Design rationale and performance evaluation of Wavelet Health Wristband: bench-top validation of a wrist-worn physiological signal recorder

Onur Dur PhD\textsuperscript{1}, Colleen Rhoades MS\textsuperscript{1}, Sally Ng MS\textsuperscript{1}, Ragwa Elsayed, MS\textsuperscript{2}, Reinier van Mourik PhD\textsuperscript{1}, Maulik Majmudar, MD\textsuperscript{3}

\textsuperscript{1}Wavelet Health, Mountain View, CA
\textsuperscript{2}Department of Biomedical Engineering, San Jose State University, San Jose, CA,
\textsuperscript{3}Healthcare Transformation Lab, Massachusetts General Hospital, Harvard Medical School, Boston, MA

*Corresponding Author:*

Onur Dur, PhD
Wavelet Health
465 Fairchild Drive Suite 228
Mountain View, CA 94043
onur@wavelethealth.com
Phone: 650.441.0397

Word count: 6549

Key words: wearable devices, digital health, remote monitoring, PPG, validation
Abstract

Background: Wearable and connected health devices along with the recent advances in mobile and cloud computing provide a continuous, convenient-to-patient and scalable way to collect personal health data remotely. The Wavelet Health Platform and the Wavelet Wristband have been developed to capture multiple physiological signals and to derive biometrics from these signals including resting heart rate, heart rate variability, and respiration rate.

Objective: This study aims to evaluate the accuracy of the biometrics estimates and signal quality of the wristband.

Methods: Measurements collected from 35 subjects using the Wavelet Wristband were compared with simultaneously recorded electrocardiogram and spirometry measurements.

Results: The heart rate, heart rate variability (SDNN) and respiration rate estimates matched within 0.6 ± 0.9 bpm, 7 ± 10 ms and 1 ± 1 brpm mean absolute deviation of the reference measurements, respectively. The quality of the raw plethysmography signal collected by the wristband, as determined by the harmonic-to-noise ratio, was comparable to that obtained from measurements from a finger-clip plethysmography device.

Conclusion: The accuracy of the biometrics estimates and high signal quality indicate that the Wristband PPG device is suitable for performing pulse wave analysis and measuring vital signs.
Introduction

Wearable and connected health devices have been increasingly popular for objectively measuring personal health and wellness by the general population; over 300 million wearable device purchases were reported in 2017 worldwide [1]. Not surprisingly, wearables have recently been adopted by clinical researchers to paint the most complete picture of patients’ health and well-being by offering continuous, long term and multi-parametric monitoring [2]. Leveraging mobile, wearable and connected devices, cloud computing and machine learning algorithms, the modern era of healthcare - known as digital health - promises new and better ways to screen, diagnose, manage and treat patients; thereby to improve health-related outcomes and decrease cost of healthcare delivery. It's well known that measuring health is the key to (i) improve screening and preventative methods, (ii) optimize disease management, drug titration and adherence to therapy, (iii) prevent hospitalization and reduce adverse events. In an era where financial incentives are becoming more aligned to improve outcomes pre- and post-operatively, monitoring patients' health in the largely opaque time period where patients are outside of hospital settings is increasingly valuable. Diverting patients early to the clinic or therapy in lieu of hospital readmission promises better outcomes for patients, reduces adverse events, and aligns with the financial incentives of the providers. Wearable and connected devices offer a promising solution to streamline health data collection needs and address the lacking personalized oversight in the current health system.

Clinical research organizations and pharmaceutical companies have already started embracing digital health. Recently several large-scale research initiatives involving parties across government, health institutions, private technology and pharmaceutical companies [3-5] have announced collecting orthogonal bio-psycho-social signals and biometrics derived from these signals in longitudinal studies to understand underlying risk factors leading to disease and hospitalizations. There is also a growing interest in the pharma community to use mobile, wearable and connected health technologies to improve the operational efficiency during the product development cycle, and also to accelerate bench-to-bedside translational science. Digital operational efficiency here refers to use of digital health to increase patient adherence, trial data collection speed and efficiency, thereby reducing the time and cost to market. Electronic patient reported outcomes (ePROs) and remote patient monitoring (RPM) are good examples of how digital health has been transforming the operational workflow of biopharma and clinical research organizations (CROs) [6]. Digital biomarker development has also been a key application of digital health in the biopharma and clinical research spaces for developing new endpoints that have not previously been possible to assess, or existing endpoints that can be measured in new and possibly better ways [7]. Recently, Green et al. has reported a digital biomarker for detecting patients with obstructive hypertrophic cardiomyopathy using signals collected from a wrist-worn wearable device and machine learning algorithms [8]. Other example applications in therapeutic areas including atrial fibrillation [9], and obstructive sleep apnea [10] can potentially be used for evaluating safety and efficacy of the drugs and interventions, screening and referring patients who can benefit from life-saving therapies, and also as a companion tool for dose titration, decision support and disease management. In summary, wearable and connected device technologies are evolving rapidly and may soon
become the new standard in health monitoring to improve and accelerate the way drugs are developed, new therapies are identified and patients are cared for.

This new era of personalized digital medicine requires devices that are convenient and engaging to the patient, and have validated accuracy to enable streamlined collection of biosignals and clinically meaningful health metrics. Many of the currently available remote monitoring options, including electrocardiography (ECG) based Holter devices, event recorders or mobile cardiac telemetry devices, require frequent replacements to extend utility beyond 7 to 14 days and often cause discomfort to the patient or complications due to their intrusive nature. There is consensus on neither the benefit of continuous monitoring with implantable loop recorders on patient outcomes nor on whether these invasive devices are suitable for large scale population level disease surveillance and health screening.

Advances in semiconductor and sensory technologies, in particular wide adoption of near infrared light spectroscopy in consumer health and clinical applications along with mobile and cloud computing have generated a new class of wearable devices for health data collection and monitoring. Wavelet Health (Mountain View, California) is the first to develop a full-stack wearable and connected device platform designed specifically to address the data collection needs of health researchers (Figure 1). The key design requirements are given below.

1) High quality signals and validated computed measures (biometrics)
2) Continuous longitudinal data
3) Accessible raw data and computed biometrics
4) Configurable device settings and mobile applications
5) Multiple device connectivity
6) Wide range of physiological measures
7) 3rd party device integration
8) Consumer-centric engaging experience
9) Patient compliance monitoring
10) Data privacy and safety

Figure 1 The Wavelet Health Platform for streamlined health data collection comprises easy-to-use wearable and connected devices to capture data; mobile and tablet software applications to
collect and transmit data; cloud servers, algorithms and web interface to analyze, store and access data.

To enable ubiquitous, passive, contextually rich health data collection, multiple wearable devices have been developed under the Wavelet Health Platform. This paper focuses on the performance evaluation of the Wavelet Wristband, which is a wrist-worn photoplethysmography (PPG) device. PPG is an optical technique for detecting blood volume changes within the microvascular bed in order to estimate physiological parameters [11]. PPG has been used since the early 1960s, particularly for pulse oximetry which is the standard-of-care tool for measuring peripheral arterial oxygen saturation (SpO₂) [12]. By positioning a light sensor and a light-emitting diode (LED) on the same plane, i.e., the reflectance type sensor configuration, wrist-worn PPG devices can perform measurements from the skin surface [13]. Light emitted by the LEDs into the wrist is mostly absorbed by the underlying tissue. The reflected light is captured by a photodiode, which is sampled many times a second to construct the PPG signal. The absorption of light varies with the changes in pulsatile arterial blood flow and generates a time-varying pulse waveform [12]. This signal can be recorded, transmitted, and used as a non-invasive longitudinal measurement of the underlying blood volume changes. Due to its good tissue penetration characteristics [14], PPG devices often rely on infrared (IR) light for estimating the relative volumetric changes in microvascular bed due to pulsatile blood flow. Using the IR signal, several biometrics including arterial pulsatility, heart rate (HR), heart rate variability (HRV), respiration rate (RR), vascular tone along with others related to the cardiovascular and autonomic nervous systems can be computed non-invasively [14]. When paired with the IR signal, the red light signal enables estimating SpO₂ which is known as pulse oximetry.

Measurement of vital signs enables detection and monitoring of a large number of conditions and diseases. Continuous HR monitoring is critical to the management of cardiovascular disease as elevated HR is an independent predictor of cardiovascular events, mortality and hospitalization for worsening heart failure [16, 17]. HRV is another clinically important metric often regarded as a measure of neurocardiac function and homeostasis [18]. Fluctuations in beat-to-beat timing arise from the interaction of different physiological systems, including heart, brain and autonomic nervous system in healthy and disease states. A relatively high resting HRV is indicative of a healthy, resilient, and responsive nervous system regulating the heart's activity, whereas reduced HRV indicates unbalanced sympathetic and parasympathetic activity negatively affecting cognitive performance [19] and physical training capacity [20]. Studies showed that lowered HRV is also a strong, independent predictor of health problems and increased risk of mortality [21-24]. Low HRV is also associated with congestive heart failure [25, 26], multiple sclerosis [27], Guillain-Barre syndrome [28], and diabetic neuropathy [29].

RR, an often overlooked vital sign, enables early detection of life threatening disease such as sleep apnea [30], pneumonia [31], sudden infant death syndrome [32, 33], or chronic obstructive pulmonary disease [34]. As abnormal RR is predictive of a future critical illness [35, 36] continuous monitoring would provide clinicians with a real time indicator of their patient’s health. Current methods for measuring RR include manually counting chest movements, estimated RR
from ECG patch devices, a spirometer, and capnography monitors [37]. These methods do not allow long term continuous monitoring due to the need for manual supervision, cost, or discomfort to the patient. Alternatively, RR can also be estimated from PPG devices by leveraging three signal processing methods: baseline wander, amplitude modulation and frequency modulation - each stem from three main physiological mechanisms: changes in tissue blood volume, stroke volume and respiratory sinus arrhythmia, respectively, caused by intrathoracic pressure changes during respiration [38-40]. Several PPG-based RR algorithms have been reported in the literature with variable levels of accuracy, i.e., mean error ~ 1-6 breaths per minute (brpm), depending on which one or combination of signal processing methods incorporated in the algorithm formulation [41, 42]. Previous work by Birrenkott et al. demonstrated the importance of establishing proper signal qualification methods to achieve accurate respiration rate estimations from PPG signal [43]. Here, we present the validation of a PPG-based RR estimation algorithm that combines the frequency modulation and baseline wander methods along with threshold-based respiratory signal qualification.

The aim of the study is to evaluate the accuracy of resting HR, HRV and RR estimates of the Wavelet Wristband and benchmark its performance as a wrist-worn PPG device compared with gold standard reference devices. In addition, the signal quality of the Wavelet Wristband is compared with a reference device.
Methods

Wristband Technology

The core technology of the Wavelet Wristband relies on PPG, i.e., non-invasive optical sensing of changes in arterial pulse volume, as described previously. During the measurement the wristband is placed snug on the arm above the wrist bone. The wristband comes with a removable sensor carriage and a plastic band as shown in Figure 2. The sensor carriage contains LEDs of two wavelengths and an optical sensor along with a battery and an inductive charging coil. Fully integrated analog front end receives and digitizes PPG signal along with the sub-millisecond resolution low jitter external clock signal. The wristband is also capable of collecting inertial motion data using the 3-axis accelerometer and 3-axis gyroscope built into the sensor carriage. The sampling rate and duty cycle of the light and motion sensors are configurable. For this validation study light sensor data is collected at 86 Hz and the motion data is collected at 10 Hz. Raw PPG and motion signal data collected from the wrist is transferred by the mobile application to the cloud server where it is processed by signal processing and machine learning algorithms.

Figure 2 The Wavelet Wristband is configurable photoplethysmography (PPG) and motion sensing device placed on the wrist to collect pulse wave signal, physiological measures computed from the signal along with physical activity. The sensor carriage can be removed from the plastic band.
The proprietary algorithms developed for extraction of HR, HRV, and RR from the PPG signal collected from wrist were described herein. The core technology leverages continuous wavelet transforms and adaptive time-frequency domain methods for detection of each heart beat with its salient morphological features [44, 45]. To eliminate false readings and poor signal quality related to motion or light artifacts, the beat-to-beat signal quality is evaluated using heuristics based on short-time Fourier Transform (stFT) of light sensor and accelerometer data, along with the harmonic to noise ratio (HNR) as described in the Signal Quality Analysis and Statistical Methods section. Once the PPG signal is qualified and decomposed to its static (DC) and pulsatile (AC) components, HR, HRV and RR are computed on 60-second non-overlapping windows of the AC signal. HR and RR algorithms return a single average value for the 60-second window duration. HRV is estimated using the standard deviation of the normal-normal (interbeat) intervals (SDNN) and also the root mean square of the successive differences between adjacent interbeat intervals (RMSSD) over 60-second non-overlapping windows [46]. Biometrics-specific heuristics such as HR and HRV confidence intervals (HRCI, HRVCI) further qualify whether the algorithm returns the estimated values or not report any valid measurements. RR is extracted by both time-domain algorithm analysis of the baseline wander, and the frequency-domain algorithm analysis of frequency modulation. The time-domain algorithm leverages a continuous wavelet transform to detect inspiration and expiration phases and measure the RR rate [47]. The frequency domain method also leverages multiple continuous wavelet transformations along with a dynamic programming ridge detection scheme to estimate respiration rate [48]. Finally, the algorithm reports the weighted average of the two RR estimates if the absolute difference is less than the desired confidence interval threshold, i.e., \( RRCI = 4 \) brpm.

Study Design

Healthy subjects (n=35) with no known cardiovascular conditions were recruited for the validation study. Participants were asked to determine their skin type using the Fitzpatrick questionnaire [49], and provided their height, weight, age, and gender. Prior to each test subjects rested in a seated position for 15 minutes in order to ensure measurement of the resting heart rate [50]. Demographics of the participants are summarized in Table 1. The study was approved under Institutional Review Board of San Jose State University (SJSU). Written informed consent was obtained from all subjects.

The parameters estimated by the wristband were compared with the simultaneously recorded gold standard measurements from the ECG and spirometry sensors. The BIOPAC system (MP160, BIOPAC, Goleta, USA), which performs 3-lead ECG and spirometry measurements, was calibrated according to manufacturer instructions. Subjects were instructed to breathe through a mouthpiece while wearing a nose clip. A finger-clip Nonin pulse oximeter (8000AA, Nonin, Plymouth, USA) was placed on the subject's right index finger to compare the quality of
the PPG signal recordings of the Wavelet Wristband to a typical clinical grade finger-clip PPG device. For each test two wristbands were placed on each participant as such the first wristband on the left wrist and the second one on the right wrist. Each test lasted between two to four minutes. For several subjects (n = 12) the test was halted before the three minutes mark due to discomfort breathing into the spirometer. All data passing the signal quality analysis were included in the results as described in the next section. Each subject repeated the test twice with five minutes rest in between. Overall, 70 tests were conducted each with two wristbands and reference ECG, spirometer and finger-clip pulse oximeter device recordings. The synchronous recordings from ECG, Nonin, and Wavelet wristband devices were aligned manually based on timestamps and agreement of interbeat intervals.

Signal Quality Analysis and Statistical Methods

Once each test session is recorded by the wristband and transferred to the cloud server, the biometrics algorithms estimate the vital signs over non-overlapping 60-second windows. Each qualified biometric assessment over a 60-second window is referred to as a valid measurement. As PPG signals are inherently sensitive to motion and light artifacts, prior to estimating any vital signs each biometrics algorithm validates the signal quality of each 60-second window based on multiple heuristics including HNR, stFT, motion, and biometrics-specific thresholds such as HRCI, HRVCI and RRCI. If any part of the PPG signal fails one of the predefined signal quality checks, the algorithms report less number of valid measurements for this test session, as such 0 to 4 valid measurements can be generated for a four minutes long session depending on the signal quality. Left and right wrist recordings were analyzed independently, and effectively 140 test sessions recorded by the wristbands were compared with the reference recordings. The total number of valid sessions and the valid measurements qualified by the algorithm for each biometric are provided in the Results section.

To evaluate the signal quality of the Wavelet wristband the HNR of the wristband IR signal was compared with the HNR computed from the IR signal of Nonin finger-clip pulse oximeter. HNR is calculated using the autocorrelation method described by Boersma et al. [51]. This frequency based signal quality assessment method yields an objective measure of the periodicity of the PPG signal from the maximum of the signal’s normalized autocorrelation function [51] The HNR is computed for 6 second overlapping windows centered one second apart. The average over all HNR windows is reported as the HNR of the 60-second recording. Recordings that fail to meet the desired HNR level do not qualify to vitals analysis. It’s important to note that the HNR criterion assumes that all signals other than the signal of interest are noise. Therefore, to accurately estimate HNR of the PPG signal, a preprocessing step is required to remove other physiological contributions to the signal such as respiration. The HNR reported herein follows the cardiovascular component of PPG signal after using a bandpass filter with lower and upper limits set to 0.3 Hz and 10Hz, respectively.

To compare the HNR of the Wavelet Wristband with the finger-clip pulse oximeter, we present the mean and standard deviation of HNR for both devices for each valid measurement as well
as the box plots to compare distributions. The reference HR and HRV were computed from the reference ECG measurements using the Python BioSPPy biosignal processing toolbox [52]. To compare the biometrics estimates by the wristband to those measured in the same time window by the ECG and spirometer, the Pearson correlation coefficients along with the Bland-Altman plots and Bland-Altman limits of agreement are presented [53]. The Bland-Altman limits of agreement took into account multiple observations collected over time from the same set of individuals using two reference devices [54]. The effect of the averaging window size on the accuracy of HR estimates and the variation in biometric measurements collected from the left versus right wrists are also evaluated.
Results

Study participant demographics are summarized in Table 1. The average and standard deviation of height, weight, age was $172 \pm 10$ cm, $74 \pm 18$ kg, $25 \pm 4$ years, respectively. The Fitzpatrick score indicates a good coverage of light, medium and dark skin tones, with slight skew to the darker pigmentation range.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>172 cm</td>
<td>10 cm</td>
</tr>
<tr>
<td>Weight</td>
<td>74 kg</td>
<td>18 kg</td>
</tr>
<tr>
<td>Age</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Gender Female</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>Fitzpatrick Score*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type I:</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Type II:</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Type III:</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Type IV:</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Type V:</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Self-reported Fitzpatrick scores classify subjects’ skin tones based on response of different skin types to ultraviolet light. Type-I: pale white, Type-II: white, Type-III: cream white, Type IV: moderate brown, Type V: dark brown.

PPG Signal Qualification

The PPG signal obtained from the Wavelet Wristband has a physiological morphology similar to those collected with the Nonin device from the finger-clip as shown in Figure 3. Compared with reference finger-clip PPG measurements, the diastolic peak is located closer to the systolic peak and the diastolic decay is steeper, in agreement with earlier studies comparing different measurement sites of PPG [55], and arterial pressure [56]. Figure 4 shows the continuous heart rate estimate of the Wavelet Wristband indicating low variability across right and left wrists, and strong agreement with the reference ECG device.
Figure 3: Signal traces recorded simultaneously from ECG, Nonin finger-clip pulse oximetry device, and two Wavelet Wristbands placed on the left and right wrists. Peaks (ECG and Nonin) or valleys (Wavelet) detected are marked.
Figure 4: Representative continuous beat-to-beat estimation heart rate from the wristbands on the left and right wrists show good agreement with the ECG heart rate measurements. Wavelet #1 and #2 labels refers to the wristband placed on the left and right wrist, respectively.

Average HNR values for each 60-second non-overlapping window are computed to assess the quality of the signal at each LED wavelength collected by the Nonin finger-clip PPG device and the wristband. Table 2 shows that the mean and standard deviation of HNR are 7.70 ± 1.99, 6.40 ± 2.16 and 4.68 ± 2.70 for Wavelet IR, Wavelet red and Nonin IR, respectively. The distribution of HNR for each group is illustrated in Figure 5 using boxplots. Note that not all Wavelet Wristband recordings had a corresponding Nonin recording available due to equipment availability on the test day. The difference histogram in Figure 5 illustrates that the pairwise mean difference (with one standard deviation) between the Wristband IR HNR and Nonin HNR was -1.34 ± 2.75 dB. Thus, the Wristband IR HNR is slightly lower than that of the finger-clip sensor.

In this study, total 366 minutes of PPG signal recordings were collected from 35 subjects. Approximately 7% of the PPG recordings were flagged invalid due to subtle arm motions during the test. 74 measurements from 23 test sessions fail to meet the desired HNR level do not qualify to vitals analysis. The signal processing algorithm generated valid HR, HRV and RR measurements for 78%, 77%, 41% of the signals, respectively. Importantly, the signal quality
The pre-processing step disqualified all readings from two Type II and one Type I subject. Also, it was later found that for one subject the ECG probe was dislocated, therefore, no valid ECG reference data was collected. The number of valid vital sign measurements for which the corresponding reference data exist, are shown in **Table 3**.

**Table 2:** The Mean and Standard Deviation of the Average Harmonic to Noise Ratio for the PPG recordings collected from the Wavelet Wristband and Reference Nonin Finger-Clip Pulse Oximeter

<table>
<thead>
<tr>
<th></th>
<th>Nonin IR (dB)</th>
<th>Wavelet IR (dB)</th>
<th>Wavelet Red (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>216 measurements</td>
<td>266 measurements</td>
<td>266 measurements</td>
</tr>
<tr>
<td></td>
<td>28 subjects</td>
<td>33 subjects</td>
<td>33 subjects</td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>7.70 ± 1.99</td>
<td>6.40 ± 2.16</td>
<td>4.68 ± 2.70</td>
</tr>
</tbody>
</table>

**Figure 5:** Boxplot comparison (left) of average harmonic to noise ratio (HNR) estimated over 60-second non-overlapping windows for each PPG wavelength: Nonin finger-clip pulse oximeter (Nonin IR), the wristband infrared (Wavelet IR), and the wristband red (Wavelet Red). The histogram for pairwise-difference of average HNR between Wavelet IR and Nonin IR signals (right).
Heart Rate Validation

Table 3 shows the mean absolute error and mean absolute percentage error of the HR, HRV and RR estimates of the Wavelet Wristband compared with the reference devices. Across 254 measurements of 60-second non-overlapping windows the mean pairwise absolute error of heart rate was 0.7 ± 0.9 BPM (0.9 ± 1.3 %) against the reference ECG. Figure 6 shows the distribution of ECG and Wavelet heart rate estimates with scatter and Bland-Altman plots. The Pearson correlation coefficient (R) for HR between wristband estimates and reference ECG measurements is 0.994, and the mean difference (bias) between Wavelet and ECG HR (with 95% confidence interval) was -0.32 ± 0.13 bpm. All measurements stay within 5% absolute percent error, except one outlier for which the Wavelet Wristband underestimated the heart rate by 7% (6.8 BPM). The Bland Altman ratio, i.e., the ratio of 1.96 * standard deviation divided by the mean of the pairwise measurement means, is equal to 0.03, which indicates good agreement between measurements [57].

Table 3: The Accuracy of the Biometric Estimates (mean ± standard deviation) of the Wavelet Wristband Compared to the Reference ECG and Respirometer Measurements

<table>
<thead>
<tr>
<th>Valid Sample Size</th>
<th>HR</th>
<th>HRV SDNN</th>
<th>HRV RMSSD</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>254 measurements 31 subjects</td>
<td>247 measurements 31 subjects</td>
<td>249 measurements 31 subjects</td>
<td>128 measurements 26 subjects</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.7 ± 0.9 bpm</td>
<td>7 ± 10 ms</td>
<td>11 ± 12 ms</td>
<td>1 ± 1 brpm</td>
</tr>
<tr>
<td>Mean absolute percentage error (%)</td>
<td>0.9 ± 1.3</td>
<td>11 ± 13</td>
<td>28 ± 30</td>
<td>2.5 ± 2.5</td>
</tr>
<tr>
<td>Mean error (bias)</td>
<td>-0.3 ± 1.1 bpm</td>
<td>-1 ± 12 ms</td>
<td>3 ± 16 ms</td>
<td>1 ± 2 brpm</td>
</tr>
<tr>
<td>Pearson correlation (R)</td>
<td>0.994</td>
<td>0.907</td>
<td>0.924</td>
<td>0.863</td>
</tr>
</tbody>
</table>

HR: heart rate, HRV: heart rate variability, RR: respiration rate, SDNN: standard deviation of normal-to-normal intervals, RMSSD: root mean square of successive differences. Valid sample size refers the number of measurements and subjects where both valid wristband and reference data was available. ± values indicate one standard deviation.
Figure 6: The distribution of measured heart rate (HR) by the ECG and the wristband (left). The Pearson correlation coefficient is shown on the right corner of this scatter plot. Bland-Altman plots of the absolute error of HR between the wristband and the simultaneously recorded ECG measurements versus the mean of the measurements in beats per minute (right). The solid black line indicates the mean value. The dotted lines marks the 95% limit of agreement at -2.6 bpm and 1.9 bpm.

Heart Rate Variability

The SDNN and RMSSD HRV are computed over 60-second non-overlapping windows for ECG and Wavelet wristband recordings. The mean absolute errors for SDNN and RMSSD were estimated as 7 ± 10 ms and 11 ± 12 ms, respectively, as shown in Table 3. The mean difference (with one standard deviation) between Wavelet and ECG-based SDNN and RMSSD were -1 ± 12 ms and 3 ± 16 ms respectively. The relationship between the Wavelet SDNN HRV estimates to the ECG is visualized with scatter plots as well as with Bland-Altman plots (Figures 7 and 8). Pearson correlation coefficients for HRV SDNN and RMSSD are estimated as 0.907 and 0.924, respectively. The Bland Altman ratios for HRV SDNN and RMSSD are 0.35 and 0.42, which indicate strong correlation between the wristband HRV estimates and the reference measurements. Relatively lower correlation for RMSSD estimates is attributed to the outliers at the high RMSSD range (> 150 ms).
Figure 7: The distribution of measured HRV SDNN by the ECG and the wristband (left). Pearson correlation coefficient is shown on the right corner of this scatter plot. Bland-Altman plots of the absolute error of HRV between the wristband and the simultaneously recorded ECG measurements versus the mean of the measurements in milliseconds (right). The solid black line indicates the mean value. The dotted lines marks the 95% limit of agreement at -25 ms and at 22 ms.

Figure 8: The distribution of measured HRV RMSSD by the ECG and the Wristband (left). Pearson correlation coefficient is shown on the right corner of this scatter plot. Bland-Altman plots of the absolute error of HRV between the wristband and the simultaneously recorded ECG measurements versus the mean of the measurements in milliseconds (right). The solid black
Respiration Rate

Figure 9 shows the comparison of the RR estimates of the wristband to the reference spirometry measurements. The Pearson correlation coefficient is 0.863 and the mean difference (bias with one standard deviation) between Wavelet and spirometer RR is 1 ± 2 brpm. The pairwise mean absolute error is 1 ± 1 brpm (2.5 ± 2.5 %) as shown in Table 3.

Figure 9: The distribution of measured respiration rate (RR) by spirometer and the wristband (left). Each data point represent the average RR over the 60-second non-overlapping measurement window. Bland-Altman plots of the absolute error of RR between the wristband and the simultaneously recorded control measurements versus the mean of the measurements in beats respiration per minute (right). The solid black line indicates the mean value. The dotted lines marks the 95% limit of agreement at -3 brpm and at 4 brpm.

To assess the agreement between the HR and HRV computed by a Wavelet Wristband on the left and right wrist, the mean absolute error is computed for subjects where both valid simultaneous left and right wrist readings are available (n=23). Table 4 shows that the mean absolute difference between right and left wrist measurements was 0.6 ± 0.8 bpm, 6 ± 10 ms, 9 ± 10 ms, and 1 ± 2 brpm for HR, HRV SDNN, HRV RMSSD, and RR respectively.
Table 4: The Pairwise Mean Absolute Difference Between Left and Right Wristbands

<table>
<thead>
<tr>
<th>Valid Sample Size (pairs)</th>
<th>HR</th>
<th>HRV SDNN</th>
<th>HRV RMSSD</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>108 measurements 23 subjects</td>
<td>102 measurements 23 subjects</td>
<td>101 measurements 23 subjects</td>
<td>48 measurements 23 subjects</td>
</tr>
<tr>
<td>Mean absolute difference</td>
<td>0.6 ± 0.8 bpm</td>
<td>6 ± 10 ms</td>
<td>9 ± 10 ms</td>
<td>1 ± 2 brpm</td>
</tr>
</tbody>
</table>

HR: heart rate, HRV: heart rate variability, RR: respiration rate, SDNN: standard deviation of normal-to-normal intervals, RMSSD: root mean square of successive differences

Influence of recording window duration on biometric estimation accuracy

For certain use-cases estimation of biometrics over durations shorter than 60 seconds may be desirable. To assess the accuracy of the biometric estimations over shorter recording durations the absolute error and absolute percentage errors from the reference were computed by re-analyzing the test and reference signals over 45 and 30 seconds long windows. Results shown in Table 5 indicate that the mean absolute error in HR remain stable, i.e., within 1 bpm, as the recording duration shortens from 60 to 30 seconds. Similarly, mean absolute error in HRV RMSSD estimates remained at 12 ms without displaying dependence to recording duration. However, the mean absolute error of SDNN was influenced by the recording duration. For 30 second recording duration mean absolute error for RR estimations increased to 2±2 brpm.

Table 5: Mean Absolute Error and Standard Deviation of Biometrics Estimations from Reference Measurements at Different Recording Window Durations

<table>
<thead>
<tr>
<th>Biometrics</th>
<th>Recording Window Duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60</td>
</tr>
<tr>
<td>HR (bpm)</td>
<td>0.7 ± 0.9, n = 254</td>
</tr>
<tr>
<td>HRV SDNN (ms)</td>
<td>7 ± 10, n = 247</td>
</tr>
<tr>
<td>HRV RMSSD (ms)</td>
<td>11 ± 12, n = 249</td>
</tr>
<tr>
<td>RR (brpm)</td>
<td>1 ± 1, n = 128</td>
</tr>
</tbody>
</table>

Discussion

Clinical utility of wearable health devices it is imperative to validate the accuracy of the metrics derived by these devices against gold standard measures, and importantly to characterize its limitations. In this study, we evaluated the performance of the Wavelet Wristband by comparing the accuracy of the estimated biometrics to the ground truth references. The mean absolute error of HR, HRV and RR was 0.6 ± 0.9 bpm, 7 ± 10 ms (SDNN) and 1±1 brpm, respectively.
Bland Altman graphs demonstrate good agreement between the wristband biometric estimates with the reference measurements. These results indicate that Wavelet Wristband can accurately estimate multiple vital signs, and provide a continuous non-invasive health monitoring solution alternative to other devices typically incorporating ECG.

To our knowledge this study is the first of its kind exploring the PPG signal quality collected by a wrist-worn reflective PPG device as opposed to more traditional devices placed on measurement sites such as the finger, where the subcutaneous tissue is perfused more strongly with dense microvasculature. Previously, several investigators reported accurate estimation of vitals using PPG signal collected from the finger in comparison with ECG in healthy [58-60] and disease settings [61-63]. Investigations by Maeda et al. [55] showed relatively high signal strength can be obtained over from the wrist but lacked signal to noise ratio assessments which is essential to estimating accurate biometrics. In this study we suggest HNR can be used to characterize PPG signal quality collected from different measurement locations, benchmark PPG devices, and aid establishing signal quality standards in PPG research. Comparable HNR of the Wavelet Wristband and Nonin finger-clip pulse oximeter devices indicates that good quality PPG signal can be collected from the wrist.

While both PPG and ECG signals convey physiological information, the underlying physiology of PPG stems primarily from hemodynamics rather than electrical activity of the heart depicted in the ECG signal. The well-defined morphology of the ECG signal allow relatively simple extraction of beat to beat intervals in the absence of artifacts related to drift, electromagnetic and biologic interferences [58]. In contrast, PPG signal hosts inherently more rounded peaks and valleys, therefore requires more sophisticated algorithms to extract physiological measures. Similar to blood pressure, PPG signal morphology depends strongly on the timing of reflected waves (i.e. pulse transit time) from the downstream vasculature [64], which is negatively correlated with vascular stiffness, and age [15]. It’s reasonable to assume changes in pulse transmit time adds another layer of challenge to accurate extraction of salient features from the PPG signal, and contributed, in part, to deviations reported herein from the reference device measurements.

PPG signal quality is affected by multiple factors including improper sensor-skin coupling due to device malposition, ambient light, pressure on skin, biological factors (blood perfusion, tissue composition, skin temperature), and is highly sensitive to motion [12, 65]. To eliminate inaccurate readings from PPG devices it is necessary to incorporate proper signal qualification checks and biometric-specific heuristics to the signal processing algorithm. Only then will the PPG device be able to estimate biometrics within desired accuracy range while providing sufficient number of biometrics readings for the designed use case. In this study between 22 to 59% of measurements were disqualified by the vitals algorithms. Important to note, cloud computing and storage of raw signals enable retrospective-processing of the physiological signals collected by digital health devices. Leveraging more advanced algorithms, this framework will allow further improvement of the accuracy and the number of the valid readings generated from the same signals.
In addition to post-processing strategies, the choice of the PPG wavelength also impacts the quality of the PPG signal and accuracy of the computed metrics. This choice is a trade-off and depends on the targeted application, but is usually in the 510 to 940 nm range corresponding to green and infrared lights, respectively. Measurements done on light skins and at normal ambient temperature (around 20 °C) have shown that reflected green light has more desired properties to maintain a good signal to noise ratio during motion compared to IR [66]. This is the main reason many consumer devices that target ambulatory heart rate measurements use green light source. The advantage of IR light over green is that it is less sensitive to skin tone variations and perfusion level due to its better tissue penetration characteristics [14]. The darker the skin pigmentation, i.e., high melanin concentration, the harder it is to receive good signal with light wavelengths shorter than 650nm. Also, for individuals with relatively lower superficial skin perfusion, thicker skin, larger wrist circumference or body mass indices, particularly in cold ambient conditions blood microcirculation is significantly lower, and it becomes advantageous to reach deeper tissues. Therefore, we selected IR light as the primary light source for the Wavelet Wristband to serve particularly for resting state biometrics monitoring across the population due to its relatively lower sensitivity to skin tone and perfusion levels. Selection of the optimal wavelength for vitals monitoring depends on target application, and is a compromise between competing factors: vulnerability to artifacts versus sensitivity to skin pigmentation and poor skin perfusion.

As a limitation of this study, the biometrics reported herein were tested at the resting condition and lacked physical activity settings. It is well known that accuracy of biometrics derived from PPG-based devices is affected by motion artifacts [67, 68]. Recent studies evaluating HR estimates from several commercially available wrist-worn PPG devices reveal variable degrees of accuracy during physical activity [69, 70]. Accuracy of the biometrics estimation during motion requires robust motion reduction algorithms [71-73]. Future research in this area needed to preserve the original signal morphology and extract more information than just ambulatory HR. Another limitation is that the subjects recruited for the study have no known health conditions and come with restricted age range and skin tones available. The correlation between measurement error and Fitzpatrick score indicate higher deviations at the extremes. Further studies are needed to demonstrate utility of the devices for larger more diverse populations. Moving forward, well-designed clinical studies are required to demonstrate the impact of new wearable and connected devices on clinical outcomes, and to build real world evidence for indications that benefit most from these new technologies.

In addition to the above technical challenges, issues involving privacy, safety, accessibility, and interoperability of the health data should also be addressed. More personal health data has been collected over last three years than last three decades combined by health organizations that are looking for ways to leverage big data to modernize medicine and reduce cost of healthcare delivery. It's striking that the current healthcare systems are unable to ingest this large volume of personal health data collected from wearable and connected devices, yet alone the privatization of health data and lack of an interoperable data infrastructure are the main culprits [74]. Also, many health device vendors and platform solution providers collect and monetize personal health data in a privatized setting lacking proper consent and public
adoption. Unequivocally, the corporate monopoly over person generated data is unethical, and bound to change with the new wave of regulations and advances in information technology. Data generators’ ownership and support for the data generation process is the key to achieve the large-scale adoption of new healthcare delivery of the digital age. Therefore, there is an important unmet need to streamline health data collection, sharing and storage. An alliance between a “motivated” network of data generators, “well-supported” health device and service providers has yet to be defined to develop an interoperable scalable health data infrastructure. Ongoing clinical studies utilizing wearable and connected devices in multiple therapeutic areas including atrial fibrillation, sleep apnea, heart failure, vascular and blood diseases, structural heart diseases, trauma and blood volume management should harness interoperable big data analysis platforms to demonstrate how data gathered with these devices translates into actionable insights that led to clinical decisions. The clinical research community, biotech industry and government agency to work collectively to demonstrate improved outcomes along with lower cost of healthcare delivery.

In conclusion, this study demonstrates accurate estimation of physiological measures including HR, HRV, RR from a wrist-worn PPG device compared with the reference ECG and spirometer devices. The quality of the PPG signal generated by the Wavelet Wristband is comparable to those obtained with commercially available finger-clip pulse oximeters. Next generation wearable and connected devices provide unprecedented means for continuous long-term remote health monitoring. Due to their noninvasive, convenient-to-patient and engaging nature these technologies have been gradually becoming part of our everyday lives, and act as companion tools for clinical decision support supplementing established gold standard methods. This new streamlined health data collection modality will enable new and better ways to measure personal health, generate insights which otherwise are not available, and ultimately improve healthcare delivery.

Acknowledgements

Authors thank professors Alessandro Bellofiore and Dan Nathan-Roberts for their support for recruiting the study subjects and conducting the experiments at San Jose State University Biomedical Engineering Laboratories.

References

Association", 6(9), p.e007245.
[14] Spigulis, J., Gailite, L. and Lihachev, A., 2007, “Multi-wavelength photoplethysmography for simultaneous recording of skin blood pulsations at different vascular depths". In Biomedical Optics (BiOS) 2007 (pp. 64301I-64301I). International Society for Optics and Photonics


