Robust PPG-based Ambulatory Heart Rate Tracking Algorithm

Nicholas Huang, and Nandakumar Selvaraj*

Abstract— **Recent advances in wearable devices with optical Photoplethysmography (PPG) and actigraphy have enabled inexpensive, accessible, and convenient Heart Rate (HR) monitoring. Nevertheless, PPG's susceptibility to motion presents challenges in obtaining reliable and accurate HR estimates during ambulatory and intense activity conditions. This study proposes a lightweight HR algorithm, TAPIR: a Time-domain based method involving Adaptive filtering, Peak detection, Interval tracking, and Refinement, using simultaneously acquired PPG and accelerometer signals. The proposed method is applied to four unique, wrist-wearable based, publicly available databases that capture a variety of controlled and uncontrolled daily life activities, stress, and emotion. The results suggest that the current HR prediction is significantly ($P<0.01$) more accurate during intense activity conditions than the contemporary algorithms involving Wiener filtering, time-frequency analysis, and deep learning. The current HR tracking algorithm is validated to be of clinical-grade and suitable for low-power embedded wearable systems as a powerful tool for continuous HR monitoring in real-world ambulatory conditions.**

Index Terms— **Photoplethysmography, Heart Rate, Adaptive Filter, Motion Artifact, Peak Detection**

I. INTRODUCTION

Heart rate (HR) ranks among the primary vital signs used to evaluate a person's health. HR monitoring has a multitude of applications, ranging from personal fitness to health monitoring. Traditional methods of HR estimation often involve performing peak detection on Electrocardiogram (ECG) signals. ECG has prominent R wave peaks that are relatively easily identifiable and generally resilient to motion artifacts. However, the traditional wired ECG recorders such as Holter/portable telemetry monitors, and the emerging adhesive based ECG devices are uncomfortable and limited for continuous long-term monitoring applications [1].

In contrast, optical-based pulse oximeter sensors are ubiquitous in clinical practice and wearable consumer health, and are available in various form-factors attached to a body site including finger, wrist, upper arm, and earlobe [2]. These devices allow convenient acquisition of blood volume pulse, also known as a Photoplethysmogram (PPG). PPG-based HR estimation under stationary conditions is widely considered as a reasonable surrogate for ECG-based HR monitoring [3]. Wearable watch/band PPG sensors are comfortable enough to be worn continuously and can capture day-to-day ambulatory patterns and pulse waveforms. In such real-world scenarios, PPG signals are often corrupted by human motion during activities of daily living and work outs. Obtaining reliable HR during motion is clinically very useful for studying exercise physiology, cardio-respiratory dynamics, and endurance of the user. To that end, motion tolerant algorithms for accurate estimation of HR using PPG during ambulatory conditions have been recently evolving [4].

ECG-based HR estimation involves detecting individual QRS complexes; similarly, peak detection methods have also traditionally been used for PPG [5]–[8]. Recent advances in HR algorithms explore the use of spectral analyses to effectively isolate frequency information. Major spectral approaches tend to include motion artifact removal (e.g. Wiener filtering [9], adaptive filtering [10], or spectral subtraction [11]), and conversion to time-frequency representation (e.g. Sparse Signal Reconstruction [12], FFT with phase vocoder [9], or Truncated Fourier Series [13]). More recently, deep learning has also been applied in this field [14].

This study presents a novel algorithm for ambulatory HR estimation based on simultaneous PPG and accelerometer signals involving four major Time-domain processing steps: Adaptive filtering, Peak detection, Interval tracking, and Refinement. In the interest of brevity, this methodology is referred to as TAPIR. The algorithm framework and validation methods are detailed in Section II. Performance results from the proposed method and contemporary algorithms involving adaptive filtering, time-frequency spectra, and deep learning are shown in Section III. Finally, the discussion and conclusion are given in Section IV.

II. MATERIALS AND METHODS

A. Databases

Several public databases that include one or two PPG signals, triaxial accelerometer signals, and HR annotations derived from an ECG signal are chosen for this study.

1) IEEE Dataset [12], [15]: The IEEE Signal Processing Cup is an annual competition with a different topic and dataset each year. In 2015, the objective was to extract HR from PPG recordings contaminated by motion artifacts. The IEEE dataset is divided into two subgroups per the design of the competition. The IEEE Train and Test datasets consist of 12 and 10 recordings respectively, each 5 min long, with subjects walking and running on a treadmill (Train dataset) or performing intense arm exercises (Test dataset).

2) Daily Life Activities (DaLiA) [14]: Overall, the IEEE dataset has high quality PPG recordings, but is limited in range of activities and duration. Two other large datasets including relatively long duration PPG recordings were recently collected and published for public use. These two databases contain PPG and triaxial acceleration recordings from an Empatica E4 wristband and an ECG signal from a RespiBAN device on the chest.

Authors are with Biofourmis Inc., Boston, MA 02110 USA. (*correspondence e-mail: kumar@biofourmis.com)

The DaLiA (Daily Life Activities) database captures a variety of daily life activities over $2-3$ hours scheduled in order: 1 Sitting, 2 Stairs, 3 Table Soccer, 4 Cycling, 5 Driving, 6 Lunch, 7 Walking, 8 Working. Each activity was performed for a specified duration, with time between activities varying across 15 subjects. HR annotations are used as ground truth for performance analysis.

3) Wearable Stress and Affect Detection (WESAD) [16]: The WESAD dataset was recorded by the same group that created the DaLiA database, employing the same recording equipment for another 15 subjects. Instead of daily life activities, these 2-hour recordings focused on the subjects' affect. Tested states were baseline, amusement, stress, and meditation conditions. The ECG signal was used to derive the reference HR using a standard peak detection method based on the Pan-Tompkins algorithm [17].

B. Heart Rate Estimation Method (TAPIR)

1) Preprocessing: The available PPG and accelerometer signals are first preprocessed [9] using a Butterworth filter with corner frequencies at 0.4 and 4 Hz, corresponding to 24 to 240 beats per minute (bpm). Afterwards, each of these signals are Z-Score normalized. For the recordings with multiple PPG signals (IEEE datasets), the PPG signals are combined by averaging the normalized signals.

2) Adaptive Filter for Motion Artifact Removal: Adaptive filters are commonly used to remove motion artifacts from PPG signals [10], [18]–[22]. Often, adaptive noise cancellation (ANC) is performed using each accelerometer axis in sequence [10], [20], [21]. However, applying the adaptive filter with all three signals at once allows the filter weights to be properly adjusted with information from all motion vectors simultaneously.

Currently, motion artifacts are removed using a least mean square (LMS) adaptive filter [23] using all 3 accelerometer axes with delays up to 250 ms as the input and the PPG signal as the desired output. The adaptive filter is adjusted so that the linear combination of the previous accelerometer data best fits the PPG in terms of lowest mean squared error. This filtered signal is an estimate of noise caused by motion and is removed from the PPG signal by subtraction.

An important parameter for the LMS adaptive filter is the step size [23] (μ, final value $8 * 10^{-5}$). A too-small step size may result in a poor filter while a too-large step size may fail to converge. Because the step size is not selected based on any individual subject's data, whenever the average absolute value of the weights exceeds 1, the weights are all immediately reset to the original value of 0.

3) Peak Detection: Peak detection is performed using a method based on the MATLAB function *mspeaks* from the Bioinformatics toolbox [24]. Peaks are initially assigned by identifying all the positive-valued local maxima in the signal. Afterwards, all intervals less than a *Refractory Period* (RP, final value of 300 ms) are identified, and the lower amplitude bordering peak for each such interval is removed. This step both reduces noise and limits the estimated HR from exceeding a physiological range.

4) Peak Adjustment by Interval Tracking on a Short Time-Scale: Peak adjustment involves tracking the average peak-to-peak interval over time. Starting from the 11^{th} interval, if the peak-to-peak time is greater than some percentage (*Short-term Upper bound* or SU, final value 180%) of the average of the previous 10 intervals, then a peak may be missing in that interval (false negative). Any local maximum with a magnitude above an amplitude threshold (*Short-term Threshold* or ST, final value 0) within that interval is chosen as another possible peak location. If the interval width is less than another percentage (*Short-term Lower bound* or SL, final value 50%) of the average of the previous 10, then it is assumed that an extraneous peak was selected (false positive). In this case, the latter peak is excluded.

5) Preliminary Heart Rate Estimation by Interval Tracking on a Long Time-Scale: A preliminary HR estimate is made using a second interval tracking procedure similar to Peak Adjustment, but with a longer time scale. As an initialization step, a sliding standard deviation is calculated starting from the first 30 intervals. If the standard deviation of those 30 intervals is below a threshold (*Baseline Consistency* or BC, final value 10 bpm), then those 30 intervals are labelled as 'good' intervals, and the long-term interval tracking procedure begins. If not, the window is shifted forward by one interval, and the standard deviation is again calculated.

The long-term interval tracking procedure closely matches the Peak Adjustment process, but with three key differences. First, a longer time scale is used, tracking the average of the previous 30 'good' intervals rather than just any previous 10 intervals. Further, a stricter threshold is used with the *Long-term Upper bound* (LU) at a final value of 140% and the *Long-term Lower bound* (LL) at 70%. Finally, rather than adding or removing peaks, intervals are simply labelled as 'good' or not depending on whether or not they fit between the upper and lower bounds. As a clarification, the standard deviation is only used in the one-time baseline initialization procedure, not in the interval tracking.

Once the intervals are evaluated using this procedure, HR is estimated using a sliding window of length 8 seconds and step size 2 seconds. For each window, all 'good' intervals with a midpoint within the window are averaged. Then the reciprocal of the average window is taken and multiplied by 60 to reach an initial HR estimate in bpm.

6) Notch Filter HR Refinement: The final step in the algorithm is to refine the initial estimate of the HR by using a notch filter-based strategy similar to that employed by Wang et al. [10]. First, a second-order notch filter is constructed of bandwidth .5 Hz at the estimated HR frequency for each 8-second window. The PPG signal is then filtered by this notch filter, and the result is subtracted from the PPG signal.

Peaks are re-obtained from the HR-focused PPG signal for each window using the Peak Detection method with the same height and minimum distance thresholds. However, only peaks located near the center of each window are kept to avoid edge effects and for combination across windows. Lastly, the Heart Rate Estimation procedure is repeated on the new set of peaks to obtain the final refined HR prediction.

C. Performance Evaluation and Statistical Analysis

Performance is evaluated by calculating the mean absolute error (MAE). All HR estimations are performed on 8-second sliding windows with a 2-second step size, in order to match annotations in the databases. Error is calculated by first subtracting the prediction from the annotated HR, then taking the absolute value of the error, and finally taking the mean across time. Statistical analysis is performed by comparing TAPIR with previous algorithms using paired sample t-tests on the MAE values across subjects.

For comparison to TAPIR, three contemporary methods of HR estimation are also evaluated. Wiener Filtering and Phase Vocoder (WFPV) results are reproduced from Temko [9] for the IEEE database; the publicly available code for WFPV is applied directly to DaLiA and WESAD. WFPV is a well-known HR estimation approach based on frequency-domain information that has previously been validated only on the IEEE dataset. In addition, HR estimation results from SpaMaPlus (another time-frequency approach) and CNN Ensemble (a deep learning method) trained with leave-one-session-out cross validation are reproduced from Reiss et al. [14] for all public datasets involved.

To evaluate the progression in accuracy at each stage of the TAPIR method, the error calculation is carried out for intermediate stages of the algorithm. First, the performance of the concurrent adaptive filter is compared to a baseline cascaded filter. Then, MAE is evaluated with additional stages of TAPIR included cumulatively.

Furthermore, HR estimation performance is also evaluated for individual activities in the DaLiA dataset. First the HR annotations and estimates are divided based on which activity was being performed at the midpoint of the corresponding 8-second window. Finally, the MAE is computed for all of these samples combined across subjects.

III. RESULTS

A. Performance Comparison with Contemporary Algorithms

Fig. 1 shows representative best (panel A) and worse (panel B) cases of TAPIR derived HR estimations among the DaLiA database involving real-world daily activities, which associate with an MAE of 2.8 bpm and 8.0 bpm, respectively,. The worst case still shows a good correspondence to that of HR annotations throughout the intense activities.

The overall MAE performances of the TAPIR method for the IEEE, DaLiA, and WESAD datasets are compared as boxplots in Fig. 2 alongside SpaMaPlus, WFPV, and CNN Ensemble contemporary methods. TAPIR offers the best HR performance in DaLiA and WESAD databases compared to the other methods. The frequency-domain based WFPV method had the best performance in the IEEE Test database but not in other datsets. The time-frequency based SpaMaPlus method was comparably poor in all the 3 datasets, while CNN ensemble algorithm had a relatively poor performance only for the IEEE Test dataset. Furthermore, SpaMaPlus and CNN methods had one or more outliers at a clinically unacceptable error range. In contrast,

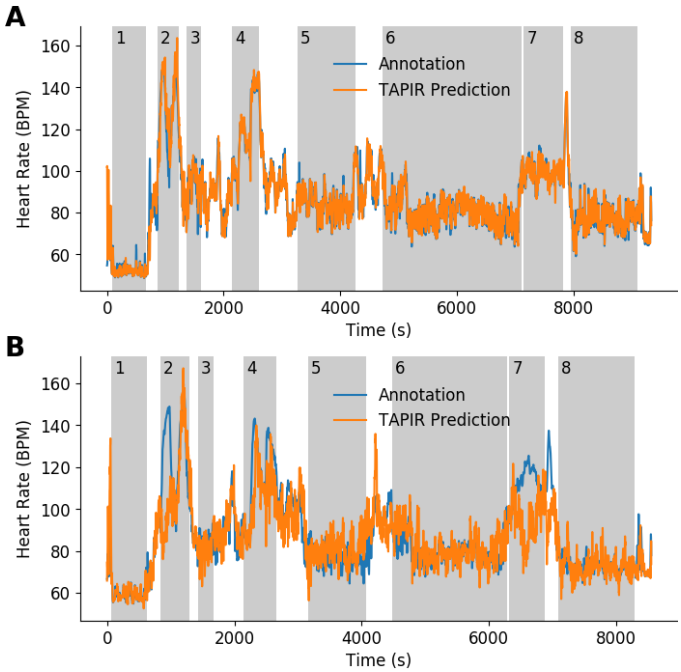

Fig. 1. Representative sample recordings of predicted HR vs reference HR from DaLiA dataset. A) Best accuracy, Subject 7. B) Worst accuracy, Subject 9. Shaded areas represent times with varying activities: 1 Sitting, 2 Stairs, 3 Table Soccer, 4 Cycling, 5 Driving, 6 Lunch, 7 Walking, 8 Working.

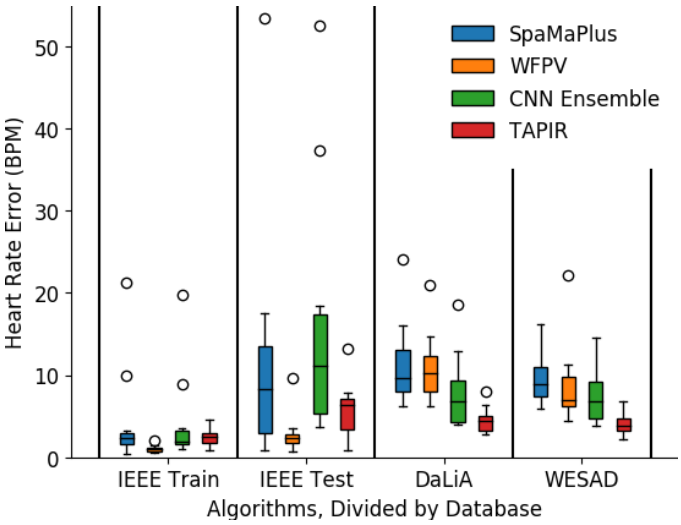

Fig. 2. Performance of various HR estimation algorithms for the evaluated datasets in terms of MAE. Boxes represent interquartile range of HR error, with the median indicated by a line in the box, and outliers are depicted as circles.

TAPIR had consistently much lower error for HR estimation in the presence of intense motion. The slightly higher MAE for TAPIR in the IEEE Test set is due to a lack of a clean stationary baseline for initializing the interval tracking procedure, but the overall accuracy remains relatively higher despite this challenge.

The MAE performance of HR estimation among the four methods for each subject in the DaLiA and WESAD datasets are listed in Tables I and II, respectively. TAPIR has significantly ($P<0.01$) lower MAE in both DaLiA and WESAD datasets (4.6 ± 1.4 bpm and 4.2 ± 1.4 bpm, respectively) and

TABLE I

HEART RATE (HR) ESTIMATION PERFORMANCE ON DALIA DATASET. VALUES ARE MEAN ABSOLUTE ERROR (MAE) IN UNITS OF BPM.

DaLiA	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	All[a]	p-value[b]
SpaMaPlus	8.9	9.7	6.4	14.1	24.1	11.3	6.3	11.3	16.0	6.2	15.2	12.0	8.5	7.8	8.3	11.1 ± 4.8	$< .001$
WVPF	9.8	8.2	8.7	11.6	20.9	12.9	6.8	11.3	14.6	6.2	13.7	10.2	7.9	7.0	10.2	10.7 ± 3.8	$< .001$
CNN ensemble	7.7	6.7	4.0	5.9	18.5	12.9	3.9	10.9	8.8	4.0	9.2	9.4	4.3	4.4	4.2	7.7 ± 4.2	$< .01$
TAPIR	4.5	4.5	3.2	6.0	5.0	3.4	2.8	6.3	8.0	2.9	5.1	4.7	3.1	5.0	4.1	4.6 ± 1.4	-

[a] HR estimation performance overall presented as mean $\pm$ 1 standard deviation across the 15 subjects.
[b] P-values from paired sample t-test against the TAPIR method.

TABLE II

HR ESTIMATION PERFORMANCE ON WESAD DATASET. VALUES ARE MAE IN UNITS OF BPM.

WESAD	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S13	S14	S15	S16	S17	All[a]	p-value[b]
SpaMaPlus	7.3	12.6	5.9	7.4	6.4	9.2	8.9	6.3	8.2	16.3	13.0	10.2	8.3	10.7	11.2	9.5 ± 2.9	$< .001$
WFPV	6.4	22.2	4.4	11.3	5.0	7.0	6.5	5.5	6.0	10.5	9.2	6.6	8.1	7.7	10.5	8.5 ± 4.3	$< .001$
CNN ensemble	5.1	14.5	7.8	7.7	3.9	6.8	4.3	4.0	8.9	11.1	6.5	5.3	4.3	12.8	9.4	7.5 ± 3.3	$< .01$
TAPIR	4.5	5.1	3.0	6.8	3.6	5.9	3.9	3.5	2.7	4.3	3.3	2.1	3.8	3.0	6.8	4.2 ± 1.4	-

[a] HR estimation performance overall presented as mean $\pm$ 1 standard deviation across the 15 subjects.
[b] P-values from paired sample t-test against the TAPIR method.

is highly consistent compared to all the other methods.

Additionally, the average and standard deviation of signed error for TAPIR are found to be -0.3 ± 1.2 bpm and 7.8 ± 2.4 bpm for DaLiA, respectively. Likewise, bias and precision of TAPIR in WESAD are -0.2 ± 1.4 bpm and 7.1 ± 2.3, respectively.

When an additional moving average filter (duration of 1 min and step size of 2 sec) is uniformly applied to both the reference HR and TAPIR based HR estimations, the TAPIR method achieves much lower MAE rates of 1.8, 4.2, 3.0, and 2.7 bpm on average for the IEEE Train, IEEE Test, DaLiA, and WESAD datasets respectively.

HR estimation error for the DaLiA dataset is further broken down by the type of activities in Table III; TAPIR outperforms the other methods for every activity type, most by a large margin.

TABLE III

HR PERFORMANCE (MAE) IN DALIA PER ACTIVITY TYPES.

	SpaMaPlus	WFPV[a]	CNN Ens.	TAPIR
Transition	14.34	11.35 (16.68)	8.77	5.47
Sitting	4.27	18.80 (2.76)	4.93	1.63
Stairs	25.42	22.80 (35.31)	16.98	12.00
Table Soccer	21.48	18.76 (27.35)	12.16	5.78
Cycling	11.97	17.98 (31.94)	12.48	5.14
Driving	6.24	5.69 (7.70)	4.96	3.49
Lunch	7.33	5.89 (8.53)	5.22	3.29
Walking	18.16	13.92 (21.30)	9.21	7.13
Working	4.91	4.16 (4.83)	3.84	2.33

[a] Parentheses indicate error values for WFPV with an extended range of potential heart rates from 60-180 bpm to 30-200 bpm.

B. Contribution of each Stage of the TAPIR Method

The performance of various stages in TAPIR are also evaluated by incorporating successive stages of the algorithm (Table IV). Cascaded adaptive noise cancelling (ANC) with peak detection is presented as a baseline, where each accelerometer axis is used in sequence; that resulted in a clinically unacceptable error range for HR estimation in all four datasets. Concurrent use of all three axes of ACC for the adaptive filter had little improvement in HR performance, but further addition of the remaining elements (peak adjustment, interval tracking and refinement) drastically reduced the error rates across all the datasets as listed.

TABLE IV

INCREMENTAL HR ESTIMATION PERFORMANCE (MAE) FOR TAPIR

	IEEE Train	IEEE Test	DaLiA	WESAD
Cascaded ANC	8.1 ± 6.7	17.3 ± 13.5	13.2 ± 4.8	13.0 ± 5.5
Concurrent ANC	5.7 ± 3.4	17.1 ± 14.0	13.1 ± 4.8	12.9 ± 5.6
+ Peak Adjustment	3.9 ± 2.7	12.2 ± 9.6	9.6 ± 3.1	7.3 ± 2.1
+ Interval Tracking	2.6 ± 6.1	6.1 ± 3.6	5.1 ± 1.7	4.4 ± 1.4
+ Refinement	2.5 ± 1.2	5.9 ± 3.5	4.6 ± 1.4	4.2 ± 1.4

IV. DISCUSSION

Monitoring HR more reliably and accurately during ambulatory conditions including activities of daily living and physical exercises has tremendous diagnostic and prognostic value for clinicians to assess one's health, fitness, and endurance. However, obtaining accurate continuous HR in the presence of motion using peripheral wearable devices with simple PPG optical sensing remains a challenge. The present novel, lightweight TAPIR algorithm achieves consistently high accuracy for HR tracking during intense physical exercise conditions as demonstrated by the validation results.

A. Comparison to Contemporary Methods

Compared to contemporary algorithms, TAPIR offers significantly lower error rates for the challenging DaLiA and WESAD databases, which most resemble real-world conditions (Tables I and II). Although WFPV shows the strongest performance for IEEE Test, it struggles with longer, more varied recordings. Of note, the WFPV approach performed poorly on the relatively low-noise 'Sitting' activity (Table III), primarily due to that method limiting HR between 60-180 bpm. This range of heart rates is well-suited to the IEEE datasets, but in DaLiA, HR sometimes falls below 60 bpm.

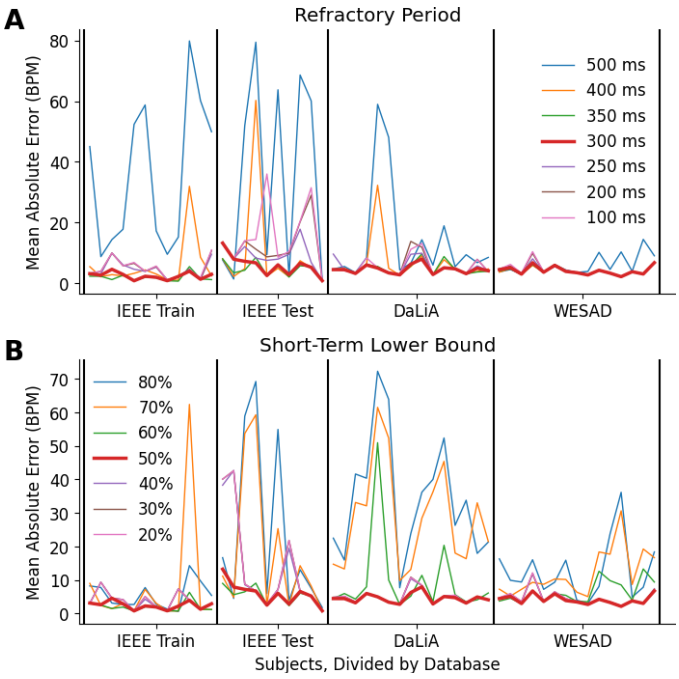

Fig. 3. Effects of varying parameters on HR error (MAE) for two example parameters. Bold red line corresponds to values used in this study. A) Refractory period RP, during Peak Detection. B) Short-term Lower bound SL, during Peak Adjustment

By extending the range to a more inclusive 30-200 bpm, the MAE for the sitting activity is reduced greatly, but at a high cost of accuracy for all of the other sections (to an average error of 14.0 bpm overall). This result further evidences that prediction models that are only trained on the IEEE dataset may be overfitted and unable to be applied to all situations.

In contrast, the SpaMaPlus and CNN Ensemble methods exhibit very poor performance with extreme outliers on the IEEE Test dataset. In addition to high noise, Reiss et al. [14] attribute some of this error to lack of sufficient training data for cross validation. This finding highlights the reliance of these methods on access to an abundance of similar training data to tune models for different settings. With its shared set of parameters across all available data, TAPIR does not suffer from this limitation.

B. Parameter Selection and Generalization

TAPIR algorithm parameters have been optimized without using machine learning to fit precisely to the data, and the parameters are kept the same across different datasets, subjects, sensor types, and physical activities. This practice stands in contrast to the deep neural network presented by Reiss et al. [14], where parameters needed to be tuned carefully to avoid overfitting and loss of generalization. The current HR algorithm achieves strong performance with the same parameters for all recordings over distinct datasets. Thus, the proposed HR estimation is validated to generalize well for various settings and recording modalities.

For example, Fig. 3 demonstrates the effect of varying two TAPIR parameters: the refractory period between two peaks (RP) and the Short-term Lower bound threshold (SL) in the peak adjustment procedure. As evidenced in the plot, the final parameter values (bold red line) offer the best performances consistently for all four datasets.

C. Merits of the Lightweight Time-Domain Method

As a time-domain method of HR estimation, TAPIR exhibits several advantages over frequency-domain strategies. For example, working with the time-domain PPG signal provides the ability to exclude noisy epochs from the HR estimate. This approach also enables the evaluation of HR variability for short intermittent segments while maintaining the continuity of HR outputs as much as possible. Further, the method is highly suitable for time-varying, irregular, and unstable HR variations characteristic of arrhythmias such as atrial fibrillation. At the same time, TAPIR is also well-suited to customize the HR estimation over long time windows.

An additional advantage that TAPIR enjoys is a high degree of computational efficiency; the algorithm primarily consists of simple calculations and a few linear filters. Further, because all computations are performed on the time-domain PPG signal, the memory requirements are relatively low that make TAPIR highly suitable for embedded systems.

D. Limitations and Future Directions

TAPIR is currently tested only on healthy subjects due to a lack of public databases containing PPG recordings from patients with conditions such as arrhythmia. Such cases could prove challenging for the algorithm because of its reliance on interval regularity. TAPIR is not alone; nearly all state of the art algorithms also explicitly use this consistency, often to limit large jumps in HR prediction [9], [10], [12], [25], [26]. While relying on heart beat consistency is vital to maintaining HR tracking in high activity conditions for healthy subjects, it could present difficulties in detection of certain cardiac disorders.

One way to overcome this challenge is to first determine whether an arrhythmia is present, whether by manual selection after diagnosis or during a clean baseline period. Using that information, the interval tracking could be adjusted or removed entirely if the heart beat is found to be too erratic. Another is to evaluate activity level using accelerometer recordings, and to have more confidence in the peaks detected when activity is low.

As previously stated, frequency-domain methods have been able to achieve high accuracy for the IEEE database. Some optimal combination with frequency-domain information could improve TAPIR's HR estimate for the IEEE database, without harming the high accuracy for DaLiA and WESAD. One possibility is to combine estimates using some measure of confidence for both TAPIR and a frequency-domain approach; for TAPIR, a straightforward metric would be the standard deviation of the intervals in each time window. Such a combination could significantly improve the already strong HR prediction presented in this work.

Overall, TAPIR is a novel method of HR monitoring using PPG and accelerometer recordings that achieves consistently high accuracy across a wide range of recording equipment, people, and activity levels, without additional training or

modification. Due to computational efficiency, ability to generalize, and resilience to even high-intensity activity, TAPIR is an ideal algorithm for wearable devices to be used during exercise and other everyday activities.

REFERENCES

[1] A. Bansal and R. Joshi, "Portable out-of-hospital electrocardiography: A review of current technologies," *Journal of Arrhythmia*, vol. 34, no. 2, pp. 129–138, Apr. 2018.

[2] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, Apr. 2014.

[3] W. H. Lin, D. Wu, C. Li, H. Zhang, and Y. T. Zhang, "Comparison of heart rate variability from PPG with that from ECG," in *IFMBE Proceedings*, vol. 42, 2014, pp. 213–215.

[4] D. Biswas, N. Simoes-Capela, C. Van Hoof, and N. Van Helleputte, "Heart Rate Estimation From Wrist-Worn Photoplethysmography: A Review," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.

[5] T. H. Fu, S. H. Liu, and K. T. Tang, "Heart rate extraction from photoplethysmogram waveform using wavelet multi-resolution analysis," *Journal of Medical and Biological Engineering*, vol. 28, no. 4, pp. 229–232, 2008.

[6] H. S. Shin, C. Lee, and M. Lee, "Adaptive threshold method for the peak detection of photoplethysmographic waveform," *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1145–1152, 2009.

[7] S. Bagha and L. Shaw, "A Real Time Analysis of PPG Signal for Measurement of SpO2 and Pulse Rate," *International Journal of Computer Applications*, vol. 36, no. 11, pp. 45–50, 2011.

[8] X. Sun, P. Yang, Y. Li, Z. Gao, and Y. T. Zhang, "Robust heart beat detection from photoplethysmography interlaced with motion artifacts based on empirical mode decomposition," in *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*, 2012, pp. 775–778.

[9] A. Temko, "Accurate Heart Rate Monitoring during Physical Exercises Using PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2016–2024, 2017.

[10] M. Wang, Z. Li, Q. Zhang, and G. Wang, "Removal of motion artifacts in photoplethysmograph sensors during intensive exercise for accurate heart rate calculation based on frequency estimation and notch filtering," *Sensors (Switzerland)*, vol. 19, no. 15, p. 3312, Jul. 2019.

[11] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1902–1910, 2015.

[12] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 522–531, 2015.

[13] H. Dubey, R. Kumaresan, and K. Mankodiya, "Harmonic sum-based method for heart rate estimation using PPG signals affected with motion artifacts," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 1, pp. 137–150, Feb. 2018.

[14] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks," *Sensors*, vol. 19, no. 14, p. 3079, Jul. 2019.

[15] Z. Zhang, "IEEE Signal Processing Cup 2015: Heart Rate Monitoring During Physical Exercise Using Wrist-Type Photoplethysmographic (PPG) Signals," 2015.

[16] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, "Introducing WeSAD, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 2018 International Conference on Multimodal Interaction*. New York, New York, USA: ACM Press, 2018, pp. 400–408.

[17] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.

[18] S. Fallet and J. M. Vesin, "Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: An approach based on adaptive filtering," *Physiological Measurement*, vol. 38, no. 2, pp. 155–170, 2017.

[19] Y. Ye, Y. Cheng, W. He, M. Hou, and Z. Zhang, "Combining Nonlinear Adaptive Filtering and Signal Decomposition for Motion Artifact Removal in Wearable Photoplethysmography," *IEEE Sensors Journal*, vol. 16, no. 19, pp. 7133–7141, Oct. 2016.

[20] D. Jarchi and A. J. Casson, "Towards Photoplethysmography-Based Estimation of Instantaneous Heart Rate during Physical Activity," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2042–2053, Sep. 2017.

[21] M. T. Islam, S. T. Ahmed, I. Zabir, C. Shahnaz, and S. A. Fattah, "Cascade and parallel combination (CPC) of adaptive filters for estimating heart rate during intensive physical exercise from photoplethysmographic signal," *Healthcare Technology Letters*, vol. 5, no. 1, pp. 18–24, 2018.

[22] S. S. Chowdhury, S. Hasan, and R. Sharmin, "Robust Heart Rate Estimation from PPG Signals with Intense Motion Artifacts using Cascade of Adaptive Filter and Recurrent Neural Network," *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pp. 1958–1963, Oct. 2019.

[23] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2005.

[24] The Mathworks Inc, "MATLAB and Bioinformatics Toolbox Release 2013a," Natick, Massachusetts, United States, 2013.

[25] S. M. Salehizadeh, D. Dao, J. Bolkhovsky, C. Cho, Y. Mendelson, and K. H. Chon, "A novel time-varying spectral filtering algorithm for reconstruction of motion artifact corrupted heart rate signals during intense physical activities using a wearable photoplethysmogram sensor," *Sensors (Switzerland)*, vol. 16, no. 1, 2015.

[26] M. A. Motin, C. K. Karmakar, and M. Palaniswami, "PPG Derived Heart Rate Estimation during Intensive Physical Exercise," *IEEE Access*, vol. 7, pp. 56 062–56 069, 2019.