Beyond Motion Artifacts: Optimizing PPG Preprocessing for Accurate Pulse Rate Variability Estimation

Yuna Watanabe
Northeastern University
Boston, MA, USA
watanabe.y@northeastern.edu

Natasha Yamane
Northeastern University
Boston, MA, USA
yamane.n@northeastern.edu

Aarti Sathyanarayana
Northeastern University
Boston, MA, USA
a.sathyanarayana@northeastern.edu

Varun Mishra*
Northeastern University
Boston, MA, USA
v.mishra@northeastern.edu

Matthew S. Goodwin*
Northeastern University
Boston, MA, USA
m.goodwin@northeastern.edu

Abstract

Wearable physiological monitors are ubiquitous, and photoplethysmography (PPG) is the standard low-cost sensor for measuring cardiac activity. Metrics such as inter-beat interval (IBI) and pulse-rate variability (PRV)—core markers of stress, anxiety, and other mental-health outcomes—are routinely extracted from PPG, yet preprocessing remains non-standardized. Prior work has focused on removing motion artifacts; however, our preliminary analysis reveals sizeable beat-detection errors even in low-motion data, implying artifact removal alone may not guarantee accurate IBI and PRV estimation. We therefore investigate how band-pass cutoff frequencies affect beat-detection accuracy and whether optimal settings depend on specific persons and tasks observed. We demonstrate that a fixed filter produces substantial errors, whereas the best cutoffs differ markedly across individuals and contexts. Further, tuning cutoffs per person and task raised beat-location accuracy by up to 7.15% and reduced IBI and PRV errors by as much as 35 ms and 145 ms, respectively, relative to the fixed filter. These findings expose a long-overlooked limitation of fixed band-pass filters and highlight the potential of adaptive, signal-specific preprocessing to improve the accuracy and validity of PPG-based mental-health measures.

CCS Concepts

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Keywords

Physiological signal processing; Wearable sensing; Photoplethysmography; Mental health

ACM Reference Format:
Yuna Watanabe, Natasha Yamane, Aarti Sathyanarayana, Varun Mishra, and Matthew S. Goodwin. 2025. Beyond Motion Artifacts: Optimizing PPG Preprocessing for Accurate Pulse Rate Variability Estimation. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '25), October 12–16, 2025, Espoo, Finland.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3714394.3756241

*These authors jointly supervised this work and share senior authorship.

UbiComp Companion '25, Espoo, Finland

ACM ISBN 979-8-4007-1477-1/2025/10
https://doi.org/10.1145/3714394.3756241

1 Introduction

Recent advances in wearable peripheral sensors allow researchers to continuously and non-invasively monitor physiological states. Photoplethysmography (PPG) has emerged as a popular method for monitoring cardiac activity because of its affordability, portability, and ease of integration in consumer devices. In the context of mental and behavioral health, PPG signals have been leveraged for myriad purposes, including, but not limited to, stress detection [2, 27–29, 36, 37, 43], depression monitoring [38], aggression onset prediction [13], and detecting craving for substances [8, 12]. These analyses often rely on inter-beat interval (IBI) and pulse rate variability (PRV) that reflect autonomic nervous system regulatory activity. Accurate IBI and PRV estimates are therefore critical—any bias at this stage propagates to downstream models, undermining prediction accuracy and validity in mental-health applications.

Despite its widespread use, PPG signal preprocessing remains non-standardized [10]. Although most pipelines begin with a band-pass filter to remove low- and high-frequency noise unrelated to cardiac activity, the choice of cutoff frequencies varies widely across studies. Prior work has attempted to identify optimal cutoffs [22, 25, 26, 40], with some proposing 0.5 to 15.0 Hz as a suitable, universal passband [40]; however, researchers have not reached a consensus. Further, optimal filtering parameters depend on individual factors (e.g., age, baseline heart rate), sensor placement, and the nature of the task being recorded [5, 40]. Given this complexity, researchers are often left to select filtering configurations that best fit their datasets, which can be arbitrary and inconsistent. Some work has investigated more adaptive denoising techniques instead of, or in addition to, applying a fixed filter, including least-mean-squares (LMS) filtering [30, 32], wavelet denoising [15], and deep learning-based approaches [3, 21, 41]. However, these methods are typically designed only to reduce motion artifacts and may not generalize well to other sources of PPG signal distortion [18]. These issues raise intertwined questions: Is adaptively denoising motion artifacts enough to yield accurate IBI and PRV estimates? If not, how much does parameterization in band-pass filtering matter, and can any single fixed filter suffice?

This paper aims to evaluate whether a one-size-fits-all band-pass filter yield accurate IBI and PRV estimates, and, if not, quantify how much person-, activity-, and mental-state–specific optimization of cutoff frequencies improves accuracy. In pursuing these goals, our work generates three key insights:

- Beyond physical activity, PPG signals contain inherent noise influenced by the individual, their specific activities, and mental states.
- Applying a fixed band-pass filter can introduce substantial error in beat detection and resulting IBI and PRV estimates.
- Choosing activity- and state-specific cutoff frequencies for each individual significantly reduces beat-detection errors and improves IBI and PRV estimation.

Our findings suggest that preprocessing pipelines should be tailored to each individual and recording context to obtain accurate IBI and PRV estimates, beyond denoising motion artifacts. Doing so can enhance the validity of physiological metrics used in mental and behavioral health monitoring and lay the groundwork for more personalized and accurate assessment tools.

2 Background & Related Work

2.1 Preprocessing and filtering strategies

Researchers routinely leverage PPG signals to identify various affective states, including stress and depression, commonly deriving metrics like IBI (or pulse rate) and PRV. Because filtering strategies depend on the specific metric of interest, we focus on preprocessing methods to extract IBI or PRV from PPG signals.

Although considerable variability in filtering configurations exists, band-pass filtering is the most commonly used preprocessing technique [9, 29, 36, 38, 43]. Low-cut frequencies used in prior work range from 0 Hz (i.e., no low-pass filtering) to 0.9 Hz, while high-cut frequencies span 1.6 to 35 Hz.

In addition to band-pass filtering, some researchers apply moving averages [36] and wavelets [38], while others opt for Savitzky–Golay filters [2]. While a few studies briefly explain their rationale for filtering (e.g., to remove non-cardiac signals) [29, 36, 38, 43], most do not justify their choice of filtering methods or cutoff frequencies, suggesting a lack of consensus on optimal preprocessing practices. This observation reinforces our conjecture that filtering parameters are often chosen heuristically rather than following standardized guidelines [18].

Given the lack of standardized preprocessing practices, several researchers have attempted to identify optimal filtering strategies. For example, Liang et al. [22] compared seven filtering algorithms with various filter orders applied to short-duration PPG signals and found that a 4th-order Chebyshev Type-II filter yielded the best improvement in signal quality indices. Other studies have focused on determining the optimal cutoff frequencies for band-pass filters (Table 1). Despite these efforts, different studies report wide variation in recommended frequency ranges, with no clear agreement.

2.2 Adaptive denoising techniques

In addition to using fixed band-pass filters, researchers have explored various adaptive techniques to denoise PPG signals. Classic approaches include Fourier series analysis [33] and wavelet denoising [15], which decompose motion-corrupted signals into frequency components, remove noise, and reconstruct cleaner PPG waveforms. Sparse signal reconstruction has also been used to isolate relevant signal components while discarding noise [44].

Another widely studied method involves LMS-based adaptive filters [30, 32] that optimize filter coefficients to minimize error between the estimated clean and observed signal. While some LMS-based approaches work without additional sensors by reconstructing noise from PPG itself [32], most rely on reference noise sources such as accelerometer data. Independent component analysis provides an alternative that separates underlying signal sources without requiring additional sensors [17, 20, 30].

Recent work has applied machine learning-based denoising methods, including generative models [21], time-delay neural networks [41], and autoencoder architectures [3] that aim to learn signal-cleaning functions directly from data.

Although these methods show promise in mitigating motion artifacts, they primarily focus on motion-related noise. In our preliminary analysis (Section 4.1), we explore whether motion artifacts are a single noise source that impacts beat-detection accuracy.

3 Methods

To evaluate the impacts of filtering parameters on beat-detection accuracy, we apply different combinations of cutoffs and compare their performance. We use two datasets to account for variability between different experimental configurations.

3.1 Datasets

We used the Wearable Stress and Affect Detection (WESAD) and Stress-free datasets, which include ECG and PPG signals collected simultaneously. This allows us to use ECG-derived beats as ground truth when evaluating the validity of PPG-derived beats.

WESAD dataset: The publicly available WESAD dataset was developed for stress and affect detection [35]. It contains ECGs collected with the RespiBAN chest sensor and PPGs collected with the Empatica E4 wristband. Fifteen participants performed four tasks to induce different affective states: baseline (20 minutes), amusement (5 minutes), stress (10 minutes), and meditation (7 minutes $\times$ 2). Although the data includes transient periods during which participants completed self-reports, we only used data from the four tasks.

Stress-free dataset: The Stress-free dataset was collected to advance stress detection [1]. The dataset contains beats from ECGs collected with the Polar H10 and PPGs collected with the Empatica E4 wristband. Thirty-five participants underwent three stressful tasks after a 10-minute baseline rest period: mental arithmetic, startle response, and a cold pressor task, each lasting 4 minutes. Participants were given a 5-minute rest period between each task, wherein they remained seated without engaging in any activity; however, we use only the baseline rest data collected at the beginning and the three stress-inducing tasks in our analyses.

Table 1: Summary of recommended low-cut and high-cut frequencies in prior work.

Publication	Low-cut	High-cut	Type of filter	Evaluation
Cassani et al. [5]	0.6 Hz (adult); 1.0 Hz (child)	3.3 Hz (adult); 2.7 Hz (child)	Bandpass filter	Signal-to-noise ratio
Karolcik and Georgious [16]	0.35 Hz	10 Hz	4th-order Chebyshev Type-II	Signal quality indices
Mejía-Mejía et al. [25]	Around 0 Hz	-	4th-order Butterworth filter	PRV accuracy
Wolling et al. [40]	0.5 Hz	15 Hz		Peak displacement
Mejía-Mejía and Kyriacou [26]	Around 0 Hz	20 Hz	FIR and IIR filters	PRV accuracy

3.2 Initial data preprocessing

We used the heartview package, a Python package for preprocessing signals [42]. We preprocessed WESAD ECG signals using an elliptic band-pass filter (1–15 Hz) and Manikandan and Soman's [24] beat-detection algorithm. Following Liang et al. [22], we preprocessed PPG data with a 4th-order Chebyshev Type-II filter. We filtered PPGs using 525 parameter combinations: low-cut frequencies ($f_{C,\text{low}}$; 0.4–1.7 Hz) and high-cut frequencies ($f_{C,\text{high}}$; 1.2–5.0 Hz), both in 0.1-Hz increments, excluding combinations where $f_{C,\text{low}}$ was larger than $f_{C,\text{high}}$. We used HeartPy's [39] algorithm to detect beats in PPGs, defining each beat location as the middle-amplitude point between foot and apex to reduce IBI and PRV estimation errors [31].

3.3 Metrics for evaluation

We evaluate three accuracy metrics on PPG signals filtered with our 525 unique filter combinations: (1) beat location accuracy (compared to ECG peaks); (2) IBI accuracy (compared to ECG-derived R-R intervals); and (3) PRV accuracy (compared to ECG-derived heart rate variability [HRV]). We then visualize how accuracy metrics vary with filtering parameters and examine differences in optimal settings across persons and tasks.

Beat location accuracy: Beat location accuracy measures how closely PPG beat locations match those of ECG beats. Following Charlton et al. [7], we label a PPG beat "correct" when it occurs within 150 ms of the nearest ECG beat. Owing to potential misalignments in sampling rate, device clocks, and waveform shape, Charlton et al. searched offsets between PPGs and ECGs up to ±10 s. Visual inspection showed our recordings drifted far less, so we limited our lag search to ±2 s in 20 ms steps and retained offsets that maximized matches. We then evaluate improvements in sensitivity (Se), positive predictive value (PPV), and F1 score (F1). Each score was calculated using the following formulas:

$$Se = \frac{n_{\text{correct}}}{n_{\text{ECG}}} \times 100, \quad PPV = \frac{n_{\text{correct}}}{n_{\text{PPG}}} \times 100, \quad F1 = \frac{2 \times PPV \times Se}{PPV + Se}$$

where n_{correct} is the number of correct PPG peaks, n_{ECG} is the total number of ECG-derived peaks, and n_{PPG} is the total number of PPG-derived peaks. A high F1 score indicates that the selected filtering parameters yield optimal accuracy in beat detection.

IBI accuracy: IBI accuracy measures how closely PPG-derived IBIs match ground-truth ECG-derived R-R intervals. We segmented the signals into 60-second windows and removed artifactual beats using the FLIRT package to avoid skewing mean IBI estimates [11]. For segments with more than 10 valid beats, we computed the average IBI and the difference between R-R intervals and IBIs. We then computed the mean absolute error (MAE) of intervals per participant and task. Lower MAE indicates higher IBI accuracy, reflecting better beat detection.

PRV accuracy: As with IBI, we assessed PRV accuracy by comparing PPG-derived PRV with ECG-derived HRV. We used RMSSD (root mean square of successive IBI differences) as the metric for variability. Following the same procedure for IBI accuracy, we divided signals into 60-second segments, removed artifactual beats, and calculated RMSSD. We then calculated the difference between ECG- and PPG-derived values and computed the MAE per participant and task. Lower MAE indicates higher PRV accuracy.

To further evaluate the potential impact of signal-by-signal adaptive preprocessing, we identify optimal filter settings at two levels: (1) across all tasks and participants ($\mathbf{F}_{\text{global}}$) and (2) per participant within each task ($\mathbf{F}_{\text{pt}}$). We ran multi-objective optimization at each level using the NSGA-II algorithm in the 'pymoo' package [4] with average F1 score, MAE IBI, and MAE RMSSD as the multiple objectives. NSGA-II may return multiple solutions for which none is strictly better in all three objectives. We chose the single best filter by performing min-max normalization for each objective and minimizing the combined metric $-\text{F1}_{\text{norm}} + \text{MAE}_{\text{IBI,norm}} + \text{MAE}_{\text{RMSSD,norm}}$. As a baseline, we also apply a fixed non-optimized band-pass filter (0.5–4.0Hz; $\mathbf{F}_{\text{base}}$), as these values align with commonly recommended ranges (Section 2.1). We then compare the resulting average performance (F1, MAE IBI, and MAE RMSSD) across the three filter settings ($\mathbf{F}_{\text{base}}$, $\mathbf{F}_{\text{global}}$, and $\mathbf{F}_{\text{pt}}$). Finally, we compare the distribution of mean IBI and RMSSD against the ECG reference across $\mathbf{F}_{\text{base}}$, $\mathbf{F}_{\text{global}}$, and $\mathbf{F}_{\text{pt}}$ using a repeated-measures ANOVA ($\alpha = .95$) followed by Bonferroni-corrected pairwise t-tests.

4 Findings and Evaluation

4.1 The non-"motion" artifacts

We first ran a preliminary correlation analysis to test whether motion artifacts fully explain beat-detection errors. We preprocessed PPG signals with a fixed 0.5–4 Hz band-pass filter, segmented the signal into 60-second windows, and computed beat-detection accuracy metrics (Section 3.3). We calculated Monitor Independent Movement Summary unit (MIMS-unit)—a metric capturing human motion from accelerometer signals [14]—and used its Area-Under-the-Curve (AUC_{MIMS}) to quantify motion artifact level.

We show the correlation between motion level (AUC_{MIMS}) and beat-detection accuracy for four representative low-motion tasks in Figure 1. While some other tasks show moderate-to-strong correlations, low-motion tasks reveal many 60-second segments with small AUC_{MIMS} yet poor accuracy (low F1; high MAE IBI and RMSSD). This finding suggests that beat-detection errors can persist even

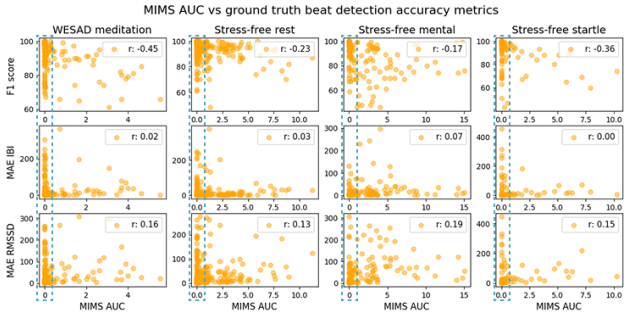

Figure 1: Correlations between motion artifacts ($\mathrm{AUC}_{\mathrm{MIMS}}$) and beat-detection accuracy (F1 score, MAE IBI & RMSSD).

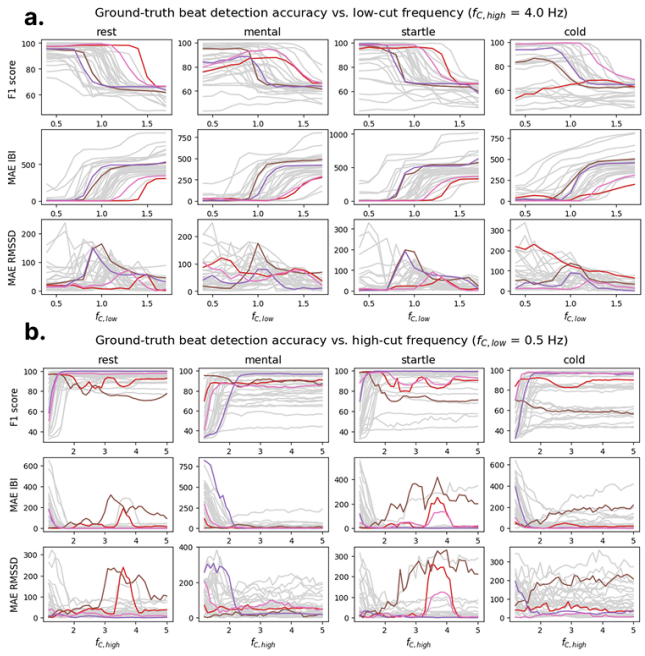

Figure 2: F1 score and MAE IBI and RMSSD changes across varying (a) low-cut and (b) high-cut frequencies in the band-pass filter for the Stress-free dataset. Each column represents a task, and each line represents a single participant's metric.

when motion artifacts are minimal, motivating our search for a more flexible filtering strategy.

4.2 Evaluating different parameters

We present F1 scores and MAE IBI and RMSSD across signal segments at varying $f_{C,\mathrm{low}}$ values for the Stress-free dataset in Figure 2a. The choice of $f_{C,\mathrm{low}}$ impacted every metric in both datasets. F1 scores drop, and MAE IBI and RMSSD increase as $f_{C,\mathrm{low}}$ varied between 0.5 and 1.5 Hz (i.e., the typical human heart rate range); however, these frequencies vary across individuals. For example, with $f_{C,\mathrm{high}}$ of 4.0 Hz, participants represented by pink and red

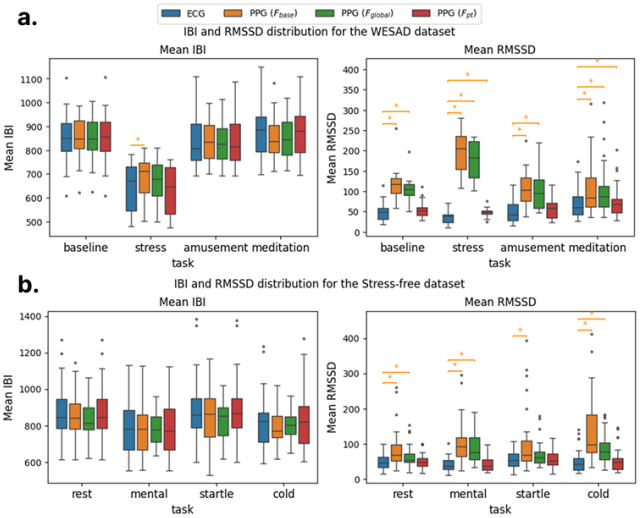

Figure 3: Distribution of mean IBI and RMSSD derived from ECGs and PPGs preprocessed with three types of filters for (a) the WESAD and (b) the Stress-free datasets. Brackets with asterisks indicate significant pairwise differences.

lines exhibited peak F1 scores around a $f_{C,\mathrm{low}}$ of 1.2 Hz, while participants represented by brown and purple lines peaked around 0.8 Hz. Although some participants showed stable F1 scores and MAE at lower $f_{C,\mathrm{low}}$ (0.5–1.5 Hz), others were sensitive to slight $f_{C,\mathrm{low}}$ changes (e.g., pink and red lines in the mental stress panels).

The choice of $f_{C,\mathrm{high}}$ also affected beat-detection accuracy (Figure 2b), though less than $f_{C,\mathrm{low}}$. Selecting a lower $f_{C,\mathrm{high}}$ (below 2 Hz) generally lowered F1 scores and raised MAE IBI and RMSSD; however, higher $f_{C,\mathrm{high}}$ impaired performance for some participants and tasks (e.g., red and brown lines in the rest and startle response panels). The optimal $f_{C,\mathrm{high}}$ also varied across participants and tasks. While Figure 2a-b only show cases where $f_{C,\mathrm{high}}$ = 4.0Hz and $f_{C,\mathrm{low}}$ = 0.5Hz, respectively, varying these parameters also affected the graph waveform, indicating a complex relationship between the them. These results suggest that an arbitrary choice may significantly impact beat-detection accuracy for some participants and tasks, indicating the need for individual- and task-specific filter parameterization. The WESAD dataset produced similar trends.

Table 2 reports *mean* performance across participants; in both datasets, $\mathbf{F}_{\mathrm{pt}}$ achieved the highest F1 scores and the lowest IBI and RMSSD errors. Relative to $\mathbf{F}_{\mathrm{base}}$, the largest mean gains were a **7.15%** in F1 (WESAD stress), a **35 ms** reduction in MAE IBI (Stress-Free cold), and a **145 ms** reduction in MAE RMSSD (WESAD stress). Comparing within *individuals* reveals still larger benefits: F1 improved by up to **28.8%** (median 2.50%, IQR 0.35–6.73%), while MAE IBI and MAE RMSSD decreased by as much as **328 ms** (median 2.96 ms, IQR 0.15–14.9 ms) and **329 ms** (median 26.5 ms, IQR 8.65–73.1 ms), respectively. We also show the mean IBI and RMSSD distribution in Figure 3. Repeated-measures ANOVA revealed a significant mean IBI difference between ECGs and PPGs only in WESAD stress; other tasks showed none, suggesting that using generic or fixed filtering thresholds may be sufficient for estimating

Table 2: Mean F1 score and MAE in IBI and RMSSD with different types of cutoff frequency optimization. Optimization per task and person produced the smallest errors in IBI and RMSSD estimations.

Dataset	Task	Fixed ($\mathbf{F}_{\text{base}}$)			Optimized across task & person ($\mathbf{F}_{\text{global}}$)			Optimized per task & person ($\mathbf{F}_{\text{pt}}$)		
		Mean F1	MAE IBI	MAE RMSSD	Mean F1	MAE IBI	MAE RMSSD	Mean F1	MAE IBI	MAE RMSSD
WESAD	baseline	69.95	17.94	68.10	71.12	15.47	55.31	**71.72**	**4.470**	**6.380**
	stress	60.96	43.36	158.7	62.64	42.49	139.98	**68.11**	**15.97**	**13.34**
	amusement	84.54	24.38	59.60	85.37	23.00	46.64	**87.05**	**3.655**	**8.729**
	meditation	90.00	29.30	40.58	90.36	28.32	34.03	**92.01**	**4.830**	**6.409**
Stress-free	cold	77.21	46.91	85.88	77.92	63.87	39.41	**83.02**	**11.95**	**8.188**
	mental	83.23	20.30	61.80	82.17	41.77	44.78	**89.03**	**3.497**	**5.128**
	rest	93.31	16.20	36.31	94.13	37.38	18.18	**97.25**	**1.156**	**2.011**
	startle	89.65	28.88	44.84	89.62	51.21	20.25	**93.94**	**2.835**	**3.031**

aggregated/high-level metrics about heart rate or IBI. In contrast, mean RMSSD differed across all tasks, indicating that fixed thresholds introduce substantial errors when measuring variability—a key indicator of autonomic nervous system activity and variable of interest for a variety of mental and behavioral health outcomes. Specifically, mean RMSSD derived from the PPGs preprocessed with $\mathbf{F}_{\text{base}}$ and $\mathbf{F}_{\text{global}}$ showed significantly different distributions compared to ECGs in almost all tasks, with medium-to-large effect sizes (Cohen's d=0.420–4.04). PPG signals preprocessed with $\mathbf{F}_{\text{pt}}$ also showed a significant RMSSD difference in the WESAD stress task, p = .0007, and meditation task, p = .0074. However, effect sizes were markedly smaller than for $\mathbf{F}_{\text{base}}$ and $\mathbf{F}_{\text{global}}$: stress task—Cohen's d = 4.04 ($\mathbf{F}_{\text{base}}$), 3.90 ($\mathbf{F}_{\text{global}}$), and 0.96 ($\mathbf{F}_{\text{pt}}$); meditation task—$d$ = 0.71 ($\mathbf{F}_{\text{base}}$), 0.62 ($\mathbf{F}_{\text{global}}$), and 0.11 ($\mathbf{F}_{\text{pt}}$). These results demonstrate that fixed cutoffs lead to substantial errors, and signal-by-signal filtering optimization can significantly enhance the accuracy and validity of beat detection and beat variability estimation.

5 Discussion

5.1 Impact of cutoff frequencies on RMSSD

In our analyses (Section 4.2), although F1 scores and MAE IBI exhibited inverse trends, MAE RMSSD followed a distinct pattern: it increased at specific low-cut frequencies and then decreased again. This behavior may be explained by the nature of the PPG waveform, which often exhibits two prominent peaks per cardiac cycle—one corresponding to the systolic wave and the other to the diastolic wave. As the low-cut frequency increases, low-frequency components are attenuated, and higher-frequency components are emphasized. Consequently, the beat-detection algorithm may erroneously identify both systolic and diastolic peaks as separate beats, doubling the beat count. Interestingly, this consistent misidentification introduces a form of regularity in the detected IBIs, resulting in RMSSD values that appear more stable and, in some cases, closer to those from correctly identified beats. This contrasts with intermediate low-cut frequencies, where beat detection is less consistent, leading to greater IBI variability and larger RMSSD errors.

5.2 Factors influencing optimal cutoffs

Heart rate is one of the most apparent factors affecting optimal cutoff frequencies, which vary with age, lung capacity, and activity level. If heart rate alone could determine optimal filtering parameters, applying frequency cutoffs well outside the plausible heart rate range would preserve signal quality. However, our results demonstrate that this is not always the case—some PPG signals remain highly sensitive to cutoff frequency settings, and broader passbands do not always yield accurate beat detection.

Additional factors may include stress and its physiological effects. Prior studies have shown that stress can alter PPG waveform morphology [6, 34]. Mental stress induces vasodilation in the forearm [23], affecting PPG amplitude. Respiration also modulates PPG amplitude, wherein inhalation reduces it, and exhalation amplifies it. These morphological changes may influence the effectiveness of different filters. Indeed, we observed that PPG signals collected during the mental arithmetic and cold pressor tasks in the Stress-free dataset and the stress task in the WESAD dataset were more sensitive to low-cut frequencies than those collected during rest. This suggests that physiological changes induced by the person's activities and mental states may impact optimal configurations.

A limitation of the present work is that we employed a single beat-detection algorithm, yet choice of algorithm may influence optimal cutoff frequencies. Some detectors are likely more tolerant of small shifts in filter settings, whereas others require carefully tuned parameters. We also only examined data from Empatica E4 wristbands collected in the lab. We believe that similar patterns can be observed in free-living data as well, once motion artifacts are removed. However, different devices may apply different preprocessing to raw data, and different on-body placement locations may exhibit distinct waveform patterns. Future work should investigate how different algorithms, devices, and sensor placements interact with preprocessing choices across in-lab and in-the-wild data.

5.3 The need for adaptive filtering

Overall, our findings suggest that cutoff frequencies significantly impact beat-detection accuracy when motion artifacts are negligible, and the optimal configuration may vary across individuals, activities, and mental states. This dependence has been largely overlooked: applying a conventional fixed filter can distort IBI and RMSSD in a context- and person-specific manner. Echoing electrodermal work showing that adaptive thresholding enhances sensitivity [19], these observations call for adaptive preprocessing for accurate pulse rate variability estimation. During model development, researchers may be able to manually verify that a fixed

filter behaves sensibly; however, once deployed in the field, they cannot inspect every new signal. If the filter systematically underestimates or overestimates PRV under stress for a user, for example, the model may misclassify that state. Tuning the filter to each signal—reflecting the user's physiology and situation—avoids systematic errors that would otherwise propagate through the pipeline, degrading the accuracy and validity of downstream mental- and behavioral-health models.

In this study, we tuned filter parameters using reference ECGs. However, such ground-truth signals are often unavailable in real-world applications. Future work should develop task- and person-aware optimization methods that operate without ECG references.

6 Conclusion

We investigate how non-motion noise in PPG signals degrades beat-detection accuracy and how filter settings affect that error. We demonstrate that PPG contains noise beyond physical movement, so motion-artifact removal alone may not ensure accurate and valid IBI and PRV estimates. We also demonstrate that using a single band-pass filter for all individuals, activities, and mental states produces substantial errors, whereas tailoring cutoffs to the person and context consistently improves metric validity across datasets and tasks. These results suggest the need for adaptive, signal-by-signal preprocessing: optimizing filter parameters for each recording can deliver more accurate IBI, RMSSD, and related physiological markers in mental- and behavioral-health research and real-world applications.

Acknowledgements

This research is supported by the National Institutes of Health, under award numbers P30DA029926 and 5R01LM014191-03, and the National Science Foundation, under IIS-2442593. ChatGPT was used to correct grammatical errors and improve flow.

References

[1] Ohida Binte Amin, Varun Mishra, Tinashe M. Tapera, Robert Volpe, and Aarti Sathyanarayana. 2025. Extending stress detection reproducibility to consumer wearable sensors. https://doi.org/10.48550/arXiv.2505.05694 arXiv:2505.05694 [cs].

[2] Aamir Arsalan and Muhammad Majid. 2021. Human stress classification during public speaking using physiological signals. *Computers in Biology and Medicine* 133 (June 2021), 104377. https://doi.org/10.1016/j.compbiomed.2021.104377

[3] Joseph Azar, Abdallah Makhoul, Raphaël Couturier, and Jacques Demerjian. 2021. Deep recurrent neural network-based autoencoder for photoplethysmogram artifacts filtering. *Computers & Electrical Engineering* 92 (June 2021), 107065. https://doi.org/10.1016/j.compeleceng.2021.107065

[4] Julian Blank and Kalyanmoy Deb. 2020. Pymoo: Multi-objective optimization in Python. *IEEE Access* 8 (2020), 89497–89509. https://doi.org/10.1109/ACCESS.2020.2990567

[5] Raymundo Cassani, Abhishek Tiwari, and Tiago H Falk. 2020. Optimal filter characterization for photoplethysmography-based pulse rate and pulse power spectrum estimation. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2020 (July 2020), 914–917. https://doi.org/10.1109/EMBC44109.2020.9175396

[6] Patrick Celka, Peter H. Charlton, Bushra Farukh, Philip Chowienczyk, and Jordi Alastruey. 2019. Influence of mental stress on the pulse wave features of photoplethysmograms. *Healthcare Technology Letters* 7, 1 (Nov. 2019), 7–12. https://doi.org/10.1049/htl.2019.0001

[7] Peter H Charlton, Kristjan Pilt, and Panicos A Kyriacou. 2022. Establishing best practices in photoplethysmography signal acquisition and processing. *Physiological Measurement* 43, 5 (May 2022), 050301. https://doi.org/10.1088/1361-6579/ac6cc4

[8] Keerthi Kumar Chintha, Premananda Indic, Brittany Chapman, Edward W Boyer, and Stephanie Carreiro. 2018. Wearable biosensors to evaluate recurrent opioid toxicity after naloxone administration: A Hilbert transform approach. *Proceedings of the Annual Hawaii International Conference on System Sciences. Annual Hawaii International Conference on System Sciences* 2018 (Jan. 2018), 3247–3252. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5782803/

[9] Margot Deviaene, Jesús Lázaro, Dorien Huysmans, Dries Testelmans, Bertien Buyse, Sabine Van Huffel, and Carolina Varon. 2018. Sleep apnea detection using pulse photoplethysmography. In *2018 Computing in Cardiology Conference (CinC)*, Vol. 45. IEEE, Piscataway, NJ, USA, 1–4. https://doi.org/10.22489/CinC.2018.134 ISSN: 2325-887X.

[10] Jessilyn Dunn, Varun Mishra, Md Mobashir Hasan Shandhi, Hayoung Jeong, Natasha Yamane, Yuna Watanabe, Bill Chen, and Matthew S. Goodwin. 2025. Building an open-source community to enhance autonomic nervous system signal analysis: DBDP-autonomic. *Frontiers in Digital Health* 6 (Jan. 2025), 1467424. https://doi.org/10.3389/fdgth.2024.1467424

[11] Simon Föll, Martin Maritsch, Federica Spinola, Varun Mishra, Filipe Barata, Tobias Kowatsch, Elgar Fleisch, and Felix Wortmann. 2021. FLIRT: A feature generation toolkit for wearable data. *Comput. Methods Programs Biomed.* 212 (Nov. 2021), 106461. https://doi.org/10.1016/j.cmpb.2021.106461

[12] August F. Holtyn, Eugene Bosworth, Lisa A. Marsch, Bethany McLeman, Andrea Meier, Elizabeth C. Saunders, Emre Ertin, Md Azim Ullah, Shahin Alan Samiei, Monowar Hossain, Santosh Kumar, Kenzie L. Preston, Massoud Vahabzadeh, Dikla Shmueli-Blumberg, Julia Collins, Jennifer McCormack, and Udi E. Ghitza. 2019. Towards detecting cocaine use using smartwatches in the NIDA clinical trials network: Design, rationale, and methodology. *Contemporary Clinical Trials Communications* 15 (Sept. 2019), 100392. https://doi.org/10.1016/j.conctc.2019.100392

[13] Tales Imbiriba, Ahmet Demirkaya, Ashutosh Singh, Deniz Erdogmus, and Matthew S Goodwin. 2023. Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism. *JAMA Netw Open* 6, 12 (Dec. 2023), e2348898. https://doi.org/10.1001/jamanetworkopen.2023.48898

[14] Dinesh John, Qu Tang, Fahd Albinali, and Stephen Intille. 2019. An open-source Monitor-Independent Movement Summary for accelerometer data processing. *Journal for the measurement of physical behaviour* 2, 4 (Dec. 2019), 268–281. https://doi.org/10.1123/jmpb.2018-0068

[15] Greeshma Joseph, Almaria Joseph, Geevarghese Titus, Rintu Mariya Thomas, and Dency Jose. 2014. Photoplethysmogram (PPG) signal analysis and wavelet de-noising. In *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD)*. IEEE, Piscataway, NJ, USA, 1–5. https://doi.org/10.1109/AICERA.2014.6908199

[16] Stefan Karolcik and Pantelis Georgiou. 2024. Optimal filtering and smoothing thresholds for high-frequency photoplethysmography signals. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, Piscataway, NJ, USA, 1–5. https://doi.org/10.1109/iscas58744.2024.10558515

[17] Byung S Kim and Sun K Yoo. 2006. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng.* 53, 3 (March 2006), 566–568. https://doi.org/10.1109/TBME.2005.869784

[18] Ian R Kleckner, Mallory J Feldman, Matthew S Goodwin, and Karen S Quigley. 2021. Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behav. Res. Methods* 53, 2 (April 2021), 518–535. https://doi.org/10.3758/s13428-020-01438-9

[19] Ian R Kleckner, Jolie B Wormwood, Rebecca M Jones, Eva Culakova, Lisa Feldman Barrett, Catherine Lord, Karen S Quigley, and Matthew S Goodwin. 2024. Adaptive thresholding increases sensitivity to detect changes in the rate of skin conductance responses to psychologically arousing stimuli in both laboratory and ambulatory settings. *Int. J. Psychophysiol.* 196 (Feb. 2024), 112280. https://doi.org/10.1016/j.ijpsycho.2023.112280

[20] Rajet Krishnan, Balasubramaniam Natarajan, and Steve Warren. 2010. Two-stage approach for detection and reduction of motion artifacts in photoplethysmographic data. *IEEE Transactions on Biomedical Engineering* 57, 8 (Aug. 2010), 1867–1876. https://doi.org/10.1109/TBME.2009.2039568

[21] Ju Hyeok Kwon, So Eui Kim, Na Hye Kim, Eui Chul Lee, and Jee Hang Lee. 2022. Preeminently robust neural PPG denoiser. *Sensors* 22, 6 (Jan. 2022), 2082. https://doi.org/10.3390/s22062082

[22] Yongbo Liang, Mohamed Elgendi, Zhencheng Chen, and Rabab Ward. 2018. An optimal filter for short photoplethysmogram signals. *Sci. Data* 5, 1 (May 2018), 180076. https://doi.org/10.1038/sdata.2018.76

[23] M Lindqvist, S Davidsson, P Hjemdahl, and A Melcher. 1996. Sustained forearm vasodilation in humans during mental stress is not neurogenically mediated. *Acta Physiol. Scand.* 158, 1 (Sept. 1996), 7–14. https://doi.org/10.1046/j.1365-201X.1996.529288000.x

[24] M Sabarimalai Manikandan and K P Soman. 2012. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed. Signal Process. Control* 7, 2 (March 2012), 118–128. https://doi.org/10.1016/j.bspc.2011.03.004

[25] Elisa Mejia-Mejia, James M May, and Panayiotis A Kyriacou. 2021. Effect of filtering of photoplethysmography signals in pulse rate variability analysis. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2021 (Nov. 2021), 5500–5503. https://doi.org/10.1109/EMBC46164.2021.9629521

[26] Elisa Mejía-Mejía and Panicos A Kyriacou. 2023. Effects of noise and filtering strategies on the extraction of pulse rate variability from photoplethysmograms.

Biomed. Signal Process. Control 80, 104291 (Feb. 2023), 104291. https://doi.org/10.1016/j.bspc.2022.104291

[27] Varun Mishra, Tian Hao, Si Sun, Kimberly N. Walter, Marion J. Ball, Ching-Hua Chen, and Xinxin Zhu. 2018. Investigating the role of context in perceived stress eetection in the wild. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1708–1716. https://doi.org/10.1145/3267305.3267537

[28] Varun Mishra, Sougata Sen, Grace Chen, Tian Hao, Jeffrey Rogers, Ching-Hua Chen, and David Kotz. 2020. Evaluating the reproducibility of physiological stress detection models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4 (Dec. 2020), 1–29. https://doi.org/10.1145/3432220 Publisher: Association for Computing Machinery (ACM).

[29] Jongwoo Park, Jongsu Kim, and Sung-Phil Kim. 2018. A study on the development of a day-to-day mental stress monitoring system using personal physiological data. In *2018 18th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, Piscataway, NJ, USA, 900–903. https://ieeexplore.ieee.org/document/8571973

[30] Fulai Peng, Zhengbo Zhang, Xiaoming Gou, Hongyun Liu, and Weidong Wang. 2014. Motion artifact removal from photoplethysmographic signals by combining temporally constrained independent component analysis and adaptive filter. *Biomed. Eng. Online* 13, 1 (April 2014), 50. https://doi.org/10.1186/1475-925X-13-50

[31] Elena Peralta, Jesus Lazaro, Raquel Bailon, Vaidotas Marozas, and Eduardo Gil. 2019. Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals. *Physiol. Meas.* 40, 2 (Feb. 2019), 025007. https://doi.org/10.1088/1361-6579/ab009b

[32] M. Raghu Ram, K. Venu Madhav, E. Hari Krishna, Nagarjuna Reddy Komalla, and K. Ashoka Reddy. 2012. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Transactions on Instrumentation and Measurement* 61, 5 (May 2012), 1445–1457. https://doi.org/10.1109/TIM.2011.2175832

[33] K. Ashoka Reddy, Boby George, and V. Jagadeesh Kumar. 2009. Use of Fourier series analysis for motion artifact reduction and data compression of photoplethysmographic signals. *IEEE Transactions on Instrumentation and Measurement* 58, 5 (May 2009), 1706–1711. https://doi.org/10.1109/TIM.2008.2009136

[34] Mantas Rinkevičius, Spyridon Kontaxis, Eduardo Gil, Raquel Bailón, Jesus Lazaro, Pablo Laguna, and Vaidotas Marozas. 2019. Photoplethysmogram signal morphology-based stress assessment. In *2019 Computing in Cardiology (CinC)*. IEEE, Piscataway, NJ, USA, Page 1–Page 4. https://doi.org/10.22489/cinc.2019.126

[35] Philip Schmidt, Attila Reiss, Robert Duerichen, Claus Marberger, and Kristof Van Laerhoven. 2018. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In *ICMI '18*. Association for Computing Machinery, New York, NY, USA, 400–408. https://doi.org/10.1145/3242969.3242985

[36] Ali Tazarv, Sina Labbaf, Stephanie M. Reich, Nikil Dutt, Amir M. Rahmani, and Marco Levorato. 2021. Personalized stress monitoring using wearable sensors in everyday settings. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Piscataway, NJ, USA, 7332–7335. https://doi.org/10.1109/EMBC46164.2021.9630224 ISSN: 2694-0604.

[37] Kobiljon Toshnazarov, Uichin Lee, Byung Hyung Kim, Varun Mishra, Lismer Andres Caceres Najarro, and Youngtae Noh. 2024. SOSW: Stress sensing with off-the-shelf smartwatches in the wild. *IEEE Internet of Things Journal* 11, 12 (June 2024), 21527–21545. https://doi.org/10.1109/JIOT.2024.3375299

[38] Batbayar Unursaikhan, Nobuaki Tanaka, Guanghao Sun, Sadao Watanabe, Masako Yoshii, Kazuki Funahashi, Fumihiro Sekimoto, Fumiaki Hayashibara, Yutaka Yoshizawa, Lodoiravsal Choimaa, and Takemi Matsui. 2021. Development of a novel web camera-based contact-free major depressive disorder screening system using autonomic nervous responses induced by a mental task and its clinical application. *Frontiers in Physiology* 12, Article 642986 (May 2021), 10 pages. https://doi.org/10.3389/fphys.2021.642986

[39] Paul Van Gent, Haneen Farah, Nicole Van Nes, and Bart Van Arem. 2019. HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation Research Part F: Traffic Psychology and Behaviour* 66 (Oct. 2019), 368–378. https://doi.org/10.1016/j.trf.2019.09.015

[40] Florian Wolling, Sudam Maduranga Wasala, and Kristof Van Laerhoven. 2021. Optimal preprocessing of raw signals from reflective mode photoplethysmography in wearable devices. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2021 (Nov. 2021), 1157–1163. https://doi.org/10.1109/EMBC46164.2021.9630955

[41] Ke Xu, Xinyu Jiang, and Wei Chen. 2020. Photoplethysmography motion artifacts removal based on signal-noise interaction modeling utilizing envelope filtering and time-delay neural network. *IEEE Sensors Journal* 20, 7 (April 2020), 3732–3744. https://doi.org/10.1109/JSEN.2019.2960370

[42] Natasha Yamane, Varun Mishra, and Matthew S Goodwin. 2024. HeartView: An extensible, open-source, web-based signal quality assessment pipeline for ambulatory cardiovascular data. In *International Conference on Pervasive Computing Technologies for Healthcare*. Springer Nature Switzerland, Cham, 107–123. https://doi.org/10.1007/978-3-031-59717-6_8

[43] Roberto Zangróniz, Arturo Martínez-Rodrigo, María T. López, José Manuel Pastor, and Antonio Fernández-Caballero. 2018. Estimation of mental distress from photoplethysmography. *Applied Sciences* 8, 1 (Jan. 2018), 69. https://doi.org/10.3390/app8010069

[44] Zhilin Zhang, Zhouyue Pi, and Benyuan Liu. 2015. TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering* 62, 2 (Feb. 2015), 522–531. https://doi.org/10.1109/TBME.2014.2359372