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) derived from inertial measurement units (IMUs) may enable wearable heart-rate (HR) monitoring. This study evaluated filtering strategies, HR-detection methods, acceleration-derived candidate signals, and 20 anatomical sensor positions during seated rest in five participants, using a synchronized photoplethysmography-based pulse oximeter reference.
Hilbert-Envelope processing substantially outperformed Direct Bandpass filtering, reducing mean MAE from 17.00 to 4.50 BPM. After quality control, FFT achieved lower MAE than peak detection (3.16 vs. 5.46 BPM), but approximately 40% of FFT windows were rejected, whereas peak detection retained more than 99.9%. Candidate-signal selection had a negligible practical effect ($\eta^2 = 0.021$); the vector magnitude performed best (2.26 ± 5.36 BPM). Body position also showed a small overall effect ($\eta^2 = 0.0358$). Position 16, below the scapula, achieved the lowest mean MAE (1.52 BPM), while Position 13 at the triceps was the best limb placement (mean MAE = 2.37 BPM). These findings support Hilbert-Envelope processing and trunk placement for IMU-based HR estimation during seated rest, although validation in larger, more diverse populations and during motion remains necessary.

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 acceleration-derived candidate signal (a_x, a_y, a_z, magnitude $|a|$, or PCA1) yields the lowest HR-estimation error under the pipeline selected in RQ1–RQ2?
RQ4: Which body position achieves the highest HR detection accuracy?

By answering these questions through mixed 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.

Participants and Protocol Five male participants, all aged 22 years, were measured while seated at rest. Their BMI ranged from 18.6 to 27.8, and their resting HR ranged from 74 to 95 BPM (Table 1). Participation was voluntary, and all participants provided informed written consent. No participant had a dependent relationship with the researchers, and withdrawal was possible at any time.

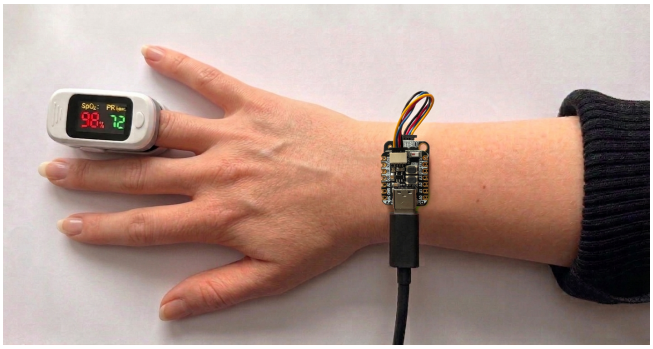


Fig. 1: Showing the apparatus (ESP32-S3 QT Py microcontroller) mounted on top of the LSM6DSOX IMU sensor breakoutboard, which is double-sided-taped to the user's wrist. A pulse oximeter provides the ground truth heart rate.

Each of the 20 body positions was recorded for 120 s, resulting in $5 \times 20 \times 120 = 12{,}000$ s of recording time. The positions were measured sequentially in numerical order from P1 to P20; the order was not randomized. The IMU sensor was fixed at each position using adhesive tape. Depending on the participant and measurement position, the sensor was attached either directly to the skin or over 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 for accelerometer placement (Fig. 2 and 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. Positions are identified by unique numbers (1–20) marked as red points in Fig. 2. The anatomical name and body surface location (anterior or posterior) for each position are provided in Table 2 to facilitate identification and reproducibility.

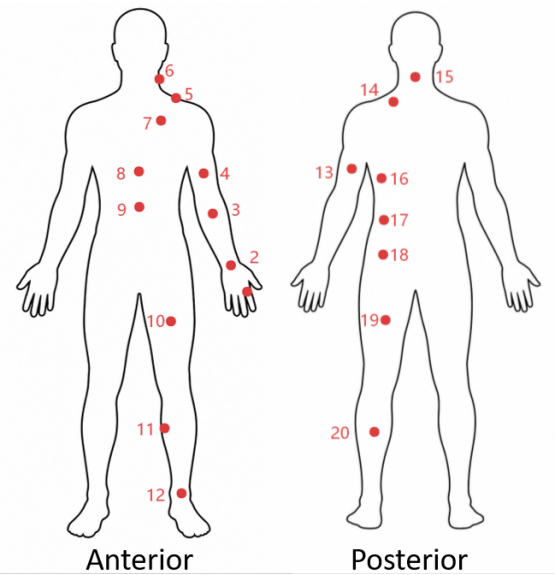


Fig. 2: Accelerometer measurement positions on the anterior (left) and posterior (right) body surface. Numbers (red points, 1–20) correspond to the positions listed in Table 2.

Table 2: Measurement positions for accelerometer placement, corresponding to the numbered positions shown in Fig. 2.

(a) Anterior positions

No.	Position
1	Index finger
2	Wrist
3	Forearm
4	Upper arm
5	Upper trapezius
6	Lateral neck
7	Upper left chest
8	Lower sternum
9	Upper abdomen
10	Anterior mid-thigh
11	Medial lower leg
12	Ankle

(b) Posterior positions

No.	Position
13	Triceps region
14	Posterior trapezius
15	Posterior neck
16	Below scapula
17	Lower back
18	Post. superior iliac spine
19	Posterior mid-thigh
20	Posterior mid-calf

2.2 Signal Processing Pipeline

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

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. Both IMU and pulse oximeter data were time-synchronized using the common PC-clock timestamp, with the IMU resampled to uniform 200 Hz via PCHIP interpolation to account for USB transmission jitter.

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]. The candidates are shown in Fig. 3 as normalized raw signals and filtered using a Hilbert envelope filtering strategy. Each candidate signal

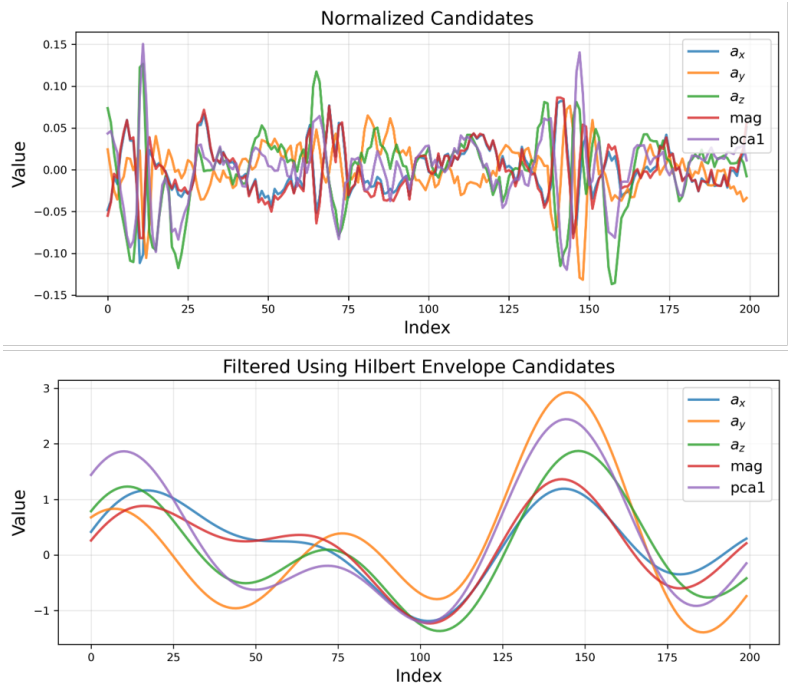


Fig. 3: 5 signal candidates a_x, a_y, a_z, magnitude, and pca1) normalized by mean subtraction (top) and filtered using a Hilbert envelope filtering strategy (bottom)

underwent three preprocessing steps: (1) NaN values were filled via linear interpolation with forward/backward edge fill to handle <2-sample gaps from timestamp cleaning; (2) outliers beyond ±4 robust standard deviations (median ± $4\times MedianAbsoluteDeviation$) were clipped to prevent bias; and (3) linear detrending via least-squares fit was applied to remove DC offset and sensor drift.

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 using a 4th-order Butterworth filter (zero-phase) 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 smoothed envelope is bandpass-filtered a second time (4th-order Butterworth, 0.75–3.0 Hz) and smoothed with a Savitzky-Golay filter (polynomial order 3, window 0.15 s).

This two-stage approach isolates cardiac mechanical impulses (4–35 Hz) before extracting their rhythmic modulation (0.75–3.0 Hz), providing higher spectral specificity than direct bandpass filtering.

Fig. 4 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.

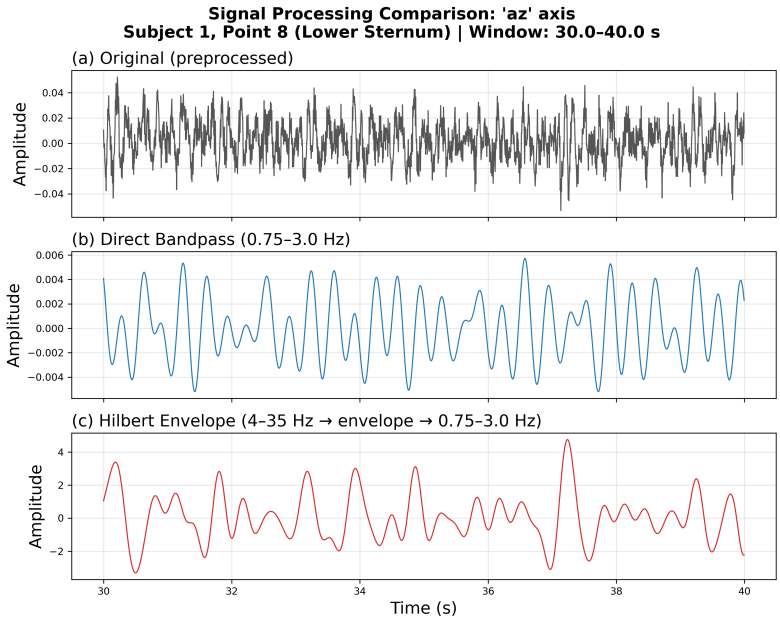


Fig. 4: Signal processing comparison for the a_z axis (Subject 1, Point 8, 10-second segment). (a) Preprocessed signal (NaN interpolation, outlier clipping at ±4 SD, and linear detrending), (b) Direct Bandpass, (c) Hilbert Envelope.

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.

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 filter strategies × 2 polarities). Two HR detection shown in Fig. 5 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 two components: (i) spectral signal-to-noise ratio (SNR; 60% weight), normalized by saturation at SNR = 10 (peak = 10× median); and (ii) spectral dominance (40% weight), defined as the peak-to-second-peak ratio, normalized by saturation at dominance = 5.

Peak Method: Peaks are detected using adaptive thresholds: minimum inter-peak distance 0.4 s (≤150 BPM), minimum prominence $0.3 \times$SD, minimum height mean + $0.2 \times$ 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 $\text{HR} = 60/\text{median(IBI)}$.

Quality is scored from three components: (i) interval regularity (50% weight), defined as $1.0 - 2.0 \times \text{CV(IBI)}$ where CV is the coefficient of variation; (ii) temporal coverage (25% weight), the span of the signal covered by detected peaks relative to window duration; and (iii) peak count adequacy (25% weight), defined as $1.0 - |\text{detected peaks/expected peaks} - 1.0|$, where expected peaks are inferred from the reference HR.

2.3 Dataset size

Theoretically, a 120-second recording segmented with 20-second windows and 5-second step size yields:

$$n_{\text{expected}} = \left\lfloor \frac{T - W}{\Delta} \right\rfloor + 1 = \left\lfloor \frac{120 - 20}{5} \right\rfloor + 1 = 21 \text{ windows per recording.} \quad (1)$$

However, actual window counts differed between filtering strategies. During "Direct Bandpass", an auto-settle algorithm is applied to remove sensor settling transients that occur immediately after sensor placement on the body. When a wearable IMU is first attached, hand movement, contact adjustment, and cable settling produce large-amplitude acceleration spikes that are unrelated to cardiac motion. These transients corrupt the first few analysis windows and produce

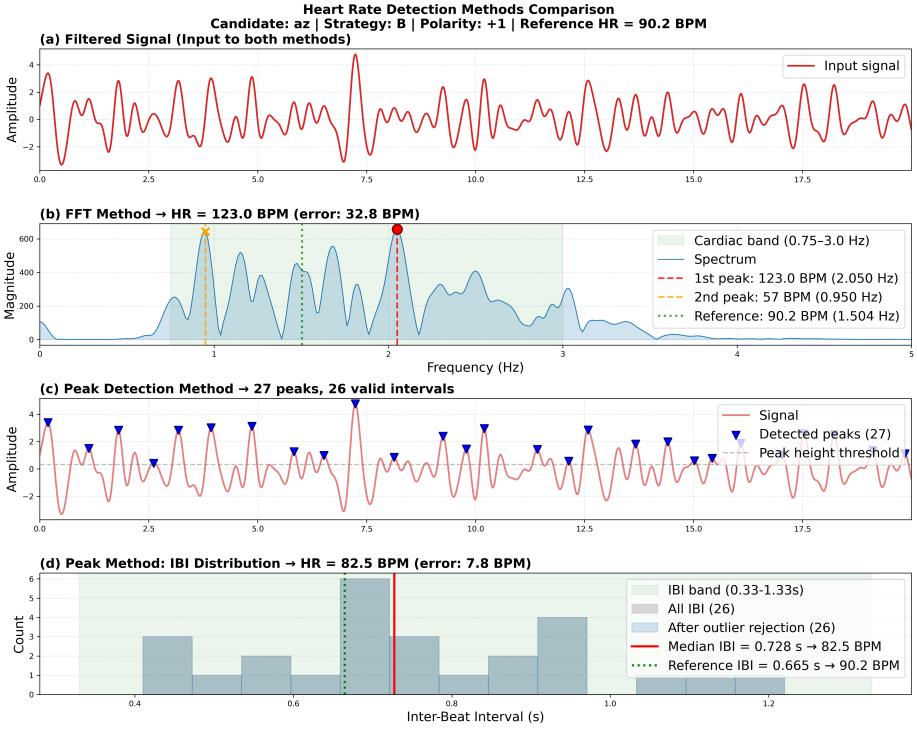


Fig. 5: ignal processing comparison for the a_z axis (Subject 1, Point 8, 10-second segment). (a) Preprocessed signal (NaN interpolation, outlier clipping at ±4 SD, and linear detrending), (b) Direct Bandpass, (c) Hilbert Envelope.

false HR estimates. To mitigate this, the auto-settle procedure automatically discards initial signal segments with high RMS energy (threshold: 3× median RMS, maximum discard: 30 s). An average of 1.24 ± 0.45 seconds per candidate signal was trimmed, reducing effective signal duration from 119.1 s to 117.85 s and yielding 19.9 ± 0.5 windows per recording per candidate.

"Hilbert Envelope" strategy applies no auto-settle procedure, retaining the full signal duration of 119.09 ± 0.3 s and yielding 20.0 ± 0.2 windows per recording per candidate, because the envelope-demodulation process naturally suppresses transient artifacts through its two-stage bandpass filtering and smoothing operations.

Each window was paired with the corresponding pulse-oximeter heart rate value from the nearest OXI timestamp. Fewer than 0.2% of windows lacked valid OXI data (caused by rare sensor disconnection during measurement), resulting in a pre-detection dataset of 19,928 valid (window, OXI reference) pairs.

Both methods employ distinct quality metrics with a common threshold of 0.3. Peak detection retained nearly all segmented windows: 19,876 (Direct Bandpass) and 20,026 (Hilbert Envelope) valid windows pre-filtering. After quality

filtering, retention remained nearly perfect: 19,876 and 20,025 windows respectively, representing $> 99.9\%$ retention regardless of filtering strategy.

The FFT method window count was doubled pre-filtering, 19,888 and 20,040 for Direct Bandpass and Hilbert Envelope, respectively, because both signal polarities (± 1) of each candidate were analyzed. However, quality filtering substantially reduced FFT retention: only 9,316 (Direct Bandpass, 59.8%) and 14,406 (Hilbert Envelope, 81.3%) windows remained valid. This 40-percentage-point difference between strategies reflects FFT's susceptibility to harmonic confusion in suboptimal spectral conditions, while peak detection proved robust across signal qualities.

Mean quality scores ($mean \pm SD$) differed markedly: 0.473 ± 0.18 of FFT vs. 0.740 ± 0.11 of Peak.

Quality filtering (threshold ≥ 0.3) fundamentally altered the interpretation of HR detection method superiority. Before filtering, Peak detection in combination with the Hilbert Envelop filtering strategy appeared superior (Peak Detection: 5.46 ± 8.49 BPM vs. FFT 7.97 ± 16.43 BPM). After filtering, FFT achieved substantially lower error (3.16 ± 9.69 BPM), reversing the ranking by 2.30 BPM. This comparison is shown in Figure 6.

The final dataset comprises 63,623 valid per-window observations.

Method	Direct Bandpass	Hilbert Envelop
FFT	35.05 ± 33.65	7.97 ± 16.43
Peak	8.33 ± 10.20	5.46 ± 8.49

(a) Before Quality Check

Method	Direct Bandpass	Hilbert Envelop
FFT	35.50 ± 36.59	3.16 ± 9.69
Peak	8.33 ± 10.20	5.46 ± 8.49

(b) After Quality Check

Fig. 6: MAE ± SD by HR Detection Method and Filtering Strategy, showing method ranking reversal after quality filtering.

2.4 Evaluation

The analyses were conducted sequentially. RQ1 compared filtering strategies across all configurations. RQ2 compared FFT and peak detection within the better-performing Hilbert Envelope strategy. RQ3 compared candidate signals using the all four combinations of filtering strategies and heart rate detection methods. RQ4 compared anatomical sensor locations using the same pipeline after pooling candidate signals, as RQ3 showed no practically meaningful overall candidate-signal effect.

For each configuration, Mean Absolute Error (MAE $= |HR_{est} - HR_{ref}|$) is computed.

Statistical Analysis. Given the large number of window-level observations, even small differences may reach statistical significance. Effect-size measures were therefore used to distinguish statistical from practical significance [12].

All research questions were evaluated using mixed-effects models with the investigated factor as a fixed effect and Subject as a random intercept. The models report a standardized z-value for the fixed-effect test and its corresponding p-value. The intraclass correlation coefficient (ICC) was reported to quantify the proportion of MAE variance attributable to differences between subjects and to assess the robustness of the results across individuals.

Eta-squared (η^2). Eta-squared describes the proportion of variance explained by the grouping factor. The adopted thresholds were small (0.01), medium (0.06), and large (0.14) [12].

Cohen's d. Cohen's d describes the standardized difference in MAE between two groups. The adopted thresholds were negligible (< 0.2), small (0.2–0.5), medium (0.5–0.8), and large (> 0.8). A positive value indicates higher MAE for the compared group and therefore worse performance.

For RQ1 and RQ2, Cohen's d was calculated for the corresponding two-group comparisons. For the multi-group comparisons in RQ3 and RQ4, pairwise Cohen's d values were calculated relative to the best-performing reference group rather than across all possible pairs. Magnitude was used as the reference candidate for RQ3, and Position 16 (below the scapula) as the reference position for RQ4. Thus, a positive d indicates that the compared candidate or position has higher MAE than the respective best-performing reference.

The dependent variable is Mean Absolute Error ($\text{MAE} = |HR_{\text{est}} - HR_{\text{ref}}|$). Independent variables are: (RQ1) Filtering Strategy: Direct Bandpass vs. Hilbert Envelope; (RQ2) HR Detection Method FFT vs. Peak (for Hilbert filtering strategy only); (RQ3) 5 Candidate Signal: a_x, a_y, a_z, magnitude, PCA1 (for FFT HR detection method and Hilbert filtering strategy only); (RQ4) Body Position: 20 anatomical locations of sensor (for FFT HR detection method and Hilbert filtering strategy only).

3 Results

3.1 RQ1: Which filtering strategy yields the best HR estimation accuracy?

Table 3 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 (63,623 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.

Mixed ANOVA Results. A mixed-effects linear model with Strategy as fixed effect and Subject as random intercept revealed highly significant differences ($z = -85.39$, $p < 10^{-5}$). The Hilbert Envelope reduced mean MAE by

Table 3: Comparison of filtering strategies across all subjects and positions.

Metric	Direct	Bandpass Hilbert Envelope	Cohen's d
Mean MAE (BPM)	17.00	4.50	+0.67 (medium)
Median MAE (BPM)	5.33	1.76	—
SD MAE (BPM)	25.66	9.08	—

12.50 BPM (Cohen's $d = +0.67$, medium effect, $\eta^2 = 0.138$; 13.8% of variance explained). The effect was consistent across all 20 anatomical positions, with per-point Cohen's d ranging from $d = +0.49$ (small, Point 8) to $d = +0.94$ (large, Point 11); 15 positions showed medium effects ($0.5 \leq d < 0.8$) and five large effects ($d \geq 0.8$).

Subject Heterogeneity. The intraclass correlation coefficient (ICC = 0.057) indicated that individual subject differences accounted for 5.7% of the total variance in MAE, with the remaining 94.3% attributed to algorithmic factors (filtering strategy, detection method, anatomical position, candidate signal, and measurement noise). This low ICC suggests that filtering strategy effects generalize robustly across individuals.

3.2 RQ2: Which HR detection method (FFT or peak-based) performs best?

To isolate the effect of detection method from filtering strategy, this analysis uses only the Hilbert Envelope pipeline. Fig. 7 summarizes the comparison of both HR detection methods with and without quality check. The number of observations for each combination is also shown in Fig. 7.

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

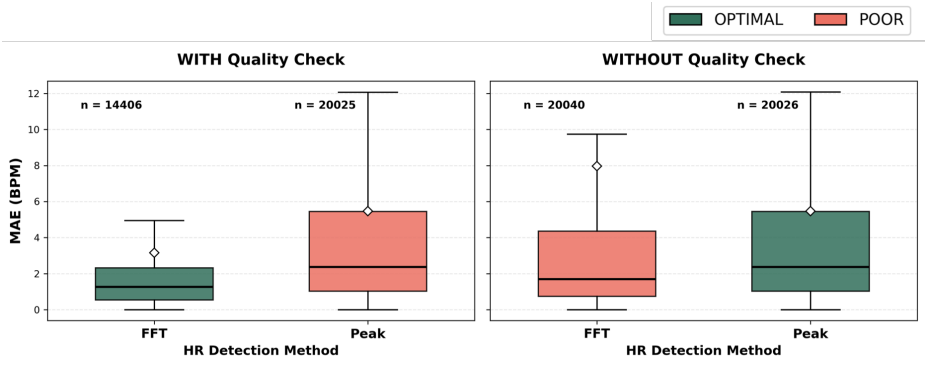


Fig. 7: Comparison of HR detection methods with and without quality check (Hilbert Envelope only).

Mixed ANOVA Results. Results were analysed both before and after applying the common quality-score threshold of 0.3. The corresponding results are summarized in Table 4.

Table 4: Comparison of FFT and peak-based HR detection with and without quality control using the Hilbert Envelope pipeline.

Metric	**With quality control**	**Without quality control**
Observations	34,431	40,066
Mixed-model z-value	18.19	-20.165
p-value	$< 10^{-5}$	$< 10^{-5}$
Cohen's d	-0.26	$+0.19$
η^2	0.1169	0.1032
ICC	0.138	0.116

Results with quality control. After applying the quality-score threshold, FFT achieved a lower mean MAE than peak detection (3.16 vs. 5.46 BPM), corresponding to an improvement of 2.30 BPM. The difference was statistically significant ($z = 18.19$, $p < 10^{-5}$), although its practical magnitude was small (Cohen's $d = -0.26$). The detection method explained 11.69% of the variance in MAE ($\eta^2 = 0.1169$, medium effect). FFT also achieved a lower median MAE (1.26 vs. 2.37 BPM). However, its error distribution remained less stable, with a higher standard deviation than peak detection (9.69 vs. 8.49 BPM).

The lower FFT error was obtained only after discarding low-quality windows. Approximately 40% of FFT windows failed the quality criterion, primarily because of spectral ambiguity and harmonic misidentification. In contrast, peak detection retained virtually all windows, with only 0.003% rejected by the quality check. Consequently, the post-filtering comparison reflects the accuracy of FFT on its retained windows rather than its performance across all recorded windows.

Results without quality control. Before quality filtering, the ranking was reversed. Peak detection achieved a lower mean MAE than FFT (5.46 vs. 7.97 BPM), as well as a lower standard deviation (8.49 vs. 16.43 BPM). The difference was statistically significant ($z = -20.165$, $p < 10^{-5}$), but the practical effect was negligible (Cohen's $d = +0.19$). The detection method explained 10.32% of the variance ($\eta^2 = 0.1032$), while the subject-level contribution was slightly lower than in the quality-controlled analysis (ICC = 0.116 vs. 0.138).

Overall, FFT provides the highest accuracy when low-quality windows can be identified and excluded. Nevertheless, this advantage comes at the cost of substantial data loss and increased variability. Peak detection provides slightly lower average accuracy but produces valid estimates for nearly all windows and maintains more consistent error across signal-quality conditions. Therefore, FFT is the preferred method when a reliable quality assessment and rejection mechanism are available, whereas peak detection is the more robust choice for continuous real-world monitoring.

3.3 RQ3: Which acceleration-derived candidate signal yields the lowest HR-estimation error under the pipeline selected in RQ1–RQ2?

Candidate-signal selection was evaluated using the Hilbert Envelope and FFT pipeline after quality filtering. The analysis included five candidate signals and 14,406 valid window-level observations. Table 5 summarizes the results.

Candidate-signal comparison. The vector magnitude achieved the lowest MAE (2.26 ± 5.36 BPM), whereas a_x showed the highest (3.63 ± 11.59 BPM). The difference of mean values is only 1.37 BPM.

Relative to the magnitude baseline, the a_z candidate did not differ significantly ($p = 0.071$). The remaining candidates showed statistically significant differences ($p < 0.001$); however, the corresponding pairwise effect sizes were negligible, with Cohen's d values ranging from 0.065 to 0.161. This indicates that the statistically significant differences are unlikely to have meaningful practical consequences.

Table 5: Comparison of acceleration-derived candidate signals using the Hilbert Envelope–FFT pipeline after quality filtering (N = 14,406, ICC = 0.071, $\eta^2 = 0.021$). The z and p-values are calculated relative to the magnitude signal baseline $|a|$.

Rank	Candidate	Mean ± SD MAE (BPM)	z	p
1	$\|a\|$	2.26 ± 5.36	baseline	–
2	a_z	2.71 ± 8.19	1.806	0.071
3	PCA1	3.61 ± 10.66	5.476	< 0.001
4	a_y	3.63 ± 11.30	5.309	< 0.001
5	a_x	3.63 ± 11.59	5.420	< 0.001

Mixed ANOVA Results. A mixed-effects model with Candidate as a fixed effect and Subject as a random intercept showed that candidate choice explained only a small proportion of the variance in MAE ($\eta^2 = 0.021$, or 2.1%). The intraclass correlation coefficient (ICC = 0.071) indicated that only 7.1% of the total variance was attributable to differences between subjects. Therefore, the candidate-signal effect was relatively consistent across individuals and practically small.

Although magnitude achieved the lowest MAE, candidate differences were negligible ($\eta^2 = 0.021$; pairwise $d = 0.065$–0.161). Therefore, all five candidates from the quality-controlled Hilbert Envelope–FFT pipeline (N = 14,406) were pooled for RQ4.

3.4 RQ4: Which body position achieves the highest HR detection accuracy?

Body-position performance was evaluated using the quality-controlled Hilbert Envelope–FFT pipeline, pooling all five candidate signals (N = 14,406 obser-

vations from five subjects). Position differences ranged from nonsignificant to highly significant ($z = 1.24$–12.707). However, the overall practical effect was small ($\eta^2 = 0.0358$), indicating that body position explained only 3.58% of the variance and had a smaller influence than the filtering strategy or HR detection method. The low intraclass correlation coefficient (ICC = 0.071) suggests that the position-related differences were robust across individuals.

Position 16, located below the scapula, achieved the lowest mean MAE (1.52 BPM) and was therefore the best-performing IMU placement. The pairwise Cohen's d values ranged from 0.144 to 0.5074, representing negligible to medium effects. Positions 16, 17, and 18 showed similar errors and consequently negligible pairwise differences. In contrast, comparisons between Position 16 and less favourable locations, such as Position 6, produced medium effects. Thus, although the overall position effect was small, specific anatomical comparisons showed practically meaningful differences.

The position-wise MAE ranking and spatial distribution are shown in Fig. 8 and Fig. 9 Overall, below-scapula placement provided the most accurate HR estimates, while the results also demonstrate that the relative performance of individual body locations can vary considerably.

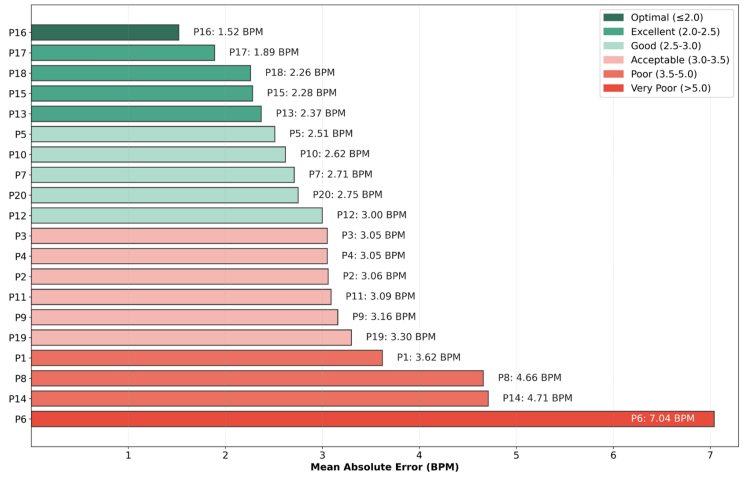


Fig. 8: Mean MAE ranking across the 20 body positions.

3.5 Inter-Subject Variability

Inter-subject variability was evaluated using the quality-controlled Hilbert Envelope–FFT pipeline with all five candidate signals pooled. The analysis included $N = 14{,}406$ retained observations across five subjects. Results are summarized in Table 6.

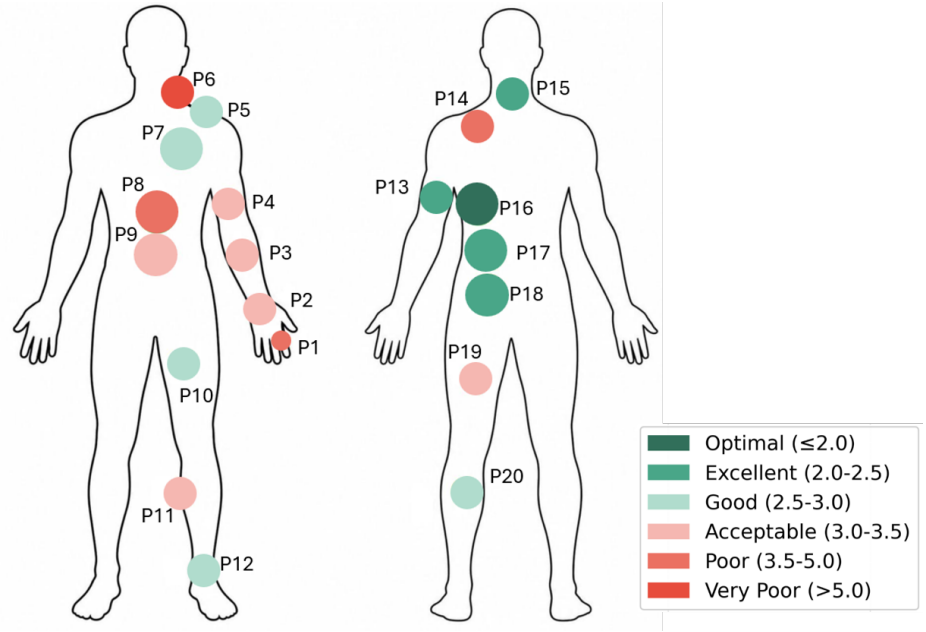


Fig. 9: Spatial distribution of HR-estimation MAE across the body (the color legend is the same as in Fig. 8.

Mean MAE ranged from 1.90 to 5.69 BPM. Subject 3 achieved the highest accuracy (1.90 ± 3.78 BPM, whereas Subject 2 showed the highest error and variability (5.69 ± 14.47 BPM. Subject 4 also showed relatively high error and variability (4.37 ± 12.97 BPM). Subjects 2 and 4 contributed the fewest retained observations (1,810 and 2,818, respectively), suggesting that more of their windows failed the FFT quality criterion.

The best-performing position varied across subjects: P16 for Subject 1, P8 for Subject 2, P13 for Subject 3, P7 for Subject 4, and P18 for Subject 5. Notably, P8 was the best position for Subject 2 but the worst for Subject 4, demonstrating a subject–position interaction.

Candidate preferences were more consistent: the magnitude signal was best for four subjects, while a_z was best for Subject 1. The worst candidate varied between a_x, a_y, and PCA1.

4 Discussion

4.1 Key Findings

RQ1: Filtering strategy. The Hilbert Envelope substantially outperformed Direct Bandpass filtering, reducing mean MAE from 17.00 to 4.50 BPM. This difference was statistically significant and practically relevant ($d = +0.67$ in the

Table 6: Per-subject HR-estimation performance using the quality-controlled Hilbert Envelope–FFT pipeline with all five candidate signals pooled. N denotes the number of retained observations after quality filtering.

Subject	N	MAE (mean ± SD)	Best pos.	Worst pos.	Best cand.	Worst cand.
1	3,148	2.85 ± 9.23	P16 (0.88)	P6 (15.57)	a_z	a_x
2	1,810	5.69 ± 14.47	P8 (1.77)	P19 (31.54)	$\|a\|$	PCA1
3	3,462	1.90 ± 3.78	P13 (0.94)	P3 (7.18)	$\|a\|$	a_y
4	2,818	4.37 ± 12.97	P7 (1.16)	P8 (16.09)	$\|a\|$	a_y
5	3,168	2.30 ± 7.05	P18 (0.78)	P20 (7.54)	$\|a\|$	a_x

pooled comparison; $\eta^2 = 0.138$). The result was consistent across body positions, indicating that separating the cardiac vibration band from its low-frequency modulation improves the robustness of IMU-based HR estimation.

RQ2: HR detection method. After quality control, FFT achieved lower mean MAE than peak detection (3.16 vs. 5.46 BPM), but the practical effect was small ($d = -0.26$). This accuracy advantage was accompanied by substantially higher variability and the rejection of many low-quality FFT windows. Therefore, FFT provides the best accuracy on retained windows, whereas peak detection is the more reliable option when continuous estimates and high data retention are required.

RQ3: Candidate signal. The vector magnitude produced the lowest error (2.26 ± 5.36 BPM), while a_x produced the highest (3.63 ± 11.59 BPM). However, the overall candidate-signal effect was small ($\eta^2 = 0.021$), and pairwise effects were negligible ($d = 0.065$–0.161). Consequently, no single candidate was selected for the position analysis, and all five candidates were pooled.

RQ4: Body position. Body position had a statistically significant but small overall effect ($\eta^2 = 0.0358$). Position 16, below the scapula, achieved the lowest mean MAE and was the best-performing placement. Pairwise Cohen's d values ranged from 0.144 to 0.5074, corresponding to negligible to medium effects. Positions 16, 17, and 18 showed similar performance, whereas comparisons with less favourable locations produced larger differences. Thus, placement remains practically relevant even though its overall variance contribution was smaller than that of filtering or HR detection.

Position 13 (triceps) was the best-performing limb placement and a practical alternative when trunk placement was not feasible, possibly because its mechanical coupling to the thorax preserves cardiac vibrations better than distal locations.

Reference uncertainty. The Wellue FS20F pulse oximeter reports an accuracy of approximately ±1–2 BPM under resting conditions [20]. Consequently, the reported IMU errors represent deviations from the PPG reference rather than absolute deviations from the true HR. Reference uncertainty may therefore contribute to the observed MAE, particularly for results close to 1–2 BPM. Differences within this range should be interpreted cautiously, and future studies should validate the proposed method against a clinical-grade ECG reference.

Inter-Subject Variability and Practical Implications. Inter-subject variability was evaluated using the quality-controlled Hilbert Envelope–FFT pipeline with all five candidate signals pooled ($N = 14{,}406$). Subject 3 achieved the lowest MAE (1.90 ± 3.78 BPM), despite having the highest BMI (27.76). This finding does not support a simple relationship between BMI, presumed tissue thickness, and HR-estimation accuracy. Individual anatomy, sensor coupling, signal quality, and placement may be more influential.

Subject 2 and Subject 4 showed the highest errors (5.69 ± 14.47 and 4.37 ± 12.97 BPM, respectively) and contributed the fewest retained observations ($N = 1{,}810$ and $N = 2{,}818$). Their lower retention may reflect a higher proportion of windows failing the FFT quality criterion, although missing or invalid recordings may also have contributed. These two subjects also had the lowest resting HRs in the cohort (74.1 and 81.7 BPM). Lower HR may increase spectral ambiguity and harmonic confusion, but this possible explanation could not be confirmed with the present data. The best-performing position varied across subjects, indicating an interaction between anatomy and sensor placement.

An additional source of variability was the non-standardized sensor attachment. Depending on the measurement position, the exposed sensor board was placed directly on the skin or over a single layer of clothing. Direct skin contact may have caused discomfort, skin friction, or involuntary movements, leading to variable contact pressure and sensor coupling. This may have contributed to the higher errors observed at the lateral neck (P6) and distal upper-limb positions (P1–P4), although the effects of attachment method and anatomical position cannot be separated in the present data.

Because only five subjects were included, these findings are exploratory and do not support stable conclusions about the effects of resting HR, BMI, or individual anatomy.

4.2 Comparison with Prior Work

Table 7 places our results in context with both IMU-based and PPG-based wearable studies.

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).

Table 7: 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	P16: MAE 1.52 BPM
PPG-based benchmark				
Nelson and Allen [10]	Wrist	Apple Watch 3	MAE 5.24 BPM; MAPE 7.21%	P13: MAE 2.37 BPM; MAPE 4.51%
	Wrist	Fitbit Charge 2	MAE 5.93 BPM; MAPE 6.93%	P13: MAE 2.37 BPM; MAPE 4.51%
Bent et al. [11]	Wrist	Consumer PPG	MAE 7.2 BPM	P13: MAE 2.37 BPM

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

This suggests that IMU-based HR detection can be competitive with PPG-based wearables under low-motion conditions.

4.3 Limitations and Future Work

Several limitations should be considered. The study included only five healthy 22-year-old men, limiting generalizability to women, other age groups, different body compositions, and clinical populations. The subject-specific findings should therefore be interpreted as exploratory. Future studies should validate the pipeline in larger and more diverse cohorts, including participants with bradycardia, tachycardia, arrhythmias, and cardiovascular disease.

Each position was recorded for only 120 s during a single seated-rest session. This duration may be insufficient to capture fully stabilized resting HR and does not reflect performance during movement, exercise, or postural changes. The reference HR range was also narrow (74.1–94.5 BPM). Longer recordings, repeated sessions, and controlled motion protocols are needed to assess test–retest reliability and performance across a wider physiological range.

Sensor attachment was not fully standardized. Depending on the position, the exposed IMU sensor board was placed directly on the skin or over a single layer of clothing rather than being enclosed in a standardized wearable housing. Direct contact with bare skin may have caused discomfort, skin friction, or involuntary movements, resulting in variable contact pressure and sensor coupling. This may have contributed to the lower accuracy observed at the lateral neck (P6) and distal positions (P1–P4), although the effects of attachment method and anatomical position cannot be separated in the present dataset.

The study evaluated a classical, fixed signal-processing pipeline and did not compare it with adaptive machine-learning or deep-learning methods. Although

this approach is interpretable and computationally efficient, it may be less robust to inter-subject, positional, and motion-related variability. Future work should evaluate adaptive data-driven methods using larger datasets and subject-independent validation.

Finally, the proposed physiological explanation for the favourable triceps performance remains hypothetical. Simultaneous IMU measurements with vascular ultrasound or local pulse-wave recordings could clarify the contribution of vascular transmission. In addition, the use of overlapping windows reduces the independence of observations; future analyses should therefore include participant-level summaries, non-overlapping validation segments, or hierarchical models that explicitly account for window dependence.

5 Conclusion

This study evaluated IMU-based HR estimation across filtering strategies, detection methods, candidate signals, and 20 body positions. The Hilbert Envelope substantially outperformed Direct Bandpass filtering, reducing mean MAE from 17.00 to 4.50 BPM. After quality control, FFT achieved the lowest mean MAE (3.16 BPM), but approximately 40% of its windows were rejected. Peak detection retained more than 99.9% of windows and was therefore the more reliable choice for continuous monitoring.

Candidate-signal selection had little practical effect ($\eta^2 = 0.021$); the magnitude signal performed best (2.26 ± 5.36 BPM), but all candidates were retained for the position analysis. Body position had a small overall effect ($\eta^2 = 0.0358$). Position 16, below the scapula, achieved the lowest mean MAE (1.52 BPM), while Position 13 at the triceps was the best-performing limb placement and a practical alternative when trunk placement is not feasible.

Overall, accurate IMU-based HR estimation is feasible during seated rest, particularly with Hilbert Envelope processing and trunk placement. However, the results are based on only five homogeneous participants and short, stationary recordings. Larger and more diverse studies with standardized sensor attachment, repeated measurements, and motion or exercise protocols are required before clinical or real-world deployment.

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