# Estimation of Heart Rate Variability Measures Using Apple Watch and Evaluating Their Accuracy

Estimation of Heart Rate Variability Measures Using Apple Watch

Ahmad, Turki
Department of Bioengineering, The University of Texas at Arlington
ahmad.turki@mavs.uta.edu

Kan, Ding
Neurology and Neurotherapeutics, UTSW Medical Center
kan.ding@utsouthwestern.edu

Rong, Zhang
Neurology and Neurotherapeutics, UTSW Medical Center
RongZhang@texashealth.org

Ming, Li
Department of Computer Science and Engineering, The University of Texas at Arlington
Ming.li@uta.edu

Kathleen, Bell
Physical Medicine and Rehabilitation, UTSW Medical Center
kathleen.bell@utsouthwestern.edu

Khosrow, Behbehani
Department of Bioengineering, The University of Texas at Arlington
kb@uta.edu

## ABSTRACT

In a pandemic crisis such as the one that the world has been experiencing since January 2020, utilizing remote mobile health monitoring can help monitor the cardiac health of healthy individuals as well as chronically ill patients. Apple Watch is a fitness tracker that measures heart rate and can facilitate this need due to its availability worldwide and relative affordability. Hence, it has the potential of being useful in assessing and monitoring one's cardiac health. This paper reports the preliminary results from an ongoing study aimed at assessing the accuracy of the Apple Watch in measuring heart rate variability. To achieve the goal of the study, concurrent measurements of electrocardiogram (ECG), as a gold standard, and instantaneous Apple Watch heartrate values were made in six (4 male) age 26.4±3.2 years old volunteer healthy subjects. Both time and frequency domain measures of heartrate variability (HRV) were obtained from the recorded ECG and estimated from the Apple Watch instantaneous heartrate values. The results indicate that it may be possible to obtain a reasonable estimate of HRV time-domain and frequency-domain metrics using Apple Watch in healthy subjects.

## CCS CONCEPTS

• **Human-centered computing**; • **Ubiquitous and mobile computing**; • **Ubiquitous and mobile devices**; • **Smartphones**;

## KEYWORDS

Wearable Health Monitor, Wearable Fitness Tracking, Exercise Therapy, Fitness Tracking Ecosystem

## 1 INTRODUCTION

One of the clinically important biomarkers of cardiac health is heart rate variability (HRV). HRV is shown to be an indicator of the neurocardiac function and is generated by heart-brain interactions in conjunction with the nonlinear dynamics of the autonomic nervous system (ANS) [1]. Since HRV can be measured non-invasively, it is a logical choice for monitoring one's cardiac health remotely. Previous research has investigated the usage of HRV monitoring in calculating or detecting sleep quality, mental stress, chronic pain, posttraumatic stress disorder, bipolar disorder, cardiac health [2], and traumatic brain injury [3]. Furthermore, other research has investigated the accuracy of the Apple Watch HRV measurements against another device (Polar chest strap; Polar Electro, Kempele, Finland) [4].

Heart rate variability is a quantitative measure of the fluctuations in the time interval between consecutive heartbeats [5]. HRV reflects neurocardiac function which is generated by heart-brain interactions and dynamic non-linear autonomic nervous system (ANS) process [5]. HRV reflects the regulation of autonomic balance, blood pressure (BP), gas exchange, gut, heart, and vascular tone [5]. HRV can be considered as a measure of ANS imbalances [5]. Specifically, the variation between the temporal position of the consecutive heartbeats is low when a person finds herself/himself in a fight-or-flight mode, and when the variation between the temporal position of consecutive beats is high a person's body is in a more relaxed state [6]. In other words, the healthier the ANS the faster the patient can switch gears and show more resilience and flexibility when responding to stimuli [6]. A low HRV is even associated with an increased risk of death and cardiovascular disease [7].

In addition to the relation of cardiac health to HRV, cognitive and emotional control have also been shown to be related to cardiac vagal tone [8] [9] [10]. In the HRV spectrum, the high-frequency band corresponds to parasympathetic cardiac activity [2]. The successful adaptation of the individual to changing environmental demands is influenced by the parasympathetic nervous system [5] [11] [8] [12]. A lack of ability to respond flexibly to varying demands could be indicated by a reduction in vagal control (i.e., decreased high frequency-HRV), decreasing the range of possible cognitive and emotional options and thus limiting the individuals' ability to generate appropriate responses and inhibit inappropriate ones [1].

The quantitative measures of HRV have been developed and they can be categorized as time-domain, frequency-domain, and non-linear measurements [5]. The time series of the interval between R peaks is called the RR interval. The time series formed by RR intervals is used for both time-domain and frequency-domain measures [5]. Non-linear measurement is usually used to quantify the unpredictability of a time series [5].

The amount of variability in a measurement of the inter-beat intervals (IBI) which is the period between successive heartbeats [5] can be used to estimate the time-domain measures of HRV. Alternatively, HRV frequency-domain measures approximate the distribution of absolute or relative power of the IBI in four frequency bands.

The ULF band range is ≤0.003 Hz indexes fluctuations in IBIs with a period from 5 min to 24 h and is measured using 24 h recordings [5]. The VLF band range is 0.0033–0.04 Hz is covered with patterns for intervals between 25 and 300 s [5]. The LF band range is 0.04–0.15 Hz is contained of patterns with periods between 7 and 25 s and is influenced by breathing from ~3 to 9 bpm [5]. In a 5 min sample, there are 12–45 complete periods of oscillation [5]. The HF or respiratory band range is 0.15–0.40 Hz and is affected by breathing from 9 to 24 bpm [5]. The ratio of LF to HF power (LF/HF ratio) may approximate the ratio between the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity under regulated conditions [5].

## 2 METHOD

Apple Watch Series 3 (Apple, Cupertino, CA USA) was used for this study's evaluation. Heart rate measuring while sitting was chosen to test the accuracy of the Apple Watch in measuring heart rate variability.

Testing the Apple Watch during sitting can elucidate the heart rate variability recordings in resting mode. The heart rate variability was compared with concurrently recorded electrocardiography (ECG) using Biopac MP 150 ECG system (BIOPAC Systems, Goleta, CA USA) with a 5 ECG electrode configuration. The ECG data were sampled using the National Instrument NI-USB-6128 analog-to-digital converter (National Instruments, Austin, TX USA). The sample rate was 256 samples/s and data were recorded using a Dell Latitude E640 laptop computer (Dell, Round Rock, TX USA) running a custom-made LabView program by National Instruments for data acquisition.

The data were stored and analyzed offline. The data reported here is obtained from 6 adult healthy subjects who were non-smoker (2F and 4M; age 26.4±3.20 years; BMI 24.7±1.92 kg/m2). The protocol

for the study was approved by our institutional review board (IRB) and the subjects signed an informed consent form which was also approved by our IRB. The subjects performed the test while they placed the Apple Watch on the right wrist. Each data recording session lasted 5 minutes. The data collected by the Apple watch were analyzed using the Fitiv Pulse application on an iPhone (MOTIFIT, Moncton, Canada). The Fitiv Pulse makes downloading the recorded heart rates from the Apple Watch into a separate computer for further analysis possible.

The recorded ECG QRS complex waveform was analyzed using the MATLAB program (version R2018; MathWorks, Natick, MA USA). For this purpose, for each recorded QRS the R wave peak was automatically detected using a peak detection algorithm available in MATLAB which is called "findpeaks". This allowed us to compute the temporal duration of IBI's for recorded heartbeats within the 5 minutes while the subject remained sitting. Since the heartbeats do not occur at regular intervals, it was necessary to interpolate between the recorded instances of heartbeats to achieve and uniformly resample the resulting interpolated curve. Similarly, the recorded heart rates from the Apple Watch were interpolated and resampled. An example of this method of obtaining uniformly sampled RR from ECG and Apple Watch is shown below. The interpolation of the ECG heartbeat data and the Apple-Watch heartbeat data were achieved using a cubic spline interpolation method which is available in MATLAB software. Both resulting interpolated heartbeat waveforms were uniformly sampled at 256 samples/s.

The time-domain and frequency-domain measures of HRV were computed using the Apple-Watch-recorded heart rate and the concurrent corresponding ECG recorded waveforms. Specifically, for the time-domain measures, we computed the mean, standard deviation, and coefficient variation.

For the frequency domain measures, we obtained the power spectral density (PSD) of the interpolated heart rates derived from the recorded ECG and the PSD of the interpolated heart rates obtained by the Apple Watch. Signal energy is found within a frequency band defined as signal power [1].

Heart rate (HR) oscillations were split up into three bands, very-low-frequency (VLF), low-frequency (LF), high-frequency (HF) bands [12]. Those bands were presented by the Task Force of the European Society of Cardiology and also by the North American Society of Pacing and Electrophysiology (1996). [7]. Using these PSD's, we computed power values for very low (0.0-0.04 Hz), low (0.04-0.15 Hz), and high frequency (0.15-0.4 Hz) range. HRV measurements in frequency-domain can be stated in absolute or relative power [5]. Absolute power is computed as second squared divided by cycles per second ($s^2$/Hz) [1]. In our study, we used absolute power computations.

## 3 DATA ANALYSIS

The temporal distance between the location of the peak of the R wave of the concurrently recorded ECG QRS complex was used to compute the RR intervals for the 5 minutes of recording in sitting posture then, the interval between detected consecutive R peaks were interpolated and uniformly resampled at 256 samples/s Figure 1 and Figure 2. To calculate the RR intervals from the recorded Apple Watch heart rates (Apple HR), we estimated the RR interval
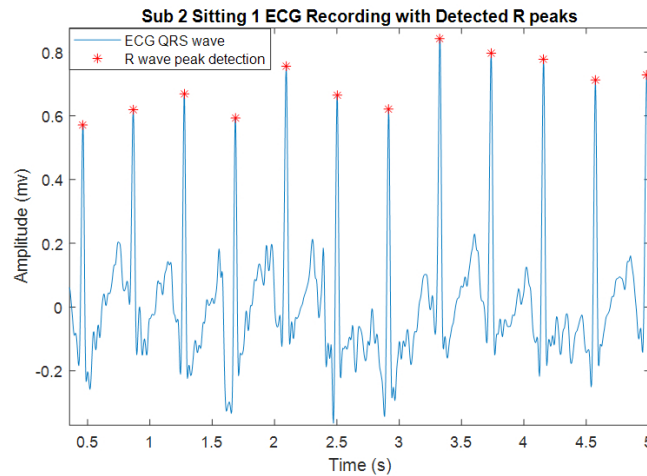
**Figure 1: Subject No. 2 RR Peaks detection (star: R wave detected peeks, line: ECG QRS wave)**

associate with each recorded heart rate (Apple RR) by the following formula:

$$Apple\ RR = \frac{60}{Apple\ HR} \tag{1}$$

where the Apple HR is measured in beats per minute and Apple RR is expressed in seconds. As in the case of ECG RR, the resulting Apple RR values for the 5 minutes of recording were interpolated and uniformly sampled at 256 samples/s.

To gauge the difference between the Apple RR and ECG RR, we compute the root means square error (RMSE) for all the interpolated value of the Apple RR and ECG RR as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} E_i^2}{n}} \tag{2}$$

where, $E_i$ = (ECG RR)$_i$ - (Apple RR)$_i$, and i signifies the interpolated sample point, and n is the number of interpolated sample points in the recorded data.

Using the interpolated values of ECG RR and Apple RR, the time-domain and the frequency-domain measures of HRV were computed. Using the time-domain and the frequency-domain measures of HRV obtained from ECG as gold standards, we performed paired T-Test of the equivalency of the mean values of these measures with their counterparts which were obtained from the Apple Watch heart rate data. Further, we applied the Bald-Altman (B&A) analysis of these HRV metrics to ascertain the agreement of the HRV derived from the recorded Apple Watch heart rates and their corresponding values obtained from the concurrently recorded ECG data.
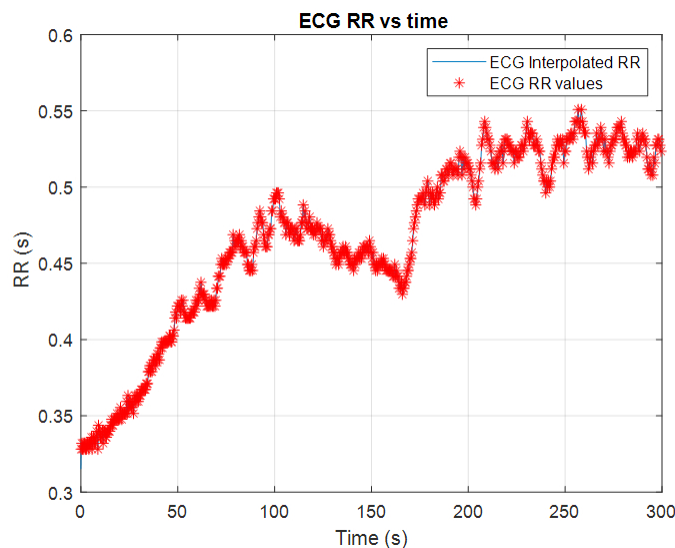


**Figure 2: Subject No 2 ECG RR interpolation (star: RR value obtained from ECG, line: interpolation between RR values)**
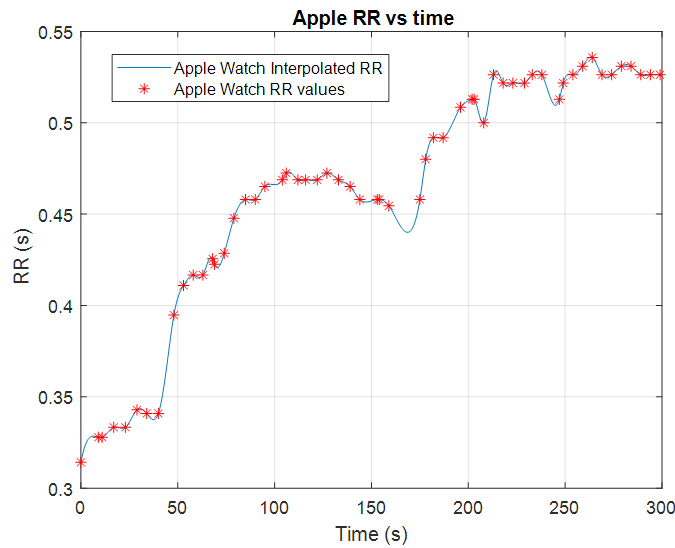
**Figure 3: Subject No.2 Apple Watch RR interpolation (star: RR value obtained from ECG, line: interpolation between RR values)**

For Bland and Altman plots, the average of each pair values of Apple HRV and ECG HRV measures is shown on the x-axis, and the difference between the pair values on y-axis, then bias (B) and 95% limits of agreement (LOA's) were calculated as follows:

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i) \quad (3)$$

$$LOA = Bias \pm 1.96SD \quad (4)$$

where, $x_i$ is the $i_{th}$ of the samples of the metric measured by the Apple Watch (i.e., Apple HRV), $y_i$ is the $i_{th}$ sample of the heart rate variability metric measured by ECG (i.e. ECG HRV), n is the number of samples compared, and SD denotes the standard deviation of the difference between the measure and reference (i.e., $y_i$ - $x_i$).

## 4 RESULTS

For data analysis, we started by detecting the peaks of R waves in the ECG QRS complex. Figure 1 shows plots of ECG for subject 2 in the sitting recordings and peak detection for the first five seconds of the first minute. After that ECG heart rate values interpolated over exercise duration. Figure 2 shows ECG RR values and their interpolation values.

Next, Apple Watch's RR values interpolated and compared with ECG's interpolated wave. Figure 3 shows interpolated Apple RR values compared to the interpolated ECG RR values.

RMSE between ECG and Apple RR measurements were computed for every subject, as it is illustrated in Figure 4. The mean± standard deviation for RMSE from all subjects' measurements was 0.0707±0.0420 seconds.
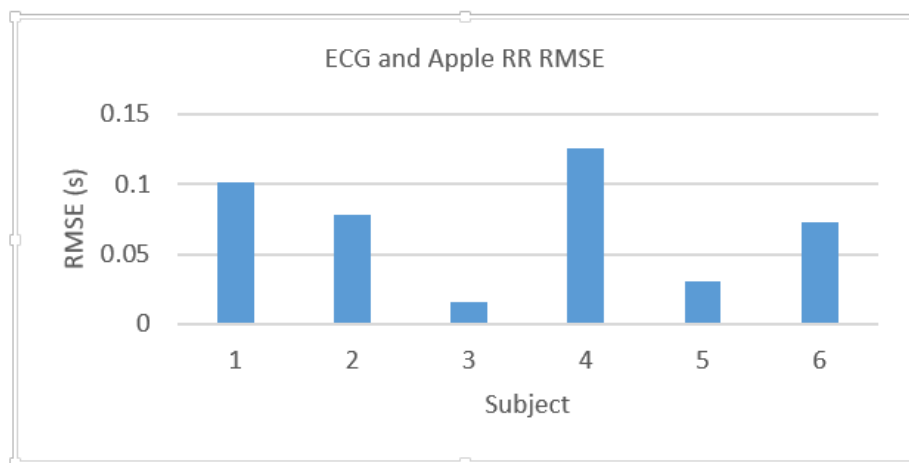


**Figure 4: ECG RR and Apple RR Root Mean Square Error**

**Table 1: Paired T-test (p=0.05) results for the HRV metric measurements between ECG and Apple Watch**

| HRV Metric | ECG (Mean ± SD) | Apple Watch (Mean ± SD) |
|---|---|---|
| SD of RR (s) | $9.23 \times 10^{-2} \pm 3.42 \times 10^{-2}$ | $9.64 \times 10^{-2} \pm 3.74 \times 10^{-2}$ |
| Mean of RR (s) | $67.51 \times 10^{-2} \pm 17.62 \times 10^{-2}$ | $66.11 \times 10^{-2} \pm 17.13 \times 10^{-2}$ |
| SDRR Coefficient of Variation (s) | $13.00 \times 10^{-2} \pm 6.00 \times 10^{-2}$ | $14.00 \times 10^{-2} \pm 3.00 \times 10^{-2}$ |
| Very-Low Frequency $(s)^2$ | $7.88 \times 10^{-3} \pm 7.18 \times 10^{-3}$ | $1.01 \times 10^{-2} \pm 7.49 \times 10^{-3}$ |
| Low Frequency $(s^2)$ | $1.30 \times 10^{-3} \pm 1.09 \times 10^{-3}$ | $2.80 \times 10^{-4} \pm 2.00 \times 10^{-4}$ |
| High Frequency $(s^2)$ | $4.40 \times 10^{-4} \pm 5.05 \times 10^{-4}$ | $4.37 \times 10^{-5} \pm 3.96 \times 10^{-5}$ |
| Low Freq./High Freq. | $4.33 \pm 1.51$ | $8.40 \pm 5.99$ |
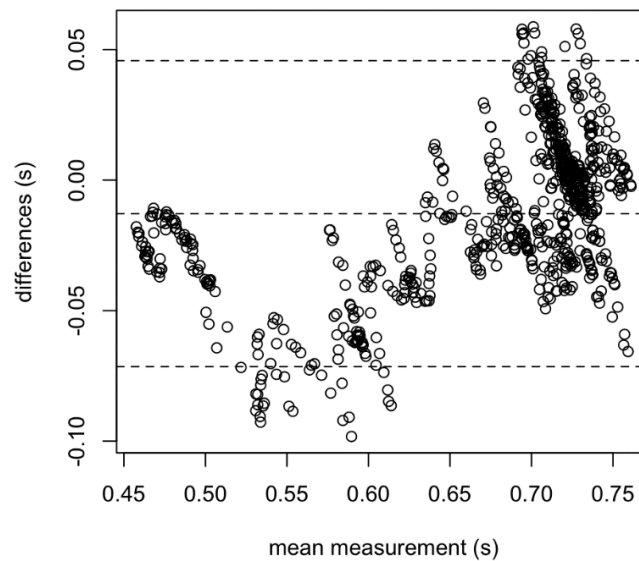


**Figure 5: ECG-RR and Apple-RR B&A plot. It is noted that the values shown are for every 100th sample point from the interpolated data to improve the visibility of the plot. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the**

To quantify Apple Watch accuracy and agreement B&A plots were studied, and statistic measures were calculated. Figure 5 shows ECG vs. Apple's B&A plots for the RR measurements. The power spectrum density for ECG and Apple watch interpolated heartrate signal are shown in Figures 6 and 7. Figure 8 shows the B&A plot mean ECG and Apple Watch RR measurements. Figures 9 and 13 shows ECG and Apple Watch HRV measurements B&A plot in time and frequency domain, respectively. Also, very low frequency, low

frequency, and high-frequency B&A plots for ECG vs. Apple are shown in 10 10, 11 and 12.

## 5 DISCUSSION

In this paper, it is proposed to extract the RR intervals for the Apple Watch be obtained by converting the instances of the measured heart rates to estimate the associated RR intervals using Eq. 1. Since the Apple Watch does not measure the heart rate at regular intervals, the cubic spline interpolation of the estimated RR intervals provides
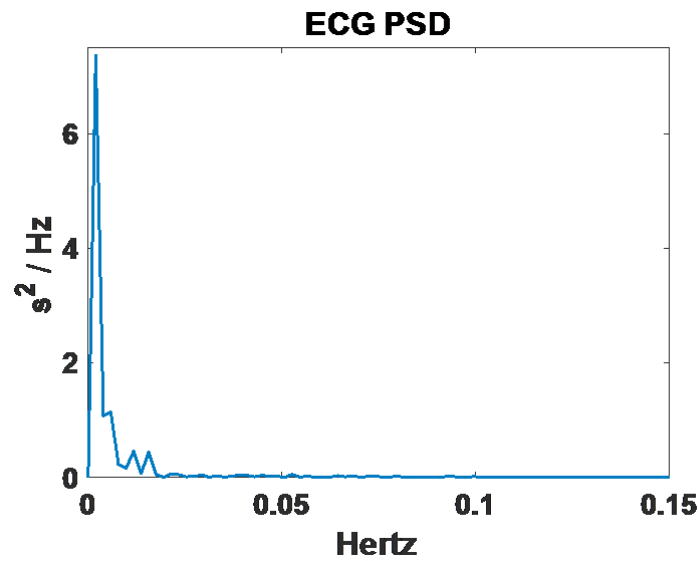
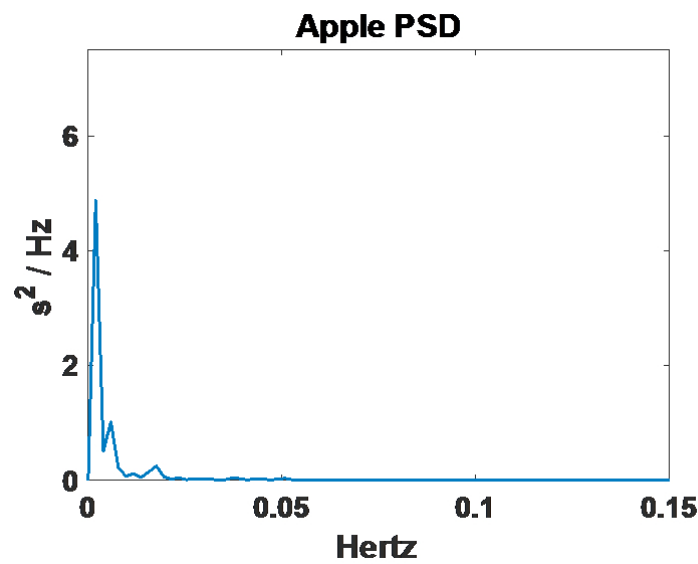Figure 6: Subject No.2 ECG power spectrum density plot



Figure 7: Subject No.2 Apple Watch power spectrum density plot

a promising method of estimating the RR intervals at a fixed sample rate and allows further analysis of the RR interval. The recordings that were obtained in this study contained a transition of the heart rate at the beginning of the recorded session, Figures 2 and 3, as before the subjects starting the sitting session for the collection of data they were walking. Hence, the values reported and analyzed reflect both the transition and the steady-state values recorded.

The RMSE calculations (Figure 4) showed that this measure of deviation between all the interpolated sample points of ECG-derived and Apple Watch-estimated RR values (i.e., approximately 76,800 values from each device for each subject) has less than 0.13

s error. The least error was with subject 3, while the highest RMSE reached up to 0.1262 s different than the gold standard of ECG-derived RR with subject 4. In the course of this study, it was observed that the watch must be worn fairly tightly around the wrist to increase readings accuracy. A loose-fitting watch could not record the heart rates accurately.

The results of the B&A analysis showed that the RR values obtained from the Apple Watch have a reasonable agreement with the ECG-derived RR. In particular, comparing the RR values which are used for extraction of the HRV metrics in Figure 5 reveals that most of the measurements lie within the limit of agreements of the
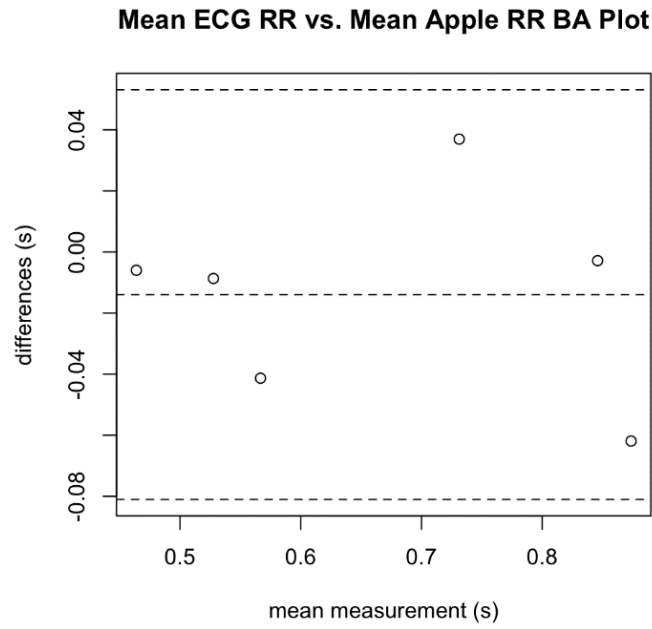
## Mean ECG RR vs. Mean Apple RR BA Plot



**Figure 8: Mean ECG RR and Mean Apple RR B&A plot. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the LOA: Bias ±1.96SD (Eq. 4)**
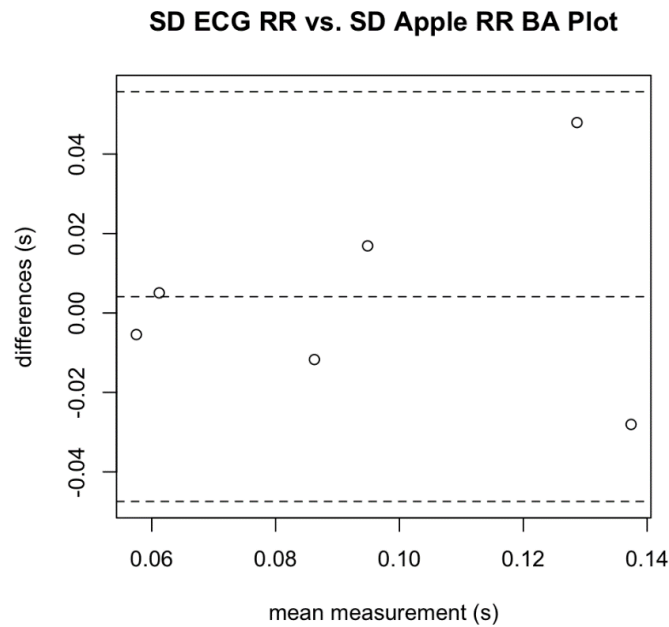
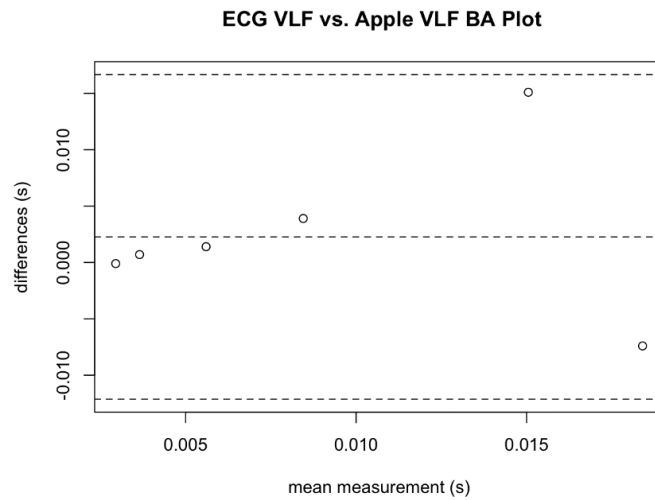## SD ECG RR vs. SD Apple RR BA Plot



**Figure 9: HRV B&A plot for Standard deviation of the RR intervals ECG and Apple. The middle line shows the Bias (Eq. 3) and the bottom lines represent the LOA: Bias ±1.96SD (Eq. 4)**

## ECG VLF vs. Apple VLF BA Plot



**Figure 10: ECG and Apple very-low-frequency HRV B&A plot. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the LOA: Bias±1.96SD (Eq. 4)**

## ECG LF vs. Apple LF BA Plot
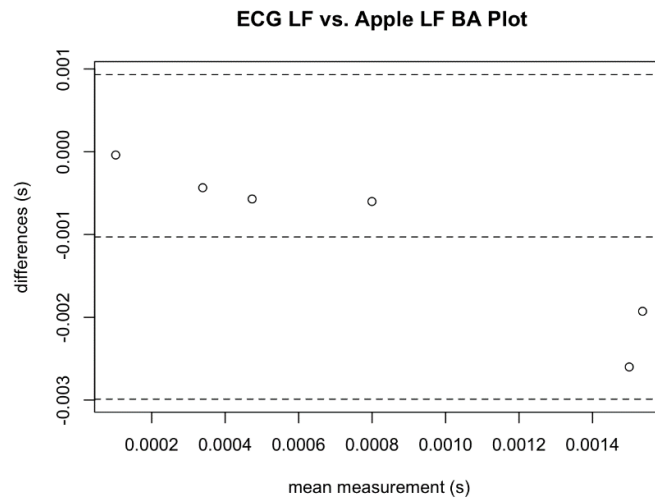


**Figure 11: ECG and Apple low-frequency HRV B&A plot. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the LOA: Bias±1.96SD (Eq. 4)**

B&A analysis. This is evidenced by both small bias value as well as smaller deviation from the ECG-derived RR (i.e., narrower 95% confidence interval).

The data reported here is obtained from 6 volunteers healthy, non-smoker, adult subjects. These preliminary results show that the level accuracy for the Apple Watch heart rate variability for the experimental setup in this study is comparable to the gold standard ECG. A paired T-Test of the mean values of the Apple Watch-derived and ECG-derived HRV metrics showed that for p= 0.05 (Table 1) no statistically significant difference between the means was detected. This lack of difference in the means was the

case for both the time-domain (i.e., SD of RR intervals) as well as the four frequency-domain metrics.

The results shown in this paper are for a relatively short recording (i.e., 5 min), but there is some evidence in the literature that short recording can still be useful [2]. Due to their relative simplicity of recording, 5 min of HRV have been broadly used and studied for several years, and mostly found to be a source of published HRV data [2]. Nonetheless, it should be mentioned that the recording length period highly influences time-domain values [5]. Shorter periods are associated with smaller values and poorly estimate 24 h

**ECG HF vs. Apple HF BA Plot**
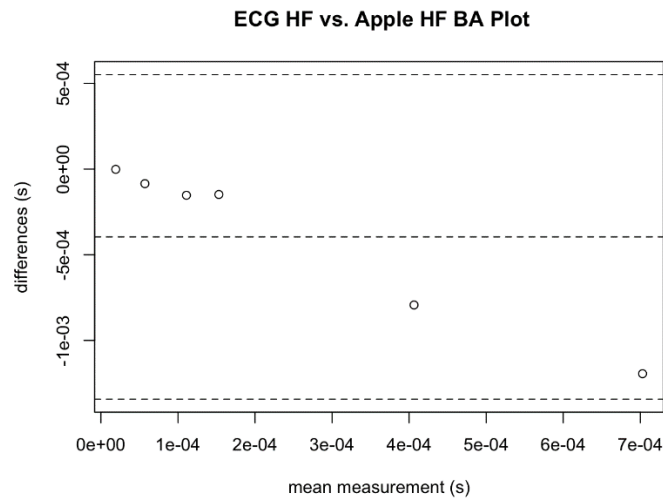


**Figure 12: ECG and Apple high-frequency HRV B&A plot. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the LOA: Bias ±1.96SD (Eq. 4)**

**ECG HRV Freq vs. Apple HRV Freq BA Plot**
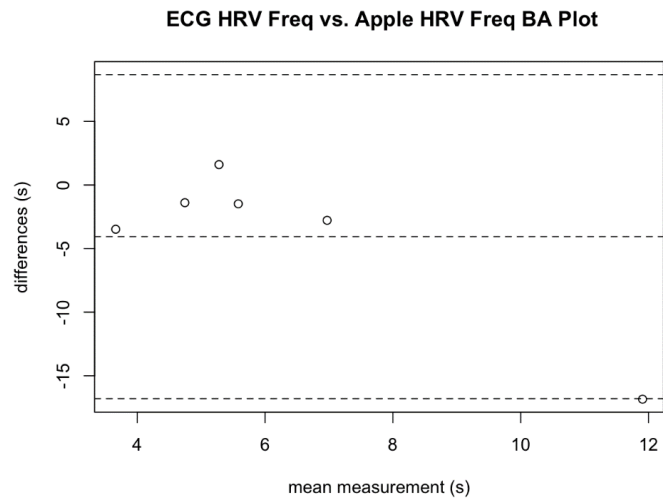


**Figure 13: ECG and Apple HRV B&A plot in the frequency domain. The middle line shows the Bias (Eq. 3) and the top and bottom lines represent the LOA: ±1.96SD (Eq. 4)**

HRV time-domain metric values [5]. Furthermore, short-term time-domain metric values are only suitable when a subject is breathing normally [5]. The recordings that were obtained in this study contained a transition of the heart rate at the beginning of the recorded session Figures 2 and 3, as before the subjects starting the sitting session for the collection of data they were walking.

HRV measurement in the time domain is measured by taking the standard deviation of the RR intervals, the B&A plot showed that measurements by the ECG and Apple Watch have a reasonable agreement. The power spectrum density plots (Figures 6 and 7) of the interpolated heartrate signal from ECG and Apple watch also

show a similar profile for the spectra for the ECG and Apple Watch measurements, particularly at frequencies greater than 0.0033 Hz, where the frequency-domain metrics are computed. As shown in Table 1, the frequency domain HRV metrics that were considered in this paper were the very-low-frequency (VLF, 0.0033–0.04 Hz), low frequency (LF, 0.04–0.15 Hz), and high frequency (HF, 0.15–0.40 Hz) of the signal power of the interpolated heart rate waveforms, and the ratio of LF/HF. Almost all of the B&A analysis for the frequency domain metrics indicated that the measurements are within the limits of agreement, as illustrated in Figure 10, 11, 12 and 13, except

one of the measurements in the LF/HF (Figure 11) which is slightly outside the lower limit of agreements.

## 6 CONCLUSION

Based on the preliminary results presented in this study, the RR values obtained from the Apple Watch follow those obtained from ECG. Moreover, the time-domain and frequency-domain metrics of HRV in the sample population presented in this paper suggest that it may be possible to estimate HRV metrics using Apple Watch in healthy subjects. However, a study with more subjects is needed to further establish this. As this study revealed promising results, in our future investigations, we are planning to test this method in a different population, for example, traumatic brain injury (TBI) patients to study the efficacy of measuring HRV using Apple Watch in the TBI patient population.

## REFERENCES

[1] G. Forte, F. Favieri and M. Casagrande, "Heart Rate Variability and Cognitive Function: A Systematic Review," *Frontiers in neuroscience,* vol. 13, 2019.

[2] E. Lam, S. Aratia, J. Wang and J. Tung, "Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study," *JMIR,* 2020.

[3] M. L. King, S. W. Lichtman, G. Seliger, F. A. Ehert and J. S. Steinberg, "Heart-rate variability in chronic traumatic brain injury," *Brain injury,* vol. 11, no. 6, 1997.

[4] D. Hernando, S. Roca, J. Sancho, Á. Alesanco and R. Bailón, "Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects," *Sensors,* vol. 18, no. 8, 2018.

[5] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in public health,* vol. 5, 2017.

[6] A. Malliani, "Heart rate variability:A challenge for a new way of thinking," *Journal of cardiac failure,* vol. 2, no. 3, 1996.

[7] S. Hillebrand, K. B. Gast, R. de Mutsert, C. A. Swenne, J. W. Jukema, S. Middeldorp, F. R. Rosendaal and O. M. Dekkers, "Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose–response meta-regression," *Europace : European pacing, arrhythmias, and cardiac electrophysiology,* vol. 15, no. 5, 2013.

[8] S. W. Porges, "Cardiac vagal tone: a physiological index of stress," *Neuroscience and Biobehavioral Reviews,* vol. 19, no. 2, 1995.

[9] A. L. Hansen, B. H. Johnsen and J. F. Thayer, "Vagal influence on working memory and attention. Int. J. Psychophysiol," *International Journal of Psychophysiology,* vol. 48, no. 3, 2003.

[10] S. Duschek, M. Muckenthaler, N. Werner and G. A. Reyes del Paso, "Relationships between features of autonomic cardiovascular control and cognitive performance," *Biological Psychology,* vol. 81, no. 2, 2009.

[11] J. F. Thayer and R. D. Lane, "A model of neurovisceral integration in emotion regulation and dysregulation," *Journal of Affective Disorders,* vol. 61, no. 3, 2000.

[12] G. A. Reyes Del Paso, M. I. González, J. A. Hernández, S. Duschek and N. Gutiérrez, "Tonic blood pressure modulates the relationship between baroreceptor cardiac reflex sensitivity and cognitive performance," *Psychophysiology,* vol. 46, no. 5, 2009.