals, while absolute error magnitude is sensitive to individual cardiac characteristics — in particular resting heart rate.

4 Discussion

4.1 Key Findings

Pipeline Design: Hilbert Envelope and Peak Detection The superiority of the Hilbert Envelope strategy ($d = +0.66$) has a clear physiological basis: cardiac mechanical energy resides in the 4–35 Hz SCG band [7, 8], while the beat-to-beat rhythm manifests as amplitude modulation of these vibrations at 0.75–3.0 Hz. Direct bandpass filtering attempts to extract this rhythm from the raw signal, where it is buried beneath respiration harmonics and sensor noise. The Hilbert Envelope approach demodulates the cardiac rhythm from its carrier, increasing effective SNR in the heart rate band.

Peak detection outperformed FFT ($d = +0.69$) because FFT assumes signal stationarity and is vulnerable to harmonic confusion, the second harmonic of the fundamental HR frequency can exceed the true peak in amplitude, which may introduce estimation bias. Peak detection avoids this by directly measuring interbeat intervals in the time domain, without assumptions about signal periodicity.

This contrasts with Haescher et al. [4, 5], who relied exclusively on FFT; our results suggest time-domain methods would improve their reported accuracy.

Irrelevance of Axis Selection The negligible effect of candidate axis choice ($\eta^2 = 0.001$) is the most practically useful finding for hardware designers. The Hilbert Envelope extracts amplitude modulation of cardiac vibrations regardless of sensor orientation, as the envelope magnitude is rotation-invariant by construction. As a result, the mean MAE difference between the best and worst axis across all positions was only 2.72 BPM, and in the optimal pipeline all pairwise Cohen's $d < 0.06$ with no significant difference between axes ($F = 3.02$, $p = 0.017$). Since any single axis yields equivalent performance, it may be sufficient to acquire only the z-axis acceleration rather than the full triaxial signal, reducing both sampling and computational load, and potentially extending battery life in resource-constrained devices.

Body Position: The Critical Design Factor Body position was the only factor showing medium effect size for estimation accuracy ($\eta^2 = 0.093$). The performance gradient from trunk to limbs follows the expected attenuation of cardiac vibrations with distance from the heart [7, 8]. The lower sternum (Point 8), located over the xiphoid process, benefits from direct bone conduction through the sternum, consistent with classical SCG literature identifying this location as optimal [8].

The trunk–limb transition represents the primary degradation boundary: upper limbs vs. trunk showed $d = +0.50$, lower limbs vs. trunk $d = +0.61$, but upper vs. lower limbs only $d = +0.06$ (negligible). Among limb positions, the triceps (Point 13, MAE = 3.63 BPM) performed comparably to several trunk positions. A plausible explanation is that the major arterial pathway from the aorta to the upper limb, the subclavian and brachial arteries, passes in close proximity to the triceps region, providing a direct vascular transmission path for cardiac mechanical activity. However, this remains a hypothesis and would require dedicated vascular imaging or pulse wave measurements to confirm.

Inter-Subject Variability and Real-World Deployment Four of five subjects showed consistent performance (MAE: 3.35–4.34 BPM), but Subject 2 exhibited substantially elevated error (MAE = 11.31 BPM, MAPE = 14.67%) driven by a persistently lower resting HR (74.1 BPM vs. 81.7–94.5 BPM for others). Longer inter-beat intervals increase the probability that spurious noise peaks are accepted by the peak detector, producing systematic overestimation. This has a direct implication for real-world deployment: users with naturally low resting heart rates (athletes, individuals with bradycardia, or older adults) may experience systematically degraded accuracy with the current peak detection approach.

Critically, position selection partially mitigates this risk. At Point 8 (lower sternum), Subject 2's MAE remained 3.22 BPM, well within acceptable range, while at Point 2 (wrist) it reached 23.41 BPM, confirming that robust sensor placement can compensate for individual physiological variability to a degree that limb placements cannot match.

4.2 Comparison with Prior Work

Table 11 places our results in context with both IMU-based and PPG-based wearable studies.

Table 11: Comparison with related HR estimation studies.

Study	Loc./cond.	Method	Reported	Our closest result
IMU-based				
Haescher et al. [4]	Wrist	IMU	SD 1.63%	P13: MAPE 4.51%
Kontaxis et al. [9]	Waist, sleep	IMU	MAE 1.32 BPM	P8: MAE 1.65 BPM
PPG-based benchmark				
Nelson and Allen [10]	Wrist	Apple Watch 3	MAE 5.24 BPM; MAPE 7.21%	P13: MAE 3.63 BPM; MAPE 4.51%
	Wrist	Fitbit Charge 2	MAE 5.93 BPM; MAPE 6.93%	P13: MAE 3.63 BPM; MAPE 4.51%
Bent et al. [11]	Wrist	Consumer PPG	MAE 7.2 BPM	P13: MAE 3.63 BPM

Note: Comparisons are approximate because studies differ in sensor placement, modality, experimental condition, and reported metrics.

The comparison should be interpreted with caution because the studies differ in sensor placement, posture, recording duration, and reported metrics. Haescher et al. [4] included 15 participants and reported an SD of 1.63%, while the closest result in the present work was P13 with a MAPE of 4.51%. A larger subject number in the present work could further improve the statistical reliability of the results. Kontaxis et al. [9] measured HR during sleep in a lying position and reported an MAE of 1.32 BPM, where motion artifacts are reduced and long overnight recordings can improve signal stability. In contrast, our measurements were performed in a sitting resting state, with P8 achieving a similar MAE of 1.65 BPM. Compared with wrist-worn PPG studies, our IMU-based P13 result, with an MAE of 3.63 BPM and a MAPE of 4.51%, showed lower error than the Apple Watch 3 in sitting conditions (MAE 5.24 BPM; MAPE 7.21%), the Fitbit Charge 2 in sitting conditions (MAE 5.93 BPM; MAPE 6.93%), and the consumer PPG wearables reported by Bent et al. [11] at rest (MAE 7.2 BPM). This suggests that IMU-based HR detection can be competitive with PPG-based wearables under low-motion conditions.

4.3 Limitations

Several limitations should be considered when interpreting these results.

First, the sample consisted of only five male participants of identical age (22 years), which severely restricts demographic generalizability; findings may not extend to female, older, or clinical populations. urthermore, measurements at each position lasted only 120 seconds (2 minutes). Recent guidance on wearable HR measurement [13] recommends minimum stabilization periods of 4 minutes for reliable resting heart rate assessment. Data from Speed et al. show that at 2 minutes of inactivity, 42.7% of subjects still exhibited heart rate fluctuations exceeding 3 bpm compared to the preceding minute, declining only to 24.4% by 3 minutes. This indicates substantial non-stabilization risk in our 120-second protocol, potentially undermining the reliability of threshold parameter calibration across subjects.

Second, all measurements were conducted at seated rest. Ambulatory conditions introduce motion artefacts that would likely affect positions differently, trunk placements may prove more robust than limb placements under movement, but this remains untested.

Third, each subject was measured in a single session, so no test-retest reliability could be assessed. Within-subject variability across days or conditions is unknown.

Fourth, the resting HR range across subjects was narrow (74.1–94.5 BPM). Performance at exercise-elevated heart rates or in the presence of cardiac arrhythmias cannot be inferred from these results. The elevated error observed for Subject 2, whose lower resting HR appeared to degrade peak detection performance, further suggests that the pipeline may not generalise uniformly across the full physiological HR range.

Finally, the use of overlapping windows violates the independence assumption of ANOVA. However, consistent results between parametric ANOVA and non-parametric Kruskal-Wallis tests ($H = 2484.74$, $p \leq 10^{-5}$) suggest that conclusions are robust to this violation.

5 Conclusion

This work established that IMU-based heart rate monitoring using Hilbert Envelope demodulation and time-domain peak detection can achieve accuracy competitive with consumer PPG wearables in controlled settings. The triceps position (pos.13) yielded a $MAE = 3.63 BPM$ ($MAPE = 4.51\%$), outperforming the Apple Watch 3 ($MAE = 5.24 BPM$) and Fitbit Charge 2 ($MAE = 5.93 BPM$) at equivalent rest, while the lower sternum (pos.8) achieved the overall lowest error (mean $MAE 1.65 BPM$). Crucially, our findings revealed that algorithm and placement design outweigh raw data dimensionality; single-axis acquisition proved sufficient (mean axis variance of only $2.72 BPM$), showing that hardware simplification is viable to extend battery life in resource-constrained wearables.

However, transitioning from controlled to real-world deployment requires addressing two primary barriers: population-specific calibration and motion artifact robustness. Because the peak-detection threshold was manually defined on a small, homogeneous cohort, future work must expand the training demographics to accommodate natural physiological variability. Furthermore, evaluating performance during ambulatory activity and evaluating adaptive thresholding strategies remain essential steps to ensure stability in other ad-hoc sensing conditions.

A Per-Position Candidate Differences (RQ3)

Table 12 lists the seven positions with the largest candidate effects within the optimal pipeline (Hilbert Envelope + Peak detection). At the remaining 13 positions $\eta^2 < 0.02$ and no pairwise d exceeded 0.30.

Table 12: Positions with meaningful candidate differences (Envelope + Peak, sorted by η^2).

Pt	Location	η^2	Best	Worst	ΔMAE	d	$\mathrm{ME}_{\mathrm{best}}$	$\mathrm{ME}_{\mathrm{worst}}$	ΔMAPE
18	Post. sup. iliac	0.084	a_z	a_x	6.10	+0.58	−1.0	+5.0	+7.6%
8	Lower sternum	0.065	pca1	a_x	1.63	+0.50	+0.2	+1.2	+2.2%
2	Wrist	0.052	a_x	a_y	8.91	+0.73	+2.6	+10.4	+11.5%
14	Post. trapezius	0.044	a_z	a_y	1.40	+0.52	−0.3	−0.1	+1.6%
7	Upper left chest	0.030	pca1	a_z	3.54	+0.35	−0.1	+3.7	+5.1%
16	Below scapula	0.026	a_x	a_y	0.78	+0.44	−0.6	−2.2	+0.9%
19	Post. mid-thigh	0.024	mag	a_x	3.96	+0.43	+3.2	+4.2	+4.3%

References

1. World Health Organization: Cardiovascular diseases (CVDs) – Fact sheet, https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), last accessed 2025/07/31
2. Timmis, A., Conrad, N., Petersen, S.E., Van Belle, E., Kaplon-Cieslicka, A., Townsend, N., Kazakiewicz, D., Kavousi, M., Torbica, A., Cvijic, M., et al.: European Society of Cardiology: cardiovascular disease statistics 2025. European Heart Journal, 1–45 (2026). https://doi.org/10.1093/eurheartj/ehag345
3. Haescher, M., Höpfner, F., Chodan, W., Kraft, D., Aehnelt, M., Urban, B.: Transforming Seismocardiograms Into Electrocardiograms by Applying Convolutional Autoencoders. In: ICASSP 2020 – IEEE International Conference on Acoustics, Speech and Signal Processing (2020). https://doi.org/10.1109/ICASSP40776.2020.9053130
4. Haescher, M., Matthies, D.J.C., Trimpop, J., Urban, B.: A study on measuring heart- and respiration-rate via wrist-worn accelerometer-based seismocardiography

(SCG) in comparison to commonly applied technologies. In: 2nd International Workshop on Sensor-Based Activity Recognition and Interaction (iWOAR), p. 2. ACM, New York (2015). https://doi.org/10.1145/2790044.2790054
5. Haescher, M., Matthies, D.J.C., Trimpop, J., Urban, B.: SeismoTracker: Upgrade any smart wearable to enable a sensing of heart rate, respiration rate, and microvibrations. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2209–2216. ACM, New York (2016). https://doi.org/10.1145/2851581.2892279
6. Hung, P.D.: Estimating respiration rate using an accelerometer sensor. In: ACM International Conference Proceeding Series, pp. 11–14. ACM, New York (2017). https://doi.org/10.1145/3156346.3156349
7. Inan, O.T., Migeotte, P.-F., Park, K.-S., Etemadi, M., Tavakolian, K., Casanella, R., Zanetti, J., Tank, J., Funtova, I., Prisk, G.K., Di Rienzo, M.: Ballistocardiography and seismocardiography: A review of recent advances. IEEE Journal of Biomedical and Health Informatics **19**(4), pp. 1414–1427 (2015). https://doi.org/10.1109/JBHI.2014.2361732
8. Taebi, A., Solar, B.E., Bomar, A.J., Sandler, R.H., Mansy, H.A.: Recent advances in seismocardiography. Vibration **2**(1), 64–86 (2019). https://doi.org/10.3390/vibration2010005
9. Kontaxis, S., Kanellos, F., Ntanis, A., Kostikis, N., Konitsiotis, S., Rigas, G.: An inertial-based wearable system for monitoring vital signs during sleep. Sensors **24**(13), 4139 (2024). https://doi.org/10.3390/s24134139
10. Nelson, B.W., Allen, N.B.: Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. JMIR mHealth and uHealth 7(3), e10828 (2019). https://doi.org/10.2196/10828
11. Bent, B., Goldstein, B.A., Kibbe, W.A., Dunn, J.P.: Investigating sources of inaccuracy in wearable optical heart rate sensors. npj Digital Medicine 3, 18 (2020). https://doi.org/10.1038/s41746-020-0226-6
12. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Lawrence Erlbaum Associates, Hillsdale (1988)
13. Speed, C., Arneil, T., Harle, R., et al.: Measure by measure: Resting heart rate across the 24-hour cycle. PLOS Digital Health 2(4), e0000236 (2023). https://doi.org/10.1371/journal.pdig.0000236
14. Fritsch, F.N., Carlson, R.E.: Monotone piecewise cubic interpolation. SIAM Journal on Numerical Analysis **17**(2), 238–246 (1980). https://doi.org/10.1137/0717021
15. Feldman, M.: Hilbert Transform Applications in Mechanical Vibration. John Wiley & Sons, Chichester (2011). https://doi.org/10.1002/9781119991656
16. Adafruit Industries: QT Py ESP32-S3 – Product documentation, https://www.adafruit.com/product/5426, last accessed 2025/07/31
17. STMicroelectronics: LSM6DSOX – iNEMO inertial module: always-on 3D accelerometer and 3D gyroscope with Machine Learning Core. Datasheet, https://www.st.com/resource/en/datasheet/lsm6dsox.pdf, last accessed 2025/07/31
18. Jolliffe, I.T.: Principal Component Analysis. 2nd edn. Springer, New York (2002). https://doi.org/10.1007/b98835
19. Savitzky, A., Golay, M.J.E.: Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry **36**(8), 1627–1639 (1964). https://doi.org/10.1021/ac60214a047
20. Wellue: FS20F Bluetooth Fingertip Pulse Oximeter – Product specifications, https://de.getwellue.com/Seiten/fs20f-Bluetooth-Fingeroximeter, last accessed 2026/07/1