

The final version of this manuscript is now available with open access at PeerJ (<https://doi.org/10.7717/peerj.13147>).

RapidHRV: an open-source toolbox for extracting heart rate and heart rate variability

Peter A. Kirk^{1,2,*}, Alex Davidson Bryan³, Sarah N. Garfinkel¹, & Oliver J. Robinson^{1,4}

¹ *UCL Institute of Cognitive Neuroscience, UK;* ² *UCL Department of Experimental Psychology, UK;* ³ *Independent Scholar;* ⁴ *UCL Clinical, Educational and Health Psychology, UK*

Author Note

This work was supported by the Leverhulme Trust as part of the Doctoral Training Program for the Ecological Study of the Brain (DS-2017-026, to P.A.K), MRC and NIHR CARP award (MR/V037676/1, to S.N.G.), and MRC Senior Non Clinical Fellowship award (MR/R020817/1, to O.J.R.).

*Corresponding author: Peter Kirk, p.kirk@ucl.ac.uk, University College London, London, UK.

Abstract

Heart rate and heart rate variability have enabled insight into a myriad of psychophysiological phenomena. There is now an influx of research attempting using these metrics within both laboratory settings (typically derived through electrocardiography or pulse oximetry) and ecologically-rich contexts (via wearable photoplethysmography, i.e. smartwatches). However, these signals can be prone to artifacts and a low signal to noise ratio, which traditionally are detected and removed through visual inspection. Here, we developed an open-source Python package, RapidHRV, dedicated to the preprocessing, analysis, and visualization of heart rate and heart rate variability. Each of these modules can be executed with one line of code and includes automated cleaning. In simulated data, RapidHRV demonstrated excellent recovery of heart rate across most levels of noise ($\geq 10\text{dB}$) and moderate-to-excellent recovery of heart rate variability even at relatively low signal to noise ratios ($\geq 20\text{dB}$) and sampling rates ($\geq 20\text{Hz}$). Validation in real datasets shows good-to-excellent recovery of heart rate and heart rate variability in electrocardiography and finger photoplethysmography recordings. Validation in wrist photoplethysmography demonstrated RapidHRV estimations were sensitive to heart rate and its variability under low motion conditions, but estimates were less stable under higher movement settings.

Keywords: Heart rate variability, Remote sensing, Python, Toolbox

RapidHRV: an open-source toolbox for extracting heart rate and heart rate variability

Evidence has outlined a link between heart rate, heart rate variability, and health-related risks, ranging from cardiac mortality to mental illness (Hillebrand et al., 2013; Jandackova et al., 2016; Makovac et al., 2016a; Pham et al., 2021). Consequently, there is now an influx of research looking into whether these measures can be derived in naturalistic settings to track clinically-relevant outcomes, namely through wearable devices (Georgiou et al., 2018; Mulcahy et al., 2019). However, a key issue when opting to use the measures in naturalistic settings are the low signal to noise ratios (e.g. photoplethysmography (PPG), a typical measure for cardiac monitoring in wrist wearables, Caizzone et al., 2017). Moreover, heart rate variability measures generally require relatively longer windows for extraction compared to heart rate (Baek et al., 2015). Thus, significant noise poses a problem for out-of-laboratory applications, as point estimates can contain large errors from technological limitations and motion artifacts within windows of extraction. In experimental settings, noise has often been dealt with through visual inspection of data (Makovac et al., 2016b; Rae et al., 2020); but when approaching time courses in relatively larger-scale samples, manual outlier detection is not a pragmatic solution.

Whilst some open-source packages are already available for the analysis of heart rate and heart rate variability, these are typically modality-specific, and not targeted at wrist-worn measures (e.g. *pyVHR* for video-based estimation, Boccignone et al., 2020). Some modality-general packages do exist, but these often still require manual visual inspection and/or can require custom scripting on the users end for tailoring to e.g. noisy, wrist-worn PPG measures ('Analysing_Smartwatch_Data' in *HeartPy*, van Gent et al., 2019; *NeuroKit2*, Pham et al., 2021). As such, these are often less suited for dealing with datasets collected across large time frames. Consequently, we set out to develop a simple yet flexible toolbox for the extraction of time-domain heart rate and heart rate variability measures with automated artifact rejection applicable across recording modalities, including wrist-worn PPG. Here, we present the development and validation of an open-source Python package, 'RapidHRV'.

Pipeline

RapidHRV was developed in Python (V 3.7.9). RapidHRV source code and tutorials are available to download through PyPi (<https://pypi.org/project/rapidhrv/>) and GitHub (<https://github.com/peterakirk/RapidHRV>). Below we provide an overview of RapidHRV's preprocessing, analysis (figure 1), and visualization. Each of the three modules only requires one function (one line of code) to run, for which we have embedded examples at the end of the relevant sections below.

–Figure 1–

Preprocessing

First, data is upsampled with cubic spline interpolation (3rd order polynomial; default = 1kHz) to increase temporal accuracy of peak detection. To mitigate potential long-term drifts in the signal, the pipeline then applies a high pass butterworth filter (0.5Hz) across the input data. Finally smoothing with a savitzky-golay filter (3rd order polynomial; default = 100ms) is applied to reduce spiking (sharp increases in the signal caused by artifacts such as motion) whilst retaining temporal precision.

Extracting Heart Rate and Heart Rate Variability

Following preprocessing, the pipeline scales the data (between 0 and 100) and runs peak detection on every window (default width = 10s; for a methodological discussion and prior validation of using ultra-short, 10s windows in heart rate variability estimation, see Munoz et al., 2015). This outputs peaks and their properties (e.g. heights, amplitudes, width; *SciPy* 'find_peaks', Virtanen et al., 2020). For ECG data however, peak detection is vulnerable to irrelevant prominent P and T waves. Specifically, traditional amplitude-based analyses may occasionally detect non-R wave peaks that demonstrate a similar or greater amplitude than R waves. Consequently, for ECG data, RapidHRV can implement k-means clustering ($k = 3$) to

discern R waves from P and T waves prevalent in the signal (*scikit-learn* ‘KMeans’, Pedregosa et al., 2011). This is implemented by reducing the minimum amplitude threshold to near-zero (i.e. 5%), running amplitude-based peak detection, then sorting peaks into three clusters using relevant properties (i.e. peak widths, heights, and prominences). R waves are then determined based on cluster centroids for peak properties, expecting R waves to hold higher prominences *and* lower widths compared to P and T waves. Figure 2 demonstrates an example (from dataset 3, see *Validation Methods*) wherein amplitude-based analyses may incorrectly identify T waves as R waves in an atypical ECG signal, and how RapidHRV’s peak clustering helps mitigate this.

–Figure 2–

As RapidHRV uses fixed movements for the sliding window, a window can start or end at any point during the cardiac signal. This can occasionally result in underestimation of the first/last peak’s amplitude as the baseline value may—for example—be set during the P wave. Therefore, RapidHRV recalculates amplitudes of the first/last peaks using baseline imputation from the neighboring peaks.

For extracting beats per minute (BPM) the number of peaks, $k - 1$, is multiplied by 60 (seconds), and divided by the difference in time between the first and last peaks, i and j :

$$\text{BPM} = ((k - 1) * 60) / (j - i)$$

The root mean square of successive differences is calculated by obtaining: (1) the interbeat interval, IBI_i , between neighboring peaks; (2) the successive differences in interbeat intervals, $\text{IBI}_i - \text{IBI}_{i+1}$; (3) the square of differences; (4) the mean of squared differences (dividing by the number peaks, $N, - 1$); and (5) the root of the mean square of successive differences (RMSSD):

$$\text{RMSSD} = \sqrt{\frac{\sum_{i=1}^{N-1} (\text{IBI}_i - \text{IBI}_{i+1})^2}{N-1}}$$

BPM and RMSSD were selected as the primary measures as they appear to be the most stable metrics when derived from ultra-short recordings (Baek et al., 2015). RapidHRV also supplements these measures with the standard deviation of N-N intervals (SDNN), standard deviation of successive differences (SDSD), proportion of successive differences greater than 20ms (PNN20), proportion of successive differences greater than 50ms (PNN50), and high-frequency power (HF; note: as this requires more data points than time-domain analyses NaN is returned if there is insufficient data).

Outlier Detection

The last phase of the pipeline is to pass measures derived from peak extraction to outlier rejection (figure 3). This is applied at the level of the sliding window. If a window is declared an outlier, heart rate and heart rate variability measures are removed from the cleaned time series. By default, RapidHRV returns both the cleaned and the uncleaned time series. In addition to default parameters listed below, the package has optional arguments embedded to allow users to override these presets. Given that not all users may be entirely comfortable manually adjusting these, RapidHRV additionally contains semantically-labeled arguments as inputs for outlier constraints ('liberal', 'moderate' [default], and 'conservative'; corresponding parameters are parenthesized under *Biological Constraints* and *Statistical Constraints*).

Biological Constraints. RapidHRV first applies restrictions to exclude data that are highly unlikely given known physiology:

1. Screening for sufficient peaks in a window (default: number of peaks > (window width / 5) + 2), floored at 3; default at 10s = 3 peaks). This is primarily for computational applicability and efficacy, screening data prior to further processing. The minimum number of peaks required to enable calculation of RMSSD is 3. As such, this is also applied to the uncleaned time series.

2. Minimal and maximal heart rate ('moderate' [default]: $30 > \text{BPM} > 190$; 'liberal': $20 > \text{BPM} > 200$; 'conservative': $40 > \text{BPM} > 180$). These boundaries were based on typical heart rate at rest and during exercise in the healthy population (Pierpont & Voth, 2004; Sandvik et al., 1995; Savonen et al., 2006).
3. Minimal and maximal heart rate variability ('moderate' [default]: $5 > \text{RMSSD} > 262$; 'liberal': $0 > \text{RMSSD} > 300$; 'conservative': $10 > \text{RMSSD} > 200$). Default arguments correspond to the minimum/maximum 2nd/98th percentiles of resting RMSSD across ages 16-89 years (van den Berg et al., 2018).

Statistical constraints. RapidHRV next applies statistical constraints to account for noisy data that may otherwise appear to provide measures within the range of known physiology:

4. Median absolute deviation (MAD) of peak heights (distance from minimum value of signal in window; i.e. 0) and prominence (amplitude from baseline height; 'moderate' [default] = 5 MAD units; 'liberal' = 7; 'conservative' = 4). Unlike Z-scoring, this quantifies each peak's height and prominence in a given window in terms of its deviation from the median value in the same window (for a discussion of median absolute deviation see Leys et al., 2013). Applying these constraints to height and prominence helps exclude windows with noise-driven inaccuracies in peak detection.
5. Median absolute deviation of interbeat intervals ('moderate' [default] = 5 MAD units; 'liberal' = 7; 'conservative' = 4). This was also implemented to account for inaccuracies in peak detection, either where spiking may cause detection of an irrelevant peak shortly after e.g. an R wave, or low signal to noise ratio may result in missing relevant peaks.
6. Time from the first peak to the last peak does not recede 50% of the fixed window width. This is to ensure the user that the *actual* length of time for extracting HR/HRV is not less than half of that which is specified in the window width argument. Given debates surrounding adequacy of different window lengths for HRV extraction (Baek et al., 2015; Munoz et al., 2015), this was implemented

primarily as a theoretical constraint (rather than for just cleaning per se) to ensure the user is not provided data that deviated significantly from their specified window.

Analysis can be executed with one line of code, which returns a pandas DataFrame (McKinney, 2012; Reback et al., 2022) containing the analyzed data.

–Figure 3–

Visualization

To allow for selected manual inspection, we have also implemented optional interactive visualizations via matplotlib (Hunter, 2007) which allow the user to plot the time course of heart rate and heart rate variability. The user can then select and view specific data points to see the window of extraction (figure 4). We have provided an example pipeline below:

```
import rapidhrv as rhv
example_signal = rhv.get_example_data() # Load example data
preprocessed = rhv.preprocess(example_signal) # Preprocess data
analyzed = rhv.analyze(preprocessed) # Analyze data
rhv.visualize(analyzed) # Visualize data
```

–Figure 4–

Validation Methods

Datasets

To validate the above pipeline we subjected it to a series of tests across both simulated and real data (table 1). We first started by testing RapidHRV's estimations in two sets of simulated data (PPGSynth; Tang et al., 2020). Next, we ran validation in real data across successively noisier modalities: electrocardiography (ECG), finger infrared photoplethysmography (IR PPG), and wrist PPG data. Database information and code generated in validation tests are available through the open science framework (<https://osf.io/7zvn9/>).

–Table 1–

1. Simulations Across Noise and Sampling Rates. We first took the pipeline forward to validation using simulated photoplethysmography (PPG) data from PPGSynth (Tang et al., 2020) in MATLAB. This allowed us to test how well RapidHRV recovered known parameters under specified conditions, such as sampling rate and noise. We produced 5 minute long synthetic datasets (1000Hz), each of which varied according to the following cardiac features:

- Mean heart rate (BPM range 60-120, increments of 5). These were selected to allow us to check the sensitivity of the pipeline for detecting values across typical resting heart rate, as well as elevations of these values.
- Heart rate variability (RMSSD range 0-100, increments of 5). Again, these were selected based on typical resting heart rate variability and moderate increases/decreases.

Following simulation of the data, we introduced noise via:

- White gaussian noise filtering (signal to noise ratios = 0.01, 10, 20, 30, 40, 50, and 60dB). Here, we selected a range of values starting from near-zero to ascertain at which level of noise the pipeline could no longer recover parameters (with near-zero expected to prevent the extraction of any meaningful metrics).

- Downsampling (frequencies = 20, 50, 100, and 250Hz). These were selected: A) to capture the range of sampling rates currently used in photoplethysmography studies (typically $\geq 20\text{Hz}$); and B) because prior work has suggested sensitivity to RMSSD deteriorates dramatically between 20-100Hz (Choi & Shin, 2017). This offered the opportunity to test the effects of cleaning on what would otherwise be considered poor quality data.

2. Simulated ‘Anxiety’. Again, we simulated data using PPGSynth. However, simulations were based on prior anxiety literature to emulate experiment-specific effects. We wanted to validate the sensitivity of the pipeline across specified sampling rates and levels of noise but in the context of a known psychophysiological effect. For this, we used an estimated Cohen’s d of .384, which was derived from a previous within-subjects threat-of-shock study (RMSSD, $t_{25} = 1.96$, polarity flipped for readability; Gold et al., 2015). To reflect decisions in experimental design, we ran a power calculation using the ‘pwr’ package (Champely et al., 2017) in R, leading us to simulate a sample of $N = 171$ ($\alpha=.001$, $1-\beta=.95$). These ‘experiments’ were simulated 10 times to ascertain confidence intervals around estimated effect sizes.

For each ‘subject’, we generated a 5 minute simulated time series (1kHz) using typical resting heart rate and heart rate variability (BPM: $\mu = 74$, $\sigma = 13$, Savonen et al., 2006; RMSSD: $\mu = 23$, $\sigma = 7$, bounded between 5-262, Nunan et al., 2010). These simulations were intended to emulate the ‘safe’ condition (no anxiety induction) in a threat-of-shock study. Next, we simulated another time series for each subject that deviated from their ‘safe’ RMSSD (heart rate held constant) based on our effect size estimate ($d = .384$). This emulated our ‘threat’ (anxiety induction) condition and contained general reductions in heart rate variability. Each time series was finally submitted to downsampling and noise filtering using the same parameters as in the previous simulations (white gaussian noise filtering to 0.01, 10, 20, 30, 40, 50, 60dB; downsampling to 20, 50, 100, and 250Hz).

3. Estimation via ECG. Following simulations, we wanted to clarify that our package was able to adequately extract measures from one of the highest standards for heart rate variability recordings, ECG. Here, we tested whether our package was able to recapitulate known age-related effects of heart rate variability. For this, we used the *Fantasia* database (Iyengar et al., 1996). This consisted of 40 subjects (20 ‘Young’ Age Range = 21-34 years; 20 ‘Old’ Age Range = 68-81 years) watching the movie ‘Fantasia’ (duration \approx 2 hours) whilst undergoing ECG recordings. Automated peak detection was previously run on this dataset, with every beat annotation verified by visual inspection. The full dataset and description is available on PhysioNet (<https://physionet.org/content/fantasia/1.0.0/>; Goldberger et al. 2000).

4. Estimation via Wrist Photoplethysmography. We next took the pipeline forward to validation in a modality considered relatively noisier than ECG, wrist photoplethysmography. Here, we used a dataset of 39 subjects (Age: Mean = 22.67; Range = 18–38 years; demographics reported prior to N=1 exclusion) watching 2 x 28 minute blocks of documentary and horror video clips undergoing finger IR PPG recording (1000Hz, de Groot et al., 2020). This allowed us to contrast psychological conditions of the experiment, testing whether RapidHRV was able to detect effects of anxiety. Moreover, in the original study, data had been preprocessed and analyzed using a commercially available software (Labchart; ADInstruments, Sydney, Australia; analyzed using built-in ‘Peak Analysis’ module). This allowed us to benchmark RapidHRV against another software. The full dataset and description is available via the Open Science Framework (<https://osf.io/y76p2/>).

5. Estimation via PPG. Finally, we analyzed the *PPG-DaLiA* dataset, which consists of 15 subjects (Age: Mean = 30.60 years; Range = 21-55) completing various activities whilst having wrist PPG recorded with an Empatica E4 device (Hz = 64) and simultaneous ECG measures with a RespiBAN (Hz = 700, Reiss et al., 2019). Over the course of 2.5 hours, participants engaged in a range of activities designed to elicit low and high motion. These were: sitting and reading; ascending/descending stairs; 1 v 1 table soccer; cycling on pavements and gravel; driving a car; lunch break (queuing, purchasing, and eating food in a cafeteria); walking;

and working at a desk (typically on a computer). This allowed us to test whether RapidHRV was able to extract heart rate and variability measures from wrist PPG and how these compared to ECG measurements. Moreover, this enabled us to highlight under what conditions estimations are optimal. For the full dataset and description, see Reiss et al. (2019).

Analyses

Unless otherwise stated, all analyses were conducted using RapidHRV's default arguments: window width = 10s; window movement = 10s; outlier method = 'moderate' (peak/height median absolute deviation = 5, interbeat interval median absolute deviation = 5, BPM range = 30-190, RMSSD range = 5-262); minimum window successful extraction = 5s, minimum amplitude for peak detection = 50, minimum distance between peaks = 250ms; for ECG data, `ecg_prt_clustering = True`). To assess performance across datasets, we used: visualizations; intraclass correlation coefficients (ICC; two-way mixed effects, absolute agreement, single measure); root-mean-square-error (RMSE); and sensitivity to experimental effects (Cohen's *d*). For ICC values, we used the following semantic labels for interpretation: $ICC < .5$ as 'poor', $.5 < ICC < .75$ as 'moderate'; $.75 < ICC < .90$ as 'good', and $.90 < ICC$ as 'excellent' (Koo & Li, 2016).

In our PPG dataset, we also calculated motion estimates as a proxy for the severity of noise present in PPG across conditions. We derived mean jerk magnitude (square root of the sum of squared changes in acceleration; Eager et al., 2016), which aimed to pick up on oscillatory ('jerk') motions in the wrist for each condition. Across all axes (x, y, and z; m/s^2), jerk vector, \vec{j} , was calculated as the change in acceleration, \vec{a} , divided by the change in time, t , between samples:

$$\vec{j}(t) = \frac{\Delta \vec{a}(t)}{\Delta t}$$

Finally, we derived jerk magnitudes by calculating the square root of the sum of squared jerks, j , for axes x, y, and z:

$$\left| \vec{j} \right| = \sqrt{j_x^2 + j_y^2 + j_z^2}$$

For each subject, jerk magnitudes were averaged across time for each condition.

Validation Results

1. Parameter Recovery in Simulated PPG

RapidHRV was able to accurately recover heart rate across most sampling frequencies and noise in our initial simulations. Accurate detection of BPM primarily started to degrade when signal to noise ratios were less than 10dB (figure 5; table 2). RapidHRV cleaning provided improvements in simulations with a signal to noise ratio of 10dB.

–Table 2–

Performance in recovery of heart rate variability was again primarily based on signal to noise ratio. At 20dB RapidHRV recovery of RMSSD was good-to-excellent for higher sampling rates ($\geq 100\text{Hz}$), whereas lower sampling rates ($< 100\text{Hz}$) required slightly lower levels of noise ($> 30\text{dB}$) for excellent recovery. RapidHRV cleaning provided clear improvements when signal to noise ratio was below 30dB (figure 3).

–Figure 5–

Key points

- RapidHRV’s estimation of simulated BPM was good-to-excellent when signal to noise ratio was 10dB or above.

- RapidHRV's estimation of simulated RMSSD was good-to-excellent when signal to noise ratio was 20dB and sampling rate was 100Hz or higher. Across all sampling rates, RMSSD recovery was excellent at 30dB or above.
- Cleaning appeared particularly helpful when noise was high and/or sampling rate was low.

2. Sensitivity to 'Anxiety' in Simulated PPG

RapidHRV's ability to estimate simulated effects was again primarily impacted by the level of SNR, where changes in RMSSD were not reliably detected at 0.01 and 10dB (figure 6; table 3). Effects were detected at signal to noise ratios of 20dB and 30dB, but this was estimated at around half that of the true value. Maximal effect sizes plateaued following 40dB for uncleaned data and 30dB for cleaned data. Moreover, reliable detection of effects in some scenarios (e.g. 50Hz, 10dB) appeared dependent on cleaning. This was consistent with our previous validation results, in that cleaning was beneficial at lower signal to noise ratios and sampling rates, but was not necessary (or could be relaxed) in cleaner data.

–Figure 6–

–Table 3–

Key points

- RapidHRV's estimation for simulated effects of Anxiety on RMSSD was excellent when signal to noise ratio was 30dB or above. Moreover, this estimation was robust at a low sampling rate (i.e. 20Hz).
- Cleaning was beneficial when noise was high, but was not necessary when noise was low.

3. Estimation via ECG

As expected, visual inspection suggested ECG data to hold a high signal to noise ratio. In line with our simulation findings, we adjusted the outlier rejection method accordingly so as not to excessively exclude data (i.e. ‘liberal’). The dataset had previously been analyzed using an automated peak detection algorithm, with every beat annotation verified by visual inspection (Iyengar et al., 1996). Subject-specific heart rate estimates were not available from the original database. However, RapidHRV heart rate estimations were able to recapitulate sample-wide summary statistics (figure 7; table 4).

–Table 4–

The analytical method used for extraction of heart rate variability in the original database (power spectral density analysis) was inconsistent with RapidHRV's time-domain heart rate variability measure (RMSSD). Despite this discrepancy, RapidHRV was able to capture previously reported effects of age on heart rate variability (Iyengar et al., 1996; estimated Cohen's d of short-term heart rate variability (i.e. α_s) ≈ 1.32), such that—in cleaned data—younger participants demonstrated higher RMSSD ($M = 58.94$, $SD = 28.82$) than older participants ($M = 25.54$, $SD = 13.02$, Cohen's $d = 1.49$; figure 6). Effects were not apparent in the uncleaned data (Cohen's $d = -0.31$).

–Figure 7–

Key points

- In a movie-watching ECG dataset, RapidHRV was able to recapitulate previously reported summary statistics of BPM.
- RapidHRV was able to reproduce previously reported effects of age on RMSSD as measured by ECG.
- Cleaning was vital for detecting effects of age on RMSSD in ECG.

4. Estimation via Finger IR PPG

Sensitivity to anxiety

In our finger IR PPG data, RapidHRV was able to capture previously reported (de Groot et al., 2020) effects of anxiety on BPM (table 5). RapidHRV additionally demonstrated an influence of anxiety on RMSSD. Effects on BPM were greater following cleaning, whereas detection of effects on RMSSD was entirely dependent on cleaning.

–Table 5–

Benchmarking

Overall, there was excellent agreement between RapidHRV and previous estimates (de Groot et al., 2020; implemented using LabChart, ADInstruments, Sydney, Australia) of BPM (ICC > .99; figure 8). For heart rate variability, there was good agreement between the two when using the cleaned time series (ICC = .89), but poor agreement when using the uncleaned time series (ICC = .32).

–Figure 8–

Key points

- RapidHRV was able to recapitulate previously reported effects of anxiety on BPM and RMSSD as measured by finger IR PPG.
- Cleaning was vital to the detection of effects of anxiety on RMSSD.

5. Estimation via Wrist-PPG During Activities

We first derived motion estimates (mean jerk magnitude; Eager et al., 2016) as a proxy for the severity of noise present in PPG. This enabled us to split conditions into low (Reading, Working, Lunch Break, Driving) and high motion activities (Table Soccer, Stairs, Walking, Cycling; figure 9).

–Figure 9–

In our initial analysis using default arguments, RapidHRV was not able to produce estimates across many of the activities for both cleaned and uncleaned data (conditions estimated: $M = 22.5\%$, $SD = 5.18\%$). Visual inspection confirmed this was due to high levels of noise and variable peak amplitude present in the signal. Therefore, we adjusted window movement to 1s (to increase the number of extraction windows in each condition), reduced minimum amplitude threshold for peak detection (30), and—in line with our simulation results above (figure 3)—tightened outlier rejection (‘conservative’). Following this, RapidHRV was able to produce estimates for all conditions in the uncleaned data and across most activities for cleaned data (conditions estimated: $M = 87.5\%$, $SD = 11.57\%$). Missing estimates from the latter were predominantly limited to high motion activities (1 missing estimate for a single subject in the driving condition).

Across low motion activities, RapidHRV PPG-based estimates converged with simultaneous (visually verified) ECG measures (Reiss et al., 2019). Agreement with low motion

activities was excellent for BPM ($ICC > .90$) and agreement with RMSSD was moderate-to-good ($.57 < ICC < .82$; table 6; figure 10). Under high motion conditions, heart rate and heart rate variability estimates showed poor agreement ($ICC < .32$), except for heart rate within the cycling condition ($ICC = .75$).

–Table 6–

–Figure 10–

Key point

- Following calibration to the data, cleaned RapidHRV wrist PPG-based estimates demonstrated moderate-to-excellent convergence with a simultaneous ECG-based analysis whilst under low motion conditions. Agreement was generally poor under high motion conditions.

Discussion

RapidHRV is an open-source toolbox for extracting heart rate and heart rate variability measures. RapidHRV was developed in response to the need for software dedicated to dealing with extensive cardiac data collected across large time frames, such as out-of-laboratory PPG recordings, which may require point estimates from very short time windows (~10 seconds). Python packages currently exist which can analyze cardiac data (e.g. *Systole*, Legrand & Allen, 2020; *NeuroKit2*, Pham et al., 2021; *pyHRV*, Gomes et al., 2019). However, outlier rejection algorithms often require visual inspection and/or extensive scripting on the user's end. While suitable for the cardiac data collected during laboratory experiments, this may not be feasible when dealing with data collected across large time-scales, such as weeks or months. Here, we

have attempted to fill this gap by developing a programmatically easy-to-use toolbox which extracts HRV measures from ultra-short windows and automates artifact detection and rejection. In general, this is applied via a series of biological and statistical constraints. Moreover, for ECG data, we have also implemented a k-means clustering algorithm for delineating P, R, and T waves. Across simulated and real datasets, we scrutinized RapidHRV, testing scenarios where it was and was not able to extract meaningful metrics. We show that signal to noise ratio, sampling rate, and recording modality had a clear impact on sensitivity of estimation. Here, we summarize these validation tests and make modality-specific recommendations for users.

Simulations

Within simulated data, RapidHRV was able to recover heart rate across most levels of noise (white gaussian noise filter $\geq 10\text{dB}$), even at relatively low sampling rates ($\geq 20\text{Hz}$). RapidHRV's recovery of heart rate variability was excellent at relatively low levels of signal to noise ratio ($\geq 20\text{dB}$), though there was degradation of performance as sampling rate decreased. Additional simulations of cardiac responses to an anxiety induction demonstrated RapidHRV estimations fully captured effects at moderate levels of noise ($\geq 30\text{dB}$) even at relatively low sampling rates (i.e. 20Hz). RapidHRV was able to partially capture effects ($\sim 50\%$ reduction in effect size) at very high levels of noise ($\geq 10\text{dB}$ when $\text{Hz} > 50$). Simulations revealed RapidHRV cleaning was particularly beneficial at lower sampling rates and higher levels of noise, but was not necessary (or could be relaxed) when signal and sampling rates were high. Moreover, these simulations were able to clarify the validity of RapidHRV's default window (10s) for estimation of heart rate variability across a longer time period (i.e. 5 minutes).

Electrocardiography

In our electrocardiography analyses, we were able to recapitulate previously reported effects of age on heart rate and heart rate variability. In line with previous analyses (Iyengar et al., 1996), RapidHRV-estimates suggested older participants had lower heart rate variability than younger participants during movie-watching. Importantly, cleaning was vital to this detection.

Finger PPG

Using RapidHRV-estimates, we noted effects of anxiety on heart rate and heart rate variability in a database of participants watching horror and documentary videos while undergoing finger infrared PPG recordings. Notably, the estimated effect size was analogous to that noted in threat-of-shock studies (Gold et al., 2015). Moreover, when contrasting subject-specific estimates, we found good-to-excellent agreement between RapidHRV and a previous analysis using a commercially available software. Effect sizes between conditions and convergence of estimates between softwares was significantly improved following RapidHRV cleaning.

Wrist PPG

In our wrist PPG validation, we noted RapidHRV was not able to produce estimates for the majority of conditions due to limitations of the default arguments. Despite applying scaling (0-100) at the level of a sliding window, PPG signals showed variable peaks amplitudes. Following alterations [`rhv.analyze(preprocessed, outlier_detection_settings="conservative", amplitude_threshold=30, window_overlap=9)`], RapidHRV-estimates during low motion activities (e.g. reading, eating lunch) demonstrated good-to-excellent agreement with a manually-verified analysis of ECG data. During high wrist-movement activities (e.g. table soccer), estimates were generally poor-to-moderate. We do note however that in one of the high motion conditions, cycling, RapidHRV-estimation of heart rate was good (though heart rate variability estimation was poor). This may reflect our proxy for motion-related artifacts (jerk magnitude), which may not always be as good an estimator of artifacts during activities which involve changes in acceleration but relatively low movements in the wrists (i.e. cycling, as hands are gripping the handle bars). Future work should seek to correlate RapidHRV quality assurance metrics with a wider range of motion estimation methods. Overall, we found adjustments to parameter arguments beneficial to PPG data, but we also note there are contextual factors and limitations (i.e. motion-related artifacts) influencing the feasibility of accurate estimation.

Overall User Recommendations

For ECG data, users may find traditional, amplitude-based analyses will not work for subjects who demonstrate atypical signal morphologies (e.g. particularly prominent P and T waves). RapidHRV includes the use of k-means clustering to help discern these components of the ECG signal, though this is not enabled by default (`ecg_prt_clustering = False`). Additionally, the cleanliness of signal, namely the stability of peak prominences, means the data may be low in artifacts, and that minor deviations could be detected as outliers. As such, when dealing with already-clean data, users may find that outlier rejection can be omitted or relaxed (e.g. `outlier method = 'liberal'`).

Results from the finger IR PPG data used in the present study did not suggest the need for alterations to default RapidHRV arguments, but did suggest that automated cleaning should be used.

Lastly, PPG data collected from naturalistic settings is typically low in signal to noise ratio, which can constrain peak detection. Consequently, lowering the minimum amplitude threshold for peak detection and decreasing window movement may help improve extraction. Furthermore, given the large amount of motion-related artifacts and the results from our simulation analyses, we recommend: a) the use of relatively conservative cleaning (e.g. `outlier method = 'conservative'`), and b) inspection of motion across conditions as an indicator of estimation accuracy (table 7).

–Table 7–

Conclusion

In the present paper, we have outlined RapidHRV: an open-source Python pipeline for the estimation of heart rate and heart rate variability. Across simulated datasets, RapidHRV showed good-to-excellent recovery of heart rate and heart rate variability at relatively high levels of noise. Estimates in electrocardiography and finger IR PPG demonstrated RapidHRV was able to

recapitulate known effects of age and anxiety, and showed excellent agreement with visually-inspected analyses and commercial software. Lastly, performance in wrist photoplethysmography data was good-to-excellent when participants were engaged in low motion activities, but we noted poor-to-moderate estimations when motion was high. Given the increased interest in the use of wearable measures of heart rate metrics and how they relate to other domains such as mental health, we hope that this toolbox will be of wide use to the community, and that the simulation and benchmarking tests provided may help inform the design and analysis of such studies.

Acknowledgements

Thank you to Kaarina Aho for statistical consultation; Nicolas Legrand for allowing us to use *Systole*'s code for deriving high frequency power; and Russell Kirk for helping derive motion estimates from smartwatch data. Appreciation goes to other open-source analysis software, namely *Systole* (Legrand & Allen, 2020) and *HeartPy* (van Gent et al., 2019), which helped inspire the development of RapidHRV.

References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- van den Berg, M. E., Rijnbeek, P. R., Niemeijer, M. N., Hofman, A., van Herpen, G., Bots, M. L., Hillege, H., Swenne, C. A., Eijgelsheim, M., Stricker, B. H., & Kors, J. A. (2018). Normal Values of Corrected Heart-Rate Variability in 10-Second Electrocardiograms for All Ages. *Frontiers in Physiology*, 9. <https://doi.org/10.3389/fphys.2018.00424>
- Biagetti, G., Crippa, P., Falaschetti, L., Saraceni, L., Tiranti, A., & Turchetti, C. (2020). Dataset from PPG wireless sensor for activity monitoring. *Data in Brief*, 29, 105044. <https://doi.org/10.1016/j.dib.2019.105044>
- Boccignone, G., Conte, D., Cuculo, V., D'Amelio, A., Grossi, G., & Lanzarotti, R. (2020). An Open Framework for Remote-PPG Methods and Their Assessment. *IEEE Access*, 8, 216083–216103. <https://doi.org/10.1109/ACCESS.2020.3040936>
- Caizzone, A., Boukhayma, A., & Enz, C. (2017). Comprehensive noise analysis in PPG read-out chains. *2017 International Conference on Noise and Fluctuations (ICNF)*, 1–4. <https://doi.org/10.1109/ICNF.2017.7985971>
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & Rosario, H. D. (2017). *pwr: Basic functions for power analysis*. <https://nyuscholars.nyu.edu/en/publications/pwr-basic-functions-for-power-analysis>
- Choi, A., & Shin, H. (2017). Photoplethysmography sampling frequency: Pilot assessment of how low can we go to analyze pulse rate variability with reliability? *Physiological Measurement*, 38(3), 586–600. <https://doi.org/10.1088/1361-6579/aa5efa>
- Eager, D., Pendrill, A.-M., & Reistad, N. (2016). Beyond velocity and acceleration: Jerk, snap and higher derivatives. *European Journal of Physics*, 37(6), 065008. <https://doi.org/10.1088/0143-0807/37/6/065008>
- Georgiou, K., Larentzakis, A., Khamis, N., Alsuhaibani, G., Alaska, Y., & Giallafos, E. (2018). Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. *Folia Medica*, 60. <https://doi.org/10.2478/folmed-2018-0012>

- Gold, A. L., Morey, R. A., & McCarthy, G. (2015). Amygdala–Prefrontal Cortex Functional Connectivity During Threat-Induced Anxiety and Goal Distraction. *Biological Psychiatry*, 77(4), 394–403. <https://doi.org/10.1016/j.biopsych.2014.03.030>
- Goldberger Ary L., Amaral Luis A. N., Glass Leon, Hausdorff Jeffrey M., Ivanov Plamen Ch., Mark Roger G., Mietus Joseph E., Moody George B., Peng Chung-Kang, & Stanley H. Eugene. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23), e215–e220. <https://doi.org/10.1161/01.CIR.101.23.e215>
- de Groot, J. H. B., Kirk, P. A., & Gottfried, J. A. (2020). Encoding fear intensity in human sweat. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1800), 20190271. <https://doi.org/10.1098/rstb.2019.0271>
- P. Gomes, P. Margaritoff, & H. P. da Silva (2019). *pyHRV: Development and evaluation of an open-source python toolbox for heart rate variability (HRV)*. *Proc. Int'l Conf. on Electrical, Electronic and Computing Engineering (IcETRAN)*, 822-828.
- Hillebrand, S., Gast, K. B., de Mutsert, R., Swenne, C. A., Jukema, J. W., Middeldorp, S., Rosendaal, F. R., & Dekkers, O. M. (2013). Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose–response meta-regression. *EP Europace*, 15(5), 742–749. <https://doi.org/10.1093/europace/eus341>
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Iyengar, N., Peng, C. K., Morin, R., Goldberger, A. L., & Lipsitz, L. A. (1996). Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 271(4), R1078–R1084. <https://doi.org/10.1152/ajpregu.1996.271.4.R1078>
- Jandackova, V. K., Britton, A., Malik, M., & Steptoe, A. (2016). Heart rate variability and depressive symptoms: A cross-lagged analysis over a 10-year period in the Whitehall II study. *Psychological Medicine*, 46(10), 2121–2131. <https://doi.org/10.1017/S003329171600060X>

- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lee, J., Scott, D. J., Villarroel, M., Clifford, G. D., Saeed, M., & Mark, R. G. (2011). Open-access MIMIC-II database for intensive care research. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 8315–8318. <https://doi.org/10.1109/IEMBS.2011.6092050>
- Legrand, N., & Allen, M. (2020). Systole: A python toolbox for preprocessing, analyzing, and synchronizing cardiac data. Retrieved from <https://github.com/emodied-computation-group/systole>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Li, B. N., Dong, M. C., & Vai, M. I. (2010). On an automatic delineator for arterial blood pressure waveforms. *Biomedical Signal Processing and Control, 5*(1), 76–81. <https://doi.org/10.1016/j.bspc.2009.06.002>
- Makovac, E., Meeten, F., Watson, D. R., Garfinkel, S. N., Critchley, H. D., & Ottaviani, C. (2016a). Neurostructural abnormalities associated with axes of emotion dysregulation in generalized anxiety. *NeuroImage: Clinical, 10*, 172–181. <https://doi.org/10.1016/j.nicl.2015.11.022>
- Makovac, E., Watson, D. R., Meeten, F., Garfinkel, S. N., Cercignani, M., Critchley, H. D., & Ottaviani, C. (2016b). Amygdala functional connectivity as a longitudinal biomarker of symptom changes in generalized anxiety. *Social Cognitive and Affective Neuroscience, 11*(11), 1719–1728. <https://doi.org/10.1093/scan/nsw091>
- McKinney W. (2012). *pandas: a foundational python library for data analysis and statistics*. San Francisco: O'Reilly Media, Inc.
- Mulcahy, J. S., Larsson, D. E. O., Garfinkel, S. N., & Critchley, H. D. (2019). Heart rate variability as a biomarker in health and affective disorders: A perspective on

- neuroimaging studies. *NeuroImage*, 202, 116072.
<https://doi.org/10.1016/j.neuroimage.2019.116072>
- Munoz, M. L., van Roon, A., Riese, H., Thio, C., Oostenbroek, E., Westrik, I., de Geus, E. J. C., Gansevoort, R., Lefrandt, J., Nolte, I. M., & Snieder, H. (2015). Validity of (Ultra-)Short Recordings for Heart Rate Variability Measurements. *PLOS ONE*, 10(9), e0138921.
<https://doi.org/10.1371/journal.pone.0138921>
- Nunan, D., Sandercock, G. R. H., & Brodie, D. A. (2010). A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults. *Pacing and Clinical Electrophysiology*, 33(11), 1407–1417.
<https://doi.org/10.1111/j.1540-8159.2010.02841.x>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Pham, T., Lau, Z. J., Chen, S. H. A., & Makowski, D. (2021). Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial. *Sensors*, 21(12), 3998.
<https://doi.org/10.3390/s21123998>
- Pierpont, G. L., & Voth, E. J. (2004). Assessing autonomic function by analysis of heart rate recovery from exercise in healthy subjects. *The American Journal of Cardiology*, 94(1), 64–68. <https://doi.org/10.1016/j.amjcard.2004.03.032>
- Pimentel, M. A. F., Johnson, A. E. W., Charlton, P. H., Birrenkott, D., Watkinson, P. J., Tarassenko, L., & Clifton, D. A. (2017). Toward a Robust Estimation of Respiratory Rate From Pulse Oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8), 1914–1923. <https://doi.org/10.1109/TBME.2016.2613124>
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Rae, C. L., Ahmad, A., Larsson, D. E. O., Silva, M., Praag, C. D. G. van, Garfinkel, S. N., & Critchley, H. D. (2020). Impact of cardiac interoception cues and confidence on voluntary

- decisions to make or withhold action in an intentional inhibition task. *Scientific Reports*, *10*(1), 4184. <https://doi.org/10.1038/s41598-020-60405-8>
- Reback, J., Jbrockmendel, McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Hawkins, S., Roeschke, M., Gfyoung, Sinhrks, Klein, A., Hoefler, P., Terji Petersen, Tratner, J., She, C., Ayd, W., Naveh, S., JHM Darbyshire, Garcia, M., ... Skipper Seabold. (2022). pandas-dev/pandas: Pandas 1.4.1 (v1.4.1) [Computer software]. Zenodo. <https://doi.org/10.5281/ZENODO.3509134>
- Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors*, *19*(14), 3079. <https://doi.org/10.3390/s19143079>
- Sandvik, L., Erikssen, J., Ellestad, M., Erikssen, G., Thaulow, E., Mundal, R., & Rodahl, K. (1995). Heart rate increase and maximal heart rate during exercise as predictors of cardiovascular mortality: A 16-year follow-up study of 1960 healthy men. *Coronary Artery Disease*, *6*(8), 667–679. <https://doi.org/10.1097/00019501-199508000-00012>
- Savonen, K. P., Lakka, T. A., Laukkanen, J. A., Halonen, P. M., Rauramaa, T. H., Salonen, J. T., & Rauramaa, R. (2006). Heart rate response during exercise test and cardiovascular mortality in middle-aged men. *European Heart Journal*, *27*(5), 582–588. <https://doi.org/10.1093/eurheartj/ehi708>
- Tang, Q., Chen, Z., Allen, J., Alian, A., Menon, C., Ward, R., & Elgendi, M. (2020). PPGSynth: An Innovative Toolbox for Synthesizing Regular and Irregular Photoplethysmography Waveforms. *Frontiers in Medicine*, *7*. <https://doi.org/10.3389/fmed.2020.597774>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>